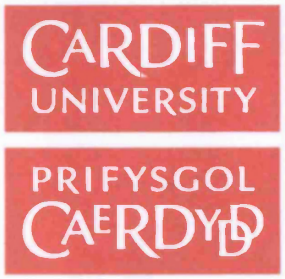
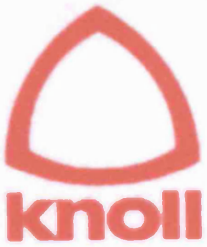
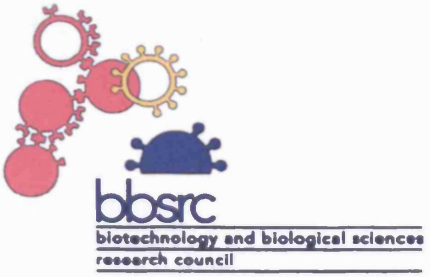






RETROVIRAL VECTOR MEDIATED GENE TRAPPING IN MICE

Thesis submitted for the Degree of Doctor of Philosophy by

Julie A. Wilkins

Cardiff University

2003

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DEDICATED TO DAD,

HE WOULD HAVE LIKED THIS!

Acknowledgements

Firstly, thank you to my supervisor Prof. Martin Evans, for advice and all round wisdom supplied throughout this period of study. Financial assistance was provided by the Biotechnology and Biological Sciences Research Council (BBSRC) and also Knoll Pharmaceuticals, to whom I am extremely grateful.

Thanks to John Dixon, CRC, Cambridge for supplying us with the retroviral constructs being studied in this project.

Bundles of thanks must go to lab C.401, Susie, Nicky, Rach, Anna, Fiona, (Michelle) and Marleeny, there's not many of us but as they say quality, not quantity. Thank you Susie for the cells and constant support; Nicky for the perfect coffees and many encouraging pep talks; Rach for being a fellow Herefordian and a terrific pal; Anna for being so generous with your magic in the lab and a great ally; Fiona and Mike for your constant guidance and a fabulous trip to Ireland, just what was needed and of course Marleeny - your collection of humorous anecdotes, THANK YOU ALL!

To my adopted adviser, Prof. Alan Clarke, thank you for supplying those all important boosts when required, greatly appreciated. James, your advice concerning my DNA concentrations will never be forgotten, thank you for saving my sanity!

Loads of thanks to Andy, my computer whizz and Jab Jab, my escape to the outside world, loves you both!

A special thank you has to go to Val, Owen and Andy. I can honestly say, without your constant support, loyalty and friendship, this thesis would never have been written. Thank you all, so much!

And lastly, but by no means least, thanks to my family. My brothers who wont believe this until they see it and to Mum, who is an infinite source of encouragement and inspiration, couldn't have done it without you, thanks Mum, love you!

ABBREVIATIONS AND CONVENTIONS

The following abbreviations are used throughout the text:

A	Adenosine
A(x)	Absorbance at x nm
AS	Alternative splicing
β-gal	Beta galactosidase
β-geo	Beta galactosidase and neomycin phosphotransferase fusion transcript
Blast	Basic local Alignment search Tool
bp	nucleotide base pairs
BSA	Bovine Serum Albumin
C	Cytosine
cDNA	Complementary DNA
CMV	Cytomegalovirus
DMSO	Dimethyl Sulfoxide
ds	double stranded
DTT	DiThioThreitol
DNA	Deoxyribonucleic Acid
EDTA	Ethylene Diamine Tetracetic Acid
ES	Embryonic stem cell
EST	Expressed Sequence Tag
FCS	Foetal Calf Serum
g	Gravity
G	Guanine
gDNA	Genomic DNA

GFP	Green fluorescent protein
HEPES	(N-[2-HydroxyEthyl]-Piperazine-N'-[2-EthaneSulphonic acid])
IRES	Internal ribosome entry site
kb	Nucleotide kilobase pairs
LB	Luria Broth (Luria Bertani Media)
LiDS	Lithium dodecyl sulfate
LIF	Leukaemia Inhibitory Factor
LTR	Long Terminal Repeat
M	Molar concentration
MEM	Modified Eagles Medium
min	Minutes
Mm	Mus musculus
MoMLV	Moloney Murine Leukaemia Virus
mRNA	Messenger RNA
n	Number of samples analysed
NCBI	National Centre for Biotechnology Information
NCS	Newborn Calf Serum
neo	Neomycin phosphotransferase
nt	Nucleotide
OD	Optical density
PBS	Phosphate Buffered Saline
pc	post coitum
PCR	Polymerase chain reaction
pp	post partum
RACE	Rapid amplification of cDNA ends

RNA	Ribonucleic acid
RT	Room temperature
RT-PCR	Reverse transcription – polymerase chain reaction
SA	Splice acceptor site
SD	Splice donor site
sec	Seconds
ss	Single stranded
STO	(SIM [Sandoz inbred Swiss mouse] thioguanine-resistant ouabain-resistant) cells
T	Thymine
TAE	Tris Acetate – EDTA buffer
TE	Tris-EDTA buffer
Tm	Melting temperature
Tris	N-Tris[HydroxyMethyl]methylglycine; N-[2-Hydroxy-1, 1-bis(hydroxymethyl)-ethyl]glycine
tRNA	Transfer RNA
U	Uracil
UTR	Untranslated region
UV	Ultraviolet
w/v	Weight by Volume
v/v	Volume by Volume

The one and three letter abbreviations for nucleotides and amino acids are those recommended for use by the Biochemical Journal. Policy of the Journal and Instructions to Authors. (1983). *Biochemical Journal.*, 209 ,1-27.

All laboratory materials employed are detailed according to their commercial name as given in the appropriate manufacturer's catalogue.

Summary

There is a great wealth of sequence information available nowadays, but this is accompanied by a serious lack of functional information. Functional genomics is an area that needs to be developed and one method being utilised is gene trapping.

In this project a gene trapping approach was employed to achieve insertional mutagenesis *in vitro* in embryonic stem (ES) cells. These early embryo derived cell lines can be manipulated *in vitro* and then returned to the embryo where they participate in the normal development of a chimeric mouse.

There are two types of gene trapping retroviral vectors being investigated, one is a shuttle vector which contains plasmid backbone between the long terminal repeats (LTRs) allowing the rescue of any trapped gene. The other is a splice acceptor (SA) vector, which has a SA infront of a promoterless β-geo gene. The provirus integrates into the genome of the ES cells and the trapped gene is tagged with the β-geo reporter gene. This enables trapped clones to be selected with G418 and expression patterns to be visualised by staining for β-galactosidase activity. Generation of a fused RNA transcript between the trapped gene and vector sequences facilitates cloning of the trapped gene. Using the β-geo sequence of the integrated vector, 3′ RACE PCR was used to amplify a segment of the trapped gene and subsequently obtain sequence data.

Two retroviral vectors of each type mentioned are examined in this project for their insertional mutagenesis ability *in vitro* in ES cells and subsequent analysis. Neither of the shuttle retroviral vectors gave any reproducible results. A comparison was made between the SA retroviral vectors, one containing an internal ribosome entry site (IRES), the other not. These vectors gave rise to resistant ES cell clones and subsequent 3′ RACE PCR sequence data.

CONTENTS

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1. Functional Genomics

The recent sequencing of the human genome (Lander et al., 2001) stands as a landmark accomplishment that heralds enormous potential. Besides providing the blueprint for all human proteins, the sequence contains all the regulatory elements that govern the developmental interpretation of the genome.

This achievement, however, is not the end of the road but rather the first step toward the functional understanding of the genome. The determined linear nucleotide sequences remain only lists of A, C, G and T, unless they are given functional significance. The coding sequences of genes can be investigated by computational methods and comparisons can be made with homologous genes in different species to try and apply functional role. However, the exact function of protein products can rarely be determined without obtaining much additional information, for example, biochemical or cell biological methods. In reported cases, some homologous genes fulfil completely different functions in different species.

An impressive demonstration of this phenomenon has been given by the functional comparison of similar molecules in the two closely related nematodes *Pristionchus pacificus* and *Caenorhabditis elegans*, which revealed that homologous genes can be recruited to serve completely new functions in new regulatory linkages (Eizinger et al., 1999). While remaining in the original developmental program, genes were found to have changed their molecular specificity and at the same time, retained other functions. Genes have jumped in or moved out of regulatory circuits and are suddenly expressed in a spatially and temporarily different manner, demonstrating that mere sequence or structural similarity does not necessarily allow conclusions on possible protein functions (Oliver, 1996).

However, it must be said that comparative analysis of genomic sequences is important as a tool to elucidate coding and regulatory regions in genomic DNA. Several comparative studies have been done on large sequence regions of human and various species (Hardison et al., 1997). It is evident from these studies that coding regions are generally well-conserved (Jareborg et al., 1999).

With so much genomic information now available and especially since the advent of the mouse genome sequence (Waterston et al., 2002) it should be taken that orthologous gene information can be used as a guide to gene function in different species, but it remains clear that the actual function of each gene needs to be studied in the entire organism. Critical information for defining gene function is often provided by generating mutations and then determining the phenotypic consequence of the given mutation on the organism under study. However, it must be considered that in diploid organisms both copies of a recessive gene must be mutated for the organism to display a mutant phenotype. A genome that contains two copies of each gene, each copy located on a separate homologous chromosome, could be protected from inactivating mutations in essential genes (Mills and Bradley, 2001). Traditionally, the phenotype-driven approach (forward genetics) has been most fruitful in organisms with comparably small genomes (Floss and Wurst, 2002).

This project began with plans for a phenotype driven screen, but due to certain methodological constraints, it developed into a genotype driven screen, where sequence information was analysed prior to micro-injection of resistant ES cells into host blastocysts.

The number of protein coding genes in an organism provides a useful first measure of its molecular complexity. Single celled prokaryotes and eukaryotes typically have a few thousand genes; for example, *Escherichia coli* (Blattner et al., 1997) has 4,300 and *Saccharomyces cerevisiae* (Goffeau et al., 1996) has 6,000. Evolution of multicellularity appears to have been accompanied by a several-fold increase in gene number. However, many were surprised when published drafts of the human genome chopped tens of thousands of genes off the long-held notion that making a human might require about 100,000 genes. The International Human Genome Sequencing Consortium found evidence for 29,691 human transcripts (Lander et al., 2001). The commercial genome project of Celera Genomics, Rockville, found 39,114 genes (Venter et al., 2001). These predictions are in line with a previous estimate (Ewing and Green, 2000) of 35,000 genes, gained by comparing a set of human expressed sequence tag (EST) contigs with human chromosome 22 and with a non-redundant set of mRNA sequences. This disparity indicates that the number of human expressed sequence (mRNA) forms was much higher than the number of genes, suggesting that

evolution of the increased physiological complexity of vertebrates may, therefore, have depended more on the combinatorial diversification of regulatory networks or alternative splicing, than on substantial increase in gene number (Ewing and Green, 2000).

Following the completion of sequencing the human genome (Lander et al., 2001), the next step must be to assign function to identified genes. The final goal of genome research today may look futuristic, but the knowledge of the function of every single gene and the interactions between them will finally allow us to understand the development and functioning of an organism as a whole. This effort will enable understanding of the molecular mechanisms underlying normal development as well as to delineate pathways where subversion of these processes culminates in human disease (Bronson and Smithies, 1994; Mills and Bradley, 2001).

1.2. Alternative splicing

Alternative splicing is an important cellular mechanism that leads to the temporal and tissue-specific expression of unique mRNAs. It was first predicted by Walter Gilbert (Gilbert, 1978) that different combinations of exons could be spliced together "alternative splicing" to produce different isoforms of a gene. This was subsequently verified by the discovery of cDNA isoforms exhibiting the addition or exclusion of whole or partial exons (Adams et al., 1996; Breitbart and Nadal-Ginard, 1987), although identification of such splice variants largely occurred on an *ad hoc* basis.

The development of large EST sequence libraries over recent years, however, provides an opportunity to examine the incidence of alternative splicing globally by searching these libraries for exon skipping, exon truncation or inclusion of sequences currently described as intronic. As ESTs are derived from fully processed mRNA, that is, after 5′ capping, splicing and polyadenylation, they provide a broad sample of mRNA diversity. In the last few years, bioinformatic studies have identified an order of magnitude more alternatively spliced genes than were found in the past 20 years and are beginning to provide a global view of alternative splicing in humans (Modrek and Lee, 2002).

Studies have consistently reported a high rate of alternative splicing in the human genome, with 35-59% of human genes showing evidence of at least one alternative splice form (Kan et al., 2001; Mironov et al., 1999). Moreover, given that the vast number of ESTs have not been adequately assessed for alternative splice variants, it seems possible that more alternative splicing exists. These studies indicate that alternative splicing is far more abundant, ubiquitous and functionally important than previously thought and there are more types of mRNA isoforms. For instance, bioinformatic studies have reported that about 25% of genes have alternative polyadenylation forms, which are mRNAs that have been cleaved and polyadenylated at different sites (Beaudoing et al., 2000).

1.3. Experimental Animal Model

The history of each of the model organisms is slightly different. Beginning in the 1940s, yeast had proven to be useful as a classical genetic model and in the 1980s it began to be used as a molecular genetic model (Schweizer et al., 1986). In the last decade, yeast has been used for its genomic advantage. The worm is the youngest model system, just turning 40. It is a very simple animal, having only 959 somatic cells. The complete cell lineage and wiring of the worm's nervous system (302 cells) are known and it was the first multicellular organism to have its genome sequenced (Sulston et al., 1992). The fly has been a favoured system for genetic studies for more than 90 years and has proven to be an excellent model to identify genes involved in evolutionary conserved development and cellular processes (http://flybase.bio.indiana.edu/). Finally, the mouse has been used as a model for genetic studies for as long as the fly. Fancy breeders of different coat colour variants were the first mouse geneticists, whose work was quickly recognised by different scientific fields in biology (http:www.informatics.jax.org/).

To understand fully the molecular nature underlying the biology of human genetic disease, which is a prerequisite before a therapeutic goal can be considered feasible, it is necessary to employ the use of experimental animals as model systems. This enables gene function to be studied *in situ*, where individual proteins play out their roles through protein interactions, signalling responses and metabolic outputs, in the context of multiple cell types and organs and different physiological and disease

states. Observations recorded in the course of *in vitro* experimentation with mammalian cell lines, although certainly highly informative as regards cellular biology, cannot be directly extrapolated as being of relevance to the genetic disorder *in vivo*. The model species of choice in this field of experimental science is the common house mouse, *Mus musculus*.

The mouse has been the experimental model generally chosen for defining gene function because of it's anatomical, physiological, reproductive and genomic similarity to humans. The biology of this species is of far more physiological relevance to human biology than that of the other "classical" genetic model systems, such as those mentioned above. The mouse's small size, fertility and short gestation period make their practical handling and storage convenient for often necessary large scale breeding programmes. Females breed well in the laboratory with an average of 5-10 pups per litter and an immediate postpartum oestrus. It is also advantageous that the fathers do not harm their young and so breeding pairs can be maintained together after litters are born. For developmental studies, the deposition of a vaginal plug allows an investigator to time all pregnancies without actually witnessing the act of copulation.

By the 1980s, the experimental embryology and molecular biology of the mouse had reached a point where their parallel development allowed a fruitful fusion. Sufficient understanding of gene structure and *in vitro* DNA sequence tailoring, combined with the efficient *in vitro* manipulation of the mouse embryo at pre-implantation stage, led to the first production of transgenic mice expressing a viral (Brinster et al., 1981) and a human (Palmiter et al., 1982) gene. Introduction of exogenous, actively expressed transgenes into the mouse was the first and most extensively used genetic approach. It has provided invaluable information about the role of oncogenes and genes involved in development or cell function in normal and disease processes. Most of these transgenic studies were gain-of-function approaches. The advent of mouse embryonic stem (ES) cells (Evans and Kaufman, 1981; Martin, 1981) opened up the possibility of loss-of-function forward and reverse genetics.

The mouse has now become the premier mammalian model organism (Anderson and Ingham, 2003) for studying development and disease. The mouse genome sequence

(Waterston et al., 2002) is now available and the mouse germline can be manipulated in almost every way imaginable, in both a random (Brown and Balling, 2001; Nolan et al., 2000; Stanford et al., 2001) and a direct fashion (Nagy et al., 2003; Yu and Bradley, 2001), to study the phenotypic consequences of perturbations. As techniques for manipulating the genome have multiplied, so have the tools for the study of complex phenotypes (Mills and Bradley, 2001).

1.4. Embryonic Stem Cells

Embryonic stem (ES) cells are established *in vitro* from explanted blastocysts and retain their normal karyotype in culture (Evans and Kaufman, 1981; Martin, 1981). They are the *in vitro* counterparts of an *in vivo* population of cells, known as the epiblast, which are specific to the early embryo (Burdon et al., 2002; Evans and Hunter, 2002).

These cells are pluripotent, which means that an individual cell can give rise to all cell types of the foetus. ES cells retain the developmental identity and potential of the epiblast even after prolonged culture (Burdon et al., 2002). This phenomenon has been shown conclusively, as when injected into host blastocysts, ES cells can colonise the embryo and generate chimeric animals that are mosaic in all their tissues. Mosaicism extends to the germ cell lineage and ES cells can contribute fully functional gametes (Bradley et al., 1984). Consequently ES cells are continually exploited as vehicles for introducing genetic modifications into the mouse germline (Akiyama et al., 2000; Bronson and Smithies, 1994; Gossler et al., 1989; Robertson et al., 1986).

The pluripotentiality of ES cells and their capacity to respond to normal development signals, implies that they constitute a particularly appropriate system for the experimental investigation of mammalian development. ES cells provide a resource that circumvents the inaccessibility of the early mammalian embryo to conventional biochemical approaches. (Nichols et al., 1990).

A fundamental problem of studying mammalian developmental genetics is that unlike other model organisms, such as flies and frogs, the murine embryo can not be readily

observed as it develops. Detecting early phenotypic defects caused by mutations is therefore difficult. This is further complicated by "*in utero*" reabsorption of nonviable offspring (Baker et al., 1997). With the breakthrough of the ES cell system, it suddenly became feasible to manipulate mouse embryonic cells *ex vivo* and generate animals carrying a single predetermined mutation. "Reverse genetic" methods, in which a gene is characterised before its mutant phenotype is known, have primarily been used to study the molecular basis of murine development. The two principal reverse genetic approaches both modify the genome of ES cells, which are eventually introduced into blastocysts to secure germline transmission of mutant alleles (Hill and Wurst, 1993; Joyner, 1991; Soriano, 1995).

One approach, targeted or site-directed mutagenesis (section 1.6.) was designed to disrupt the function of specific murine genes, which have been identified by virtue of their homology to genes of other organisms. The principle drawback of this approach is that the sequence of the targeted gene must be known in order to carry out insertional or replacement mutagenesis (Bronson and Smithies, 1994; Soriano, 1995).

The other approach, gene trapping (section 1.7.) was pioneered in mammalian cells by combining both, insertional mutagenesis and the ES cell system (Gossler et al., 1989). Gene trapping was designed to identify novel, developmentally regulated genes in mouse embryos (Friedrich and Soriano, 1991; Skarnes et al., 1992; Wurst et al., 1995).

Historically, ES cells have been isolated and maintained on feeder layers of mitotically inactivated embryonic fibroblasts (Robertson, 1987). The feeder cells are thought to have generalised effects, such as medium detoxification, in addition to more specific effects in promoting cell attachment and viability and preventing ES cell differentiation. However, the presence of feeders is not always ideal when attempting genetic manipulation of the ES cells (Hooper et al., 1987).

ES cells undergo symmetrical self-renewal, which means they produce two identical stem cell daughters when they divide. Self-renewal entails the suppression of differentiation during proliferation (Burdon et al., 2002). The differentiation of ES cells in vitro can be inhibited by the presence of leukaemia inhibitory factor (LIF) or

related cytokines that can activate signal transduction from cell-surface receptors. The actions of LIF are mediated via heterodimerisation of two members of the class I cytokine receptors, the low affinity LIF receptor (LIF-R) and gp130 (Davis et al., 1993; Gearing et al., 1991) LIF sustains self-renewal through activation of the transcription factor STAT3 (Niwa et al., 1998).

Currently there is considerable interest in the prospect of exploiting this potential in analogous human pluripotent cells (Thomson et al., 1998) to generate specific, differentiated types of cell for drug development, for therapies based on cell replacement and for delivering gene therapies (Xu et al., 2001).

1.5. Retroviral Insertional Mutagenesis

In 1976, the introduction of exogenous retroviral DNA into the mouse germ line was first reported (Jaenisch, 1976) and 'insertional mutagenesis" began to be hotly pursued in the mouse. As retroviral infection usually occurred after the one- or two-cell stage of embryonic development, multiplicity of infection was extremely variable and could lead to the generation of chimeric animals. However, a mutation could be recovered by additional breeding if an insertion was transmitted through the germ line (Jenkins and Copeland, 1985; Spence et al., 1989). This insight that retroviral insertion could alter endogenous genes and their expression, leading to tumorigenesis and leukaemia (Jaenisch, 1976) in mutagenized mice, spurred the cloning of retroviral insertion sites to recover the affected, tumorigenic genes and validated the potential usefulness of insertional mutagenesis.

1.5.1. Retroviral Life cycle

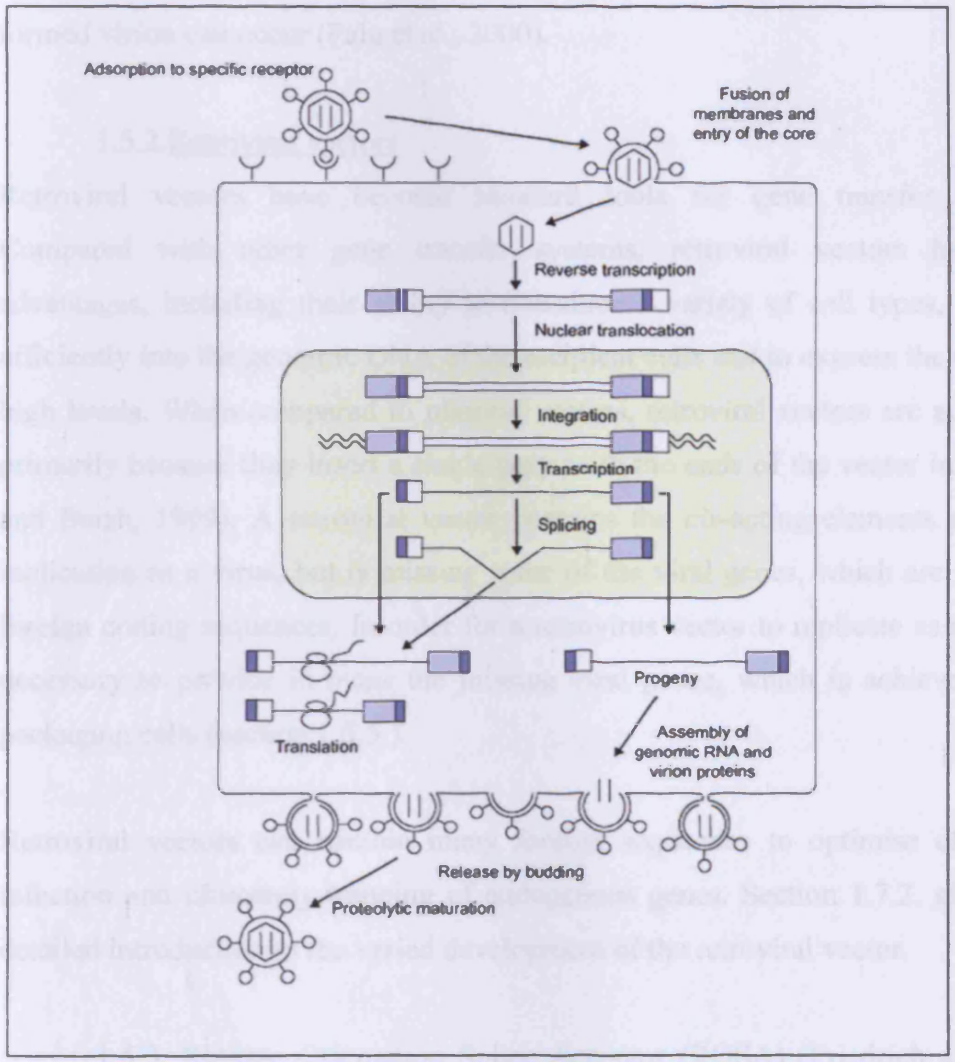
Retroviruses are a family of RNA enveloped viruses with the ability to reverse transcribe their genome from RNA to DNA (Weiss, 1998). Virions measure about 120nm in diameter and contain two identical copies of positive strand RNA genome complexed with nucleocapsid proteins. Reverse transcriptase, integrase and protease proteins are also contained in the nucleocapsid, the inner portion of the virion. A protein shell, formed by capsid proteins, encloses the nucleocapsid and delimits the viral core. Matrix proteins form a layer outside the core and interact with an envelope consisting of lipid bilayer, which surrounds the viral core particle. The envelope

originates from the cellular membrane and incorporates viral envelope glycoproteins. The envelope glycoprotein, responsible for the virus interaction with the specific receptor, is the only viral protein on the surface of the particle. It is formed by two subunits held together by di-sulphide bonds and noncovalent interactions. One transmembrane subunit anchors the protein into the envelope membrane and the surface subunit contains the receptor binding function (Palu et al., 2000).

The retroviral life cycle is divided into two distinct phases, an early phase involving entry, reverse transcription of the viral RNA into DNA and the insertion of the DNA into the host genome to establish the integrated provirus. Late phase involves the expression of the viral RNAs and proteins, the assembly of the virion particles and the release of these particles by budding through the plasma membrane (fig. 1.1.) (Goff, 2001).

The infection process begins with the interaction of the viral particles with the cell surface. The binding of the viral glycoprotein to a specific receptor complex on the cell surface causes virus and cell membranes to fuse and the viral particle to be internalised. The cellular tropism depends on the native envelope glycoprotein, for example ecotropic MoMLV, being utilised in this study, has a strictly murine host range and native receptor is the cationic amino acid transporter CAT-1 for arginine, lysine and orthinine (Kim et al., 1991; Wang et al., 1991). After virus-cell membrane fusion, the virus core is released into the cytoplasm. The viral core is partially degraded to form a large nucleoprotein particle containing the viral genome (pre-integration complex) and is transported into the nucleus. The process of entry differs between oncoretroviruses and lentiviruses (Lewis and Emerman, 1994). For HIV-1 it was established that efficient transport to the nucleus is active and uses the cellular nuclear import machinery (Bukrinsky et al., 1992). However, MoMLV cannot transit through the pores of the nuclear membrane, it has to wait until the nuclear membrane dissolves during cell division. These retroviruses are mitosis-dependent which means the pre-integration complexes will not infect non-dividing cells (Roe et al., 1993).
As the vectors being studied in this project are based on MoMLV, it was essential that the ES cells being infected were actively proliferating during infection.

Figure 1.1. Schematic representation of the retroviral life cycle (adapted from *Retroviruses*, edited by JM Coffin, SH Hughes and HE Varmus, Cold Spring Harbor Laboratory Press, 1997).



Upon entry into the nucleus the proviral genome flanked by two identical long terminal repeats (LTRs) containing the *cis*-acting regions, is randomly integrated into the host genome by the viral integrase. Synthesis of viral RNA by the cellular RNA polymerase, can then be initiated from the LTR promoter. However, in the case of the vectors studied in this project the LTR promoter is deleted and synthesis is driven by a cytomegalovirus (CMV) promoter situated within the bacterial backbone of the vector (see Appendix I for vector diagram). A single RNA species is produced which can then be processed into genomic RNA or multiply spliced, giving rise to individual

mRNA transcripts. Translation of viral proteins can then begin and this is followed by the formation of virion particles at the cell surface. The capsid and envelope are thought to assemble at the same time and once the viral genome is packaged into the nascent virion, along with the non-structural viral proteins, budding of the newly formed virion can occur (Palu et al., 2000).

1.5.2. Retroviral vectors

Retroviral vectors have become standard tools for gene transfer technology. Compared with other gene transfer systems, retroviral vectors have several advantages, including their ability to transduce a variety of cell types, to integrate efficiently into the genomic DNA of the recipient cells and to express the transgene at high levels. When compared to plasmid vectors, retroviral vectors are advantageous primarily because they insert a single copy with the ends of the vector intact (Neilan and Barsh, 1999). A retroviral vector contains the *cis*-acting elements required for replication as a virus, but is missing some of the viral genes, which are replaced by foreign coding sequences. In order for a retrovirus vector to replicate as a virus, it is necessary to provide in *trans* the missing viral genes, which is achieved by using packaging cells (section 1.5.5.).

Retroviral vectors can contain many foreign sequences to optimise efficiency of infection and ultimately trapping of endogenous genes. Section 1.7.2. gives a more detailed introduction to the varied development of the retroviral vector.

1.5.3. Reverse Orientation Splice Acceptor (ROSA) (Friedrich and Soriano, 1991)

Constructs lacking a promoter and including a reporter gene (β-geo) encoding a fusion protein with both β-galactosidase and neomycin phosphotransferase activity were designed, so that activation of the reporter gene depends on proviral insertion downstream of a flanking genomic promoter. To allow for expression of the reporter gene when inserted into an intron, a splice acceptor sequence was included in the construct. In order to aid cloning and analysis, the SAβ-geo construct was inserted into a self-inactivating retroviral vector pGen$^-$. This vector lacks the viral enhancer sequences and contains a bacterial *supF* gene in the LTRs to facilitate cloning of the

provirus and flanking sequences (Soriano et al., 1991). The splice acceptor gene trap construct was inserted in reverse orientation relative to the viral transcription. This was necessary to avoid removing the viral Ψ packaging sequence by splicing from a viral upstream splice acceptor sequence. The resulting retroviral vector is termed ROSAβ-geo (reverse orientation splice acceptor β-geo).

1.5.4. MoMLV Long Terminal Repeats (LTRs)

The long terminal repeat (LTR) is the control centre for gene expression of retroviruses. As may be expected because of the integrated phase of their life cycle, retroviruses have somewhat typical eukaryotic promoters with transcriptional enhancers and some also have regulatory elements responsive to either viral or specialised cellular trans-activating factors, for example hormonal. All of the requisite signals for gene expression are found in the LTRs: Enhancer, promoter, transcription initiation (capping), transcription terminator and polyadenylation signal (fig. 1.2.). Expression directed by the viral LTR signals is carried out entirely by host cell enzymes.

The enhancer and other transcription regulatory signals are contained in the U3 region of the 5′ LTR and the TATA box is located roughly 25 bp from the beginning of the R sequence (Speck and Baltimore, 1987).

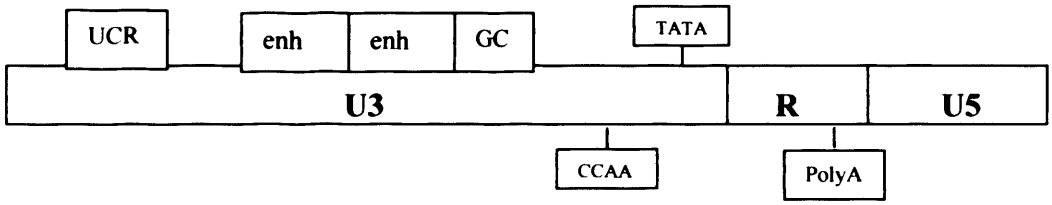


Figure 1.2. Schematic structure of the Moloney murine leukaemia virus long terminal repeat (LTR). U3, R and U5 are 449, 68 and 77 bp, respectively. Noted are the negative-acting upstream control region (UCR), the 75 bp direct repeats which have enhancer activity (enh), a GC rich region required for maximal enhancer activity, canonical promoter elements and the signal for polyadenylation (Poly A).

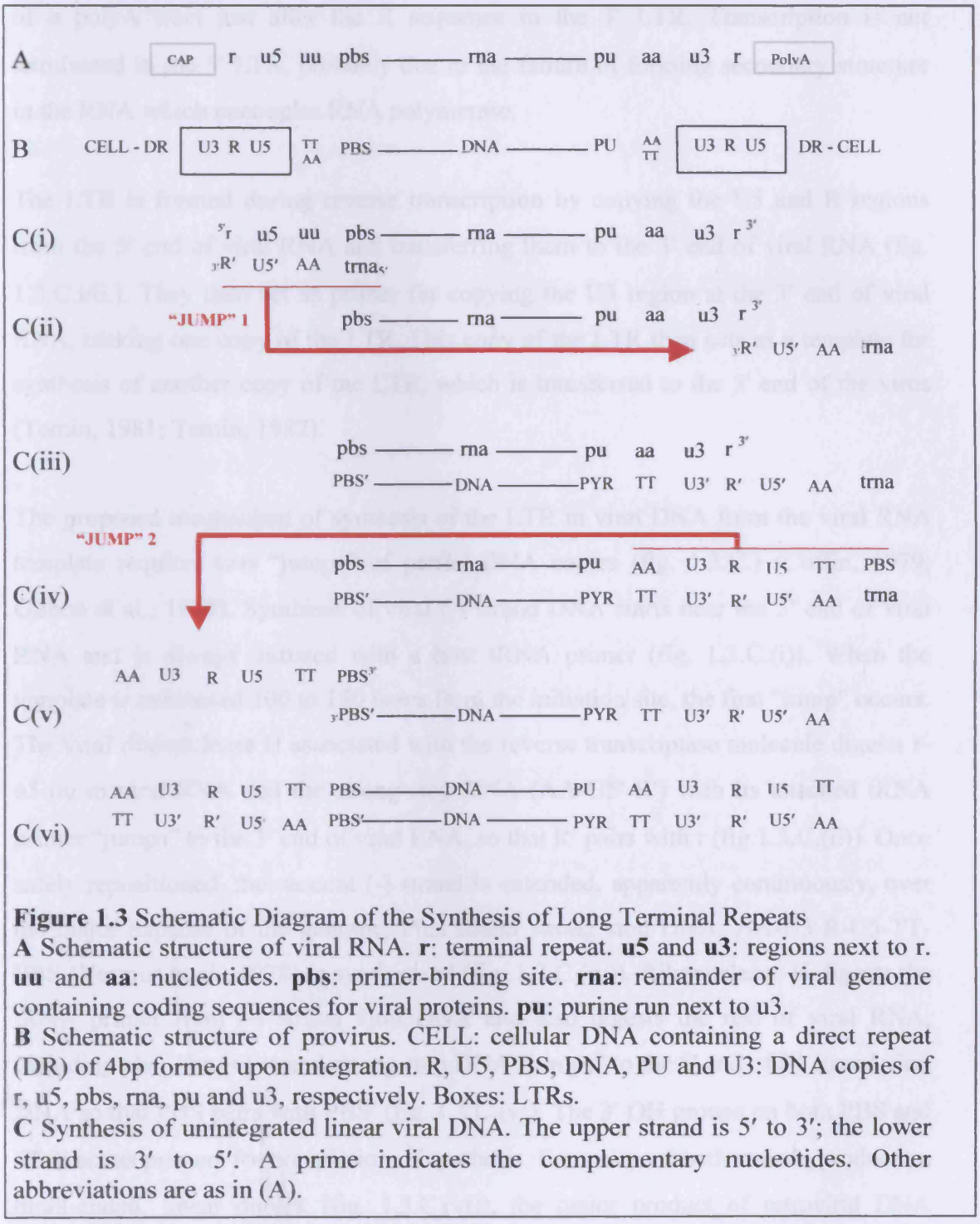


Figure 1.3 Schematic Diagram of the Synthesis of Long Terminal Repeats
A Schematic structure of viral RNA. **r**: terminal repeat. **u5** and **u3**: regions next to r. **uu** and **aa**: nucleotides. **pbs**: primer-binding site. **rna**: remainder of viral genome containing coding sequences for viral proteins. **pu**: purine run next to u3.
B Schematic structure of provirus. CELL: cellular DNA containing a direct repeat (DR) of 4bp formed upon integration. R, U5, PBS, DNA, PU and U3: DNA copies of r, u5, pbs, rna, pu and u3, respectively. Boxes: LTRs.
C Synthesis of unintegrated linear viral DNA. The upper strand is 5′ to 3′; the lower strand is 3′ to 5′. A prime indicates the complementary nucleotides. Other abbreviations are as in (A).

The integrated provirus has two LTRs and the 5′ LTR normally acts as an RNA polymerase II promoter, instructing the host polymerase when and where to initiate viral RNA synthesis. The transcript begins at the beginning of R, is capped and proceeds through U5 and the rest of the provirus, usually terminating by the addition

of a polyA tract just after the R sequence in the 3′ LTR. Transcription is not terminated in the 5′ LTR, probably due to the failure of forming secondary structure in the RNA which uncouples RNA polymerase.

The LTR is formed during reverse transcription by copying the U5 and R regions from the 5′ end of viral RNA and transferring them to the 3′ end of viral RNA (fig. 1.3.C.i/ii.). They then act as primer for copying the U3 region at the 3′ end of viral RNA, making one copy of the LTR. This copy of the LTR then acts as a template for synthesis of another copy of the LTR, which is transferred to the 5' end of the virus (Temin, 1981; Temin, 1982).

The proposed mechanism of synthesis of the LTR in viral DNA from the viral RNA template requires two "jumps" of partial DNA copies (fig. 1.3.C.) (Coffin, 1979; Gilboa et al.; 1979). Synthesis of viral (-) strand DNA starts near the 5′ end of viral RNA and is always initiated with a host tRNA primer (fig. 1.3.C.(i)). When the template is exhausted 100 to 180 bases from the initiation site, the first "jump" occurs. The viral ribonuclease H associated with the reverse transcriptase molecule digests r-u5-uu in viral RNA and the strong stop DNA (AA-U5′-R′) with its attached tRNA primer "jumps" to the 3′ end of viral RNA, so that R′ pairs with r (fig.1.3.C.(ii)). Once safely repositioned, the nascent (-) strand is extended, apparently continuously, over the major expanse of the genome. Plus strand strong stop DNA, AA-U3-R-U5-TT-PBS (Varmus et al., 1978) is synthesised (fig. 1.3.C.(iv)). Ribonuclease H digests the tRNA primer from (-) strand viral DNA and also digests the rest of viral RNA, including pbs. The (+) strand strong stop DNA "jumps" to the 5′ end of (-) strand viral DNA so that PBS pairs with PBS′ (fig. 1.3.C.(v)). The 3′ OH groups on both PBS and PBS′ act as primers for completion of synthesis. Extension of both strands produces a blunt-ended, linear duplex (fig. 1.3.C.(vi)), the major product of retroviral DNA synthesis in the cytoplasm of infected cells (Varmus, 1982).

Retrovirus vectors have been used to infect a variety of cell types in tissue culture as well as cells of the developing mouse embryo. Such vectors have been used to transduce and express foreign genes in infected cells and have been used as cell

lineage markers, as insertional mutagens and as a tag for specific chromosomal regions for many years now (Tsukiyama et al., 1989).

The MoMLV based retroviral constructs used in this project have enhancers and promoters deleted, relying on a cytomegalovirus (CMV) promoter (see Appendix I for vector diagrams) to drive replication in the packaging cells. The removal of these elements prevents any interference during the trapping process, for example, the viral promoter disturbing neighbouring genes and making interpretation of a mutant phenotype even more difficult.

1.5.5. Packaging Cells

Virions are complete virus packages, which consist of a protein coat surrounding the virus genome. This enables viral genomes to be carried from cell to cell, protecting them from inhospitable environments in which the virus cannot replicate (Bender et al., 1987). Packaging cell lines are designed to synthesise all retroviral proteins required for assembly of infectious virion, but should not produce any replication competent virus. However, some packaging cells that are initially helper free can convert to helper virus positive during prolonged passage. This is possibly due to recombination between sequences used to make the packaging cells and endogenous retrovirus like elements found in most eukaryotic cells. Also, introduction of particular retroviral vectors into some packaging cells can result in helper virus production probably due to recombination between homologous regions of the vector and the defective virus used to make the packaging cells (Miller, 1990; Weiss and Tailor, 1995).

Several early packaging cell lines were made by using helper viruses from which the retroviral packaging signal had been deleted (Cone and Mulligan, 1984; Mann et al., 1983; Miller et al., 1985). Although packaging cells of this type have been used with success in a number of studies, most notably the ψ-2 cell line (Mann et al., 1983), these lines suffer from important deficiencies. Firstly, although deletion of the packaging signal reduces the amount of helper virus RNA that is packaged into virion, the block is far from complete. If the deleted helper genome is packaged, there are no further blocks to infection of other cells, because the signals for reverse transcription

and integration are in tact. Indeed, virus from this type of packaging cell is capable of rescuing retroviral vectors in the absence of overt helper virus production (Cone and Mulligan, 1984; Miller and Buttimore, 1986), presumably due to the transfer of the deleted helper virus.

The retroviral vector used with a packaging line has a strong influence on the possibility of helper virus production. Avoidance of homologous overlap between vector and helper virus sequences in the packaging cells decreases the chance of helper virus production (Miller and Rosman, 1989; Miller et al., 1986). This can be accomplished by removing as much of the helper virus sequences from the vector as possible. In the case of retroviral vectors based on murine leukaemia virus, overlap at the 5′ end of the vector with helper sequences is difficult to avoid because the packaging signal of these viruses extends into the gag region of the helper virus (Adam and Dusty Miller, 1988; Armentano et al., 1987; Bender et al., 1987).

Alternatively, the areas of homologous overlap between vector and helper virus sequence in packaging cells can be avoided by using a vector derived from a helper virus that has little sequence similarity with the helper virus used to construct the packaging cells. For example, a vector based on spleen focus-forming virus (SFFV) has no tendency to produce helper virus when introduced into murine leukaemia virus based PA12 packaging cells, even though there is extensive overlap between the gag regions of the vector and helper DNA (Miller and Buttimore, 1986). In contrast, a vector based on the same murine leukaemia virus as the PA12 cells and having similar overlap as the SFFV-based vector readily produced helper virus in the PA12 cells (Miller and Buttimore, 1986). A possible difficulty with this strategy is potential incompatibility between a vector based on a different helper virus than the packaging cells, which might result in low titre.

In this project the BOSC (Pear et al., 1993) packaging cell line is being used to introduce in *trans* the necessary viral genes. To circumvent the problems of stable producer cell lines, this packaging cell line is transiently transfected to produce high-titre, helper-free infectious retroviruses. The retroviral vectors contain the *cis*-acting elements required for replication as a virus, but are missing some of the viral genes,

which are replaced by foreign coding sequences. The BOSC cells contain the *gag*, *pol* and *env* retroviral genes and are therefore capable of producing retrovirus from the experimental constructs on their own. However, there is probably only a single copy of each gene per BOSC cell and therefore greater titres can be achieved by co-transfecting pHIT plasmids which provide further copies of the *gag*, *pol* and *env* genes (see Appendix I for plasmid information). The pHIT 60 plasmid contains the murine leukaemia virus gag/pol cassette under the control of the human cytomegalovirus immediate early promoter, whereas pHIT 123 contains the human cytomegalovirus promoter driving murine leukaemia virus ecotropic envelope.

1.6. Gene Targeting

With the advent of "reverse genetic" methods to study the molecular basis of murine development, it was no longer necessary to rely on random mutagenesis for the generation of mutations. These are methods in which a gene is characterised before its mutant phenotype is known. In stark contrast to the hit and miss strategies is that of targeted mutagenesis by homologous recombination in ES cells. This approach modifies the genomes of ES cells, which are eventually introduced into host blastocysts to secure germline transmission of mutant alleles. Targeted or site-directed mutagenesis was designed to disrupt the function of specific murine genes which are known by virtue of their homology to genes of other organisms (Capecchi, 1989).

The ability to manipulate the mouse genome using gene targeting has revolutionised the capacity to assess mammalian gene function and to generate models for human diseases. Gene targeting in the mouse is extremely useful for determining the function of individual genes, but only a few thousand genes have been subjected to this analysis. Although it is technically possible to generate individual mouse strains that are deficient for each of the estimated 30,000 to 40,000 genes within the mouse genome, this endeavour would require around 100 years to achieve at the current rate (Mills and Bradley, 2001).

There are other drawbacks to this approach, one being that the targeted gene must be known in order to carry out insertional or replacement mutagenesis. Another is the inability to reliably predict the exact biological process that will be affected.

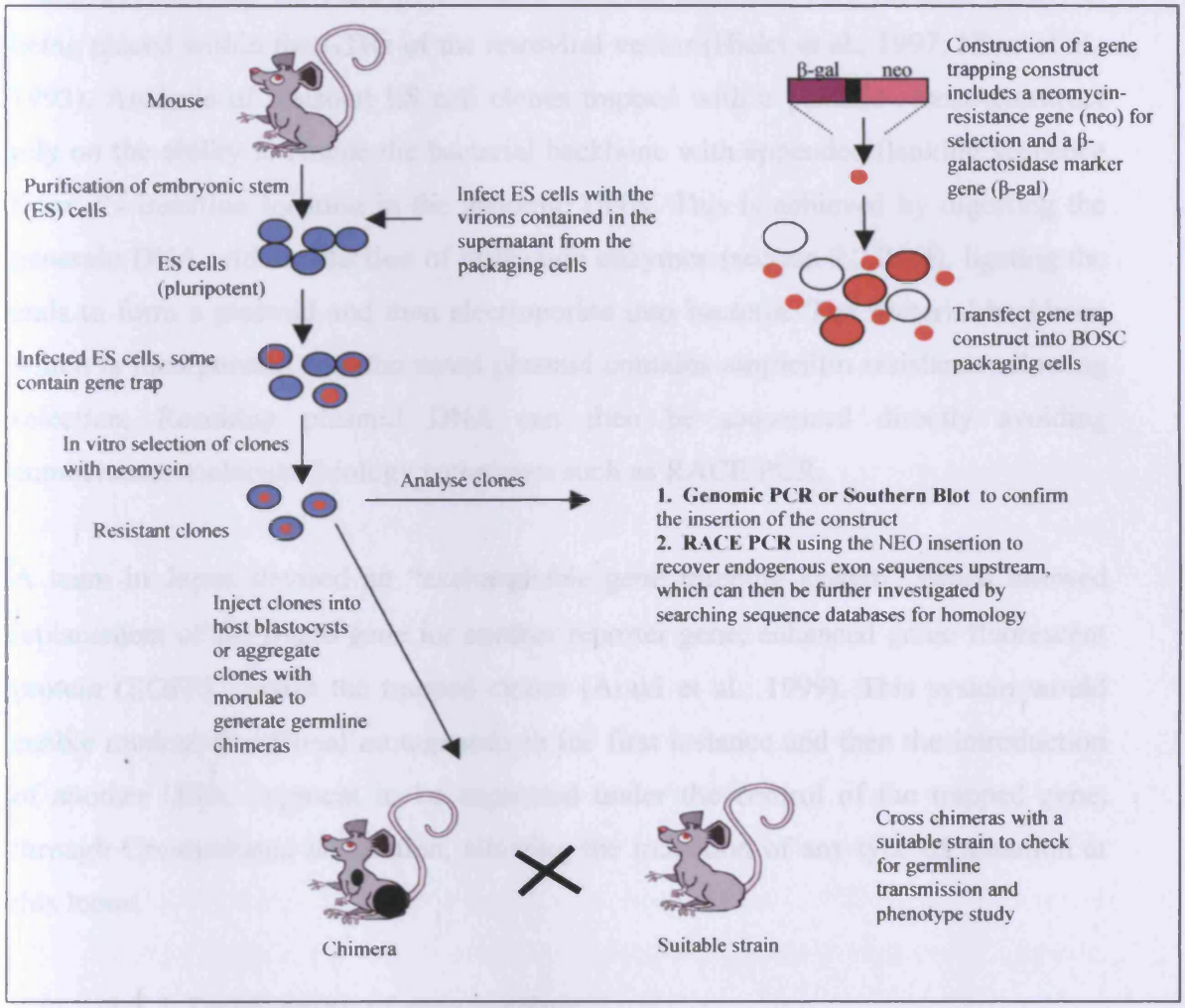
Sometimes, genetic redundancy causes diminished phenotypes, which can hinder the functional analysis of the mutation. At other times, the targeted gene is required for early embryonic viability. The resulting failure of a mutant embryo to survive prevents the function of the gene in later developmental processes from being studied. In addition, the molecular lesion generated does not always have the predicted, and/or an easily interpretable effect. The advent of conditional recombinase systems, which allow gene expression to be controlled in a temporally and spatially manner, has helped to overcome this limitation (Dymecki, 1996; Lakso et al., 1992; Orban et al., 1992).

Although being a very time consuming method of mutagenesis, having to design and generate the targeting constructs, isolating clones, breeding and then analysing the phenotype (Bronson and Smithies, 1994; Mills and Bradley, 2001; Mitchell et al., 2001; Soriano, 1995), gene targeting has led to unprecedented insights into gene function. However, the null mutations that are generally created by investigators using homologous recombination strategies often do not resemble the types of molecular lesions found in disease. So, despite the success of gene targeting, the interest in random mutagenesis screens intensified.

1.7. Gene Trapping

Gene trapping was designed to identify novel, developmentally regulated genes in mouse embryos and potentially offers a powerful tool to generate numerous insertional mutations in mice and to perform phenotypic screens for developmental mutants (fig 1.4.). The trapping principle is a form of insertional mutagenesis specifically designed to disrupt gene function by producing intragenic integration events.

Figure 1.4. Schematic outline of gene trapping strategy



The basic strategy of gene trapping (Friedrich and Soriano, 1991; Gossler et al., 1989; Hicks et al., 1997; von Melchner et al., 1992; Wiles et al., 2000; Zambrowicz et al., 1998) is to introduce into ES cells some marker transgenes that will only function if they have integrated into an active gene. Such an insertion usually results in the disruption of this gene. The disrupted gene is also tagged by the integration, so it can be readily identified by sequence and database analysis.

Gene trapping methodology is a powerful strategy for cloning and identifying functional genes, as it marks a gene with a tag and simultaneously generates a corresponding genetic variation for that particular locus. Therefore, gene trapping is an extremely useful tool for functional genomics, establishing correlation between the physical and genetic maps of the genome.

1.7.1. Plasmid Rescue

Plasmid rescue (fig. 3.9.) is a style of gene trapping that relies on a bacterial backbone being placed within the LTRs of the retroviral vector (Hicks et al., 1997; Niwa et al., 1993). Analysis of resistant ES cell clones trapped with a plasmid rescue construct rely on the ability to rescue the bacterial backbone with appended flanking sequence from it's insertion location in the genomic DNA. This is achieved by digesting the genomic DNA with a selection of restriction enzymes (section 2.2.2.xii), ligating the ends to form a plasmid and then electroporate into bacteria. The bacterial backbone which is incorporated into the novel plasmid contains ampicillin resistance allowing selection. Resulting plasmid DNA can then be sequenced directly avoiding cumbersome molecular biology techniques such as RACE PCR.

A team in Japan devised an "exchangeable gene trapping system" which allowed replacement of the β-geo gene for another reporter gene, enhanced green fluorescent protein (EGFP), within the trapped clones (Araki et al., 1999). This system would enable random insertional mutagenesis in the first instance and then the introduction of another DNA fragment to be expressed under the control of the trapped gene, through Cre-mediated integration, allowing the induction of any type of mutation at this locus.

1.7.2. Development of gene trapping:

Each functional genomics approach has its strengths and weaknesses and only by taking advantage of each one will the functions of the mammalian genome be understood. Gene trapping takes the middle path between random and molecularly defined mutations. Gene trap vectors have evolved from enhancer-trap vectors, a molecular tool used to identify and characterise mammalian enhancer sequences from cell lines (Weber et al., 1984). This technique was quickly modified in both *Drosophila* and the mouse to generate vectors that adopt the transcriptional regulatory apparatus of endogenous genes to report the endogenous expression of a gene that flanks the vector insertion site (Allen et al., 1988; Kothary et al., 1988; O'Kane and Gehring, 1987).

O'Kane and co-workers developed an approach for the *in situ* detection of genomic elements that regulate transcription in *Drosophila melanogaster* (O'Kane and Gehring, 1987). The random generation of operon fusions in prokaryotes involves integrating a promoterless reporter gene, whose expression can easily be detected and assayed at many different positions in a target genome. This is possible due to the reporter gene coming under the control of a random selection of chromosomal promoters and inspired a similar approach in *Drosophila*. This approach would enable the isolation and characterisation of genes, simply by knowing or postulating their pattern of expression. It was realised that in eukaryotes however, that for the efficient expression of a reporter gene, not only would an active promoter be required, but the reporter gene would normally have to be the first gene in the fusion transcript. They decided to express the reporter gene, *lacZ* gene of *Escherichia coli*, from a weak promoter and to integrate this promoter fusion by P-element transformation into the *Drosophila* genome. An increase in expression of the reporter gene was expected to reflect the activity of nearby enhancer like elements in the genome that can act at a distance on the weak promoter. Using P-element-based vectors, thousands of insertions have been generated to monitor chromosomal loci active during embryogenesis. The *lacZ* reporter provides a sensitive and easily assayable gene product to detect expression in whole embryos. Approximately 65% of the lines tested were found to express the *lacZ* reporter in a restricted pattern during embryogenesis. Furthermore, about 15% of the P-element insertions caused recessive mutations that resulted in visible phenotypes (O'Kane and Gehring, 1987).

Gossler and co-workers devised a strategy for identifying regions of the mouse genome that are transcriptionally active in a temporally and spatially restricted manner during development (Gossler et al., 1989). They based their experiments on previous work in which the integration of the *lacZ* gene linked to a weak promoter was shown to detect *cis*-acting elements in the mouse genome that activated β-galactosidase expression in the developing spinal cord and caused a neurological mutation (Kothary et al., 1988). Also, similar developmental activation of *lacZ* constructs was reported in *Drosophila* (O'Kane and Gehring, 1987), as mentioned above and other transgenic mouse strains (Allen et al., 1988).

Gossler and co-workers used two types of construct, one being an enhancer trap construct, which relies on *cis*-acting regulatory elements close to the site of integration to activate a weak promoter and initiate expression of the *lacZ* gene (Gossler et al., 1989). It was expected that some integration events would cause a mutation in an endogenous gene. The ability to prescreen ES cells for desired insertion events, made possible the design of a second type of vector called the gene trap. The essential feature of the "gene trap" design is the placement of a splice acceptor in front of a promoterless *lacZ* gene. Integrations of the vector into the intron of a gene in the correct orientation were predicted to create *lacZ* fusion transcripts and if the reading frame of the endogenous genes and *lacZ* are the same, an active β-galactosidase fusion protein should be produced. Insertion of the gene trap vector was expected to produce a mutation in the host genome in all cases. Vectors similar to the original bacterial vectors, termed promoter traps, have also been used (von Melchner et al., 1992). As they contain only the coding sequences of the reporter gene, they are expected to require insertions into exons to activate reporter gene expression.

Expression patterns provided clues to the function of developmentally regulated genes but final proof only came from analysing mutant phenotypes. In *Drosophila*, the majority of enhancer trap insertions tested expressed β-galactosidase in a pattern similar to the adjacent endogenous gene (Bellen et al., 1989). The requirement for gene/promoter trap insertions to occur within transcription units, predicts an accurate portrayal of endogenous gene expression.

In 1991 Glenn Friedrich and Phillippe Soriano (Friedrich and Soriano, 1991) first developed the ROSA strategy (introduced in section 1.5.3.). They developed a general strategy for selecting insertion mutations in mice. Constructs lacking a promoter and including a β-gal gene, or a reporter gene encoding a protein with both β-gal and neomycin phosphotransferase activity were designed so that activation of the reporter gene depended on its insertion within an active transcription unit. Insertion events created a mutation in the tagged gene and allowed its expression to be followed by β-gal activity. Results obtained from this promoter trap experiment suggested that a substantial proportion of mammalian genes identified by this approach are not essential for development.

Skarnes and colleagues demonstrated the capacity of a gene trap vector containing a splice acceptor site to generate *lacZ* fusion transcripts, accurately report endogenous gene expression and to mutate the endogenous gene at the site of integration. It was confirmed that gene trap vectors are able to create spliced fusion transcripts with endogenous genes and prevent the synthesis of normal transcripts at the site of integration (Skarnes et al., 1992).

New trap vectors were developed to trap genes in murine ES cells (Niwa et al., 1993). The polyA addition signal of the *neo* gene was removed so that they needed to trap an endogenous polyA signal for expression of the *neo* gene. It was found that the frequency of gene-trap events of this type of vector was about five times higher than with the vector containing the polyA signal and only one copy of the trap vector was integrated in most cases (Niwa et al., 1993). By employing a polyA trap, intragenic vector integration is selected for. A mRNA transcribed from a selectable marker gene lacking a polyA signal in a gene trap vector is stabilised only when the gene trap vector captures a cellular polyA signal. Since polyA trapping occurs independently of the expression of the target genes, any gene could potentially be identified at almost equal probability regardless of the relative abundance of its transcripts in target cells (Ishida and Leder, 1999). Other gene-trap strategies enrich for intragenic integration events from the beginning, like the promoter/enhancer trap in which the expression of a promoterless selectable marker is dependent on the capture of an active transcriptional promoter/enhancer in the target cells. While this is a very reliable way to focus exclusively on functional genes, transcriptionally silent genes in the target cells are missed by this strategy.

The trap vector both serves as a sequence tag for the identification of the mutated locus and can provide a reporter-selection marker for monitoring endogenous gene activity. The presence of splice donor (SD) or splice acceptor (SA) elements leads to the generation of a fusion protein, even in the case of an intron integration.

A large number of gene trap insertions can be isolated in mouse ES cells and are immediately accessible in molecular characterisation using high throughput methods to generate sequence tags for each mutated gene. New mutations can be generated in mice at a rate far exceeding that of conventional gene targeting and also recessive

lethal phenotypes can be identified at all developmental stages and can be easily maintained. In addition, various criteria can be used to pre-select genes of interest and their functional analysis can proceed rapidly based on a confluence of sequence, mapping, expression and phenotype information (Mitchell et al., 2001; Soriano, 1995).

The gene trap approach has had a number of successes in establishing mutant mice strains. One example is the "*jumonji*" mouse displaying abnormal groove formation on the neural plate and a defect in neural tube closure (Takeuchi et al., 1995). The gene trap construct contained a splicing acceptor sequence at the 5′ terminus of the neomycin phosphotransferase gene, to enhance trapping efficiency. An internal ribosome entry site was inserted between neo and *lacZ* to enable the dicistronic expression of two genes. The vector contained no promoter sequences so only trapped cells would be able to survive selection with G418. The sequence for nuclear transporting peptide derived from SV40 large T gene was placed at the 5′ terminus of the *lacZ* sequence. This sequence is expected to concentrate *lacZ* products in the nucleus and facilitate detection of cells expressing trapped genes (Kimura et al., 1994).

Another example used the ROSA βgeo vector to disrupt the α-enolase gene by insertion of the splice trap vector into the first intron (Couldrey et al., 1998). This mutation resulted in an early recessive embryonic lethality. Mice heterozygous for this mutation showed no obvious phenotype.

New vectors are being designed to trap specific classes of protein. For example, Skarnes and colleagues used protein sorting and the fact that β-gal activity is abolished in the endoplasmic reticulum to design a vector that specifically traps genes that encode secreted and transmembrane proteins and are expressed in ES cells (Skarnes et al., 1995). By encompassing a variety of attributes in gene trapping constructs it will be possible to achieve saturation mutagenesis of the genome.

In this gene trapping project a selection of retrovirus based vectors are being used that contain a variety of sequence information to aid integration and subsequent

characterisation. Compared to plasmid vectors, retroviral vectors are advantageous, primarily because integration leads to a short direct duplication of host sequences at the site of the single proviral insert with the ends of the vector intact and does not result in other rearrangements in the host genome. Plasmid-based vectors are susceptible to nucleases during electroporation and recombination into the genome. Deletions from the ends of the vector reduce the efficiency of cloning and make primer selection for subsequent analysis more difficult. Also, none of the vectors utilised in this study contain a promoter driven resistance cassette. The promoter could disturb neighbouring genes, which would make interpretation of the mutant phenotype even more difficult (Jaenisch, 1988).

1.7.3. Insertion location of retroviral vectors

Retroviral DNA integrates with high efficiency into the genome of the host cell, where it persists as a stable genetic element, the provirus. It must be considered, however, that target site preferences for retroviral integration may impair the ability to sample the genome effectively (Neilan and Barsh, 1999).

Chowdury and colleagues demonstrated in a large-scale gene trapping experiment using 5'RACE PCR, that gene trap vectors integrate stochastically into the genome (Chowdhury et al., 1997). Results showed that there was no preference for insertion site into individual mouse genes when using the SAβGeo or the IRESβGeo vectors. Insertions were found either at the 5'-end, at the 3'-end or in the middle of a gene. It would be expected that most insertions at the 5'-end of a trapped gene would completely knockout the corresponding gene. In other cases, interesting phenotypes may be obtained representing partial inactivation of the gene products or as a result of a dominant gain of function. A consideration with 3' insertions would be a resulting large fusion complex between the resistance marker and the trapped gene, leading to the inactivation of the resistance marker due to inadequate protein folding.

Other evidence indicates that retroviral integration is not entirely random but occurs preferentially into regions close to deoxyribonuclease I-hypersensitive sites, which are characteristic for active genes. It is possible therefore that the chromatin conformation of a given gene can influence its chance of being mutated. An expressed gene with an

opened chromatin conformation may represent a more likely target for integration and insertional mutagenesis than inactive heterochromatic genes (Craigie, 1992; Rohdewohld et al., 1987; Scherdin et al., 1990; Vijaya et al., 1986).

Areas of high chromatin enzymatic accessibility, such as DNase I hypersensitive sites are often associated with enhancers or locus control regions. A preference for 5′ integration events would generate preferable null mutations as opposed to 3′ integrations which only partially disrupt the function of the trapped gene (Floss and Wurst, 2002). Other useful features for 5′ insertion of the retroviral vector are, interpretation of mutant phenotypes are less likely to be complicated by any dominated negative effects of truncated protein products of the affected locus and the potential for production of truncated proteins retaining sufficient normal activity to rescue the phenotype is reduced (Thomas et al., 2000).

Individual vector designs have their own inherent biases that will influence the range of target genes isolated. Gene trapping vectors are designed to detect insertions within transcribed regions of the genome. A variety of vectors have been developed that differ in their individual requirements for reporter gene activation. "Promoter trap" vectors simply consist of a promoterless reporter gene that is activated after insertions into exons (Friedrich and Soriano, 1991). In contrast, "gene trap" vectors contain a splice acceptor sequence upstream of a reporter and are activated after insertion into introns (Komada et al., 2000; Skarnes et al., 1992; Takeuchi et al., 1995). Gene trap vectors are more efficient than promoter trap vectors, but will favour the detection of genes composed of large intronic regions and are likely to miss genes possessing few or no introns.

Gene trap vectors are also constrained by the reading frame imposed by the splice acceptor sequence. Non-incorporation of an ATG site into the gene trap vector means that the activation of the reporter gene depends on the production of an in-frame fusion to the N terminus of the target gene. However, to overcome this problem some gene trap vectors have an internal ribosome entry site (IRES) inserted 5′ of the reporter gene, to initiate translation independently of the upstream open reading frame (Shirai et al., 1996; Takeuchi et al., 1995).

1.7.4 Use of polybrene in gene trapping

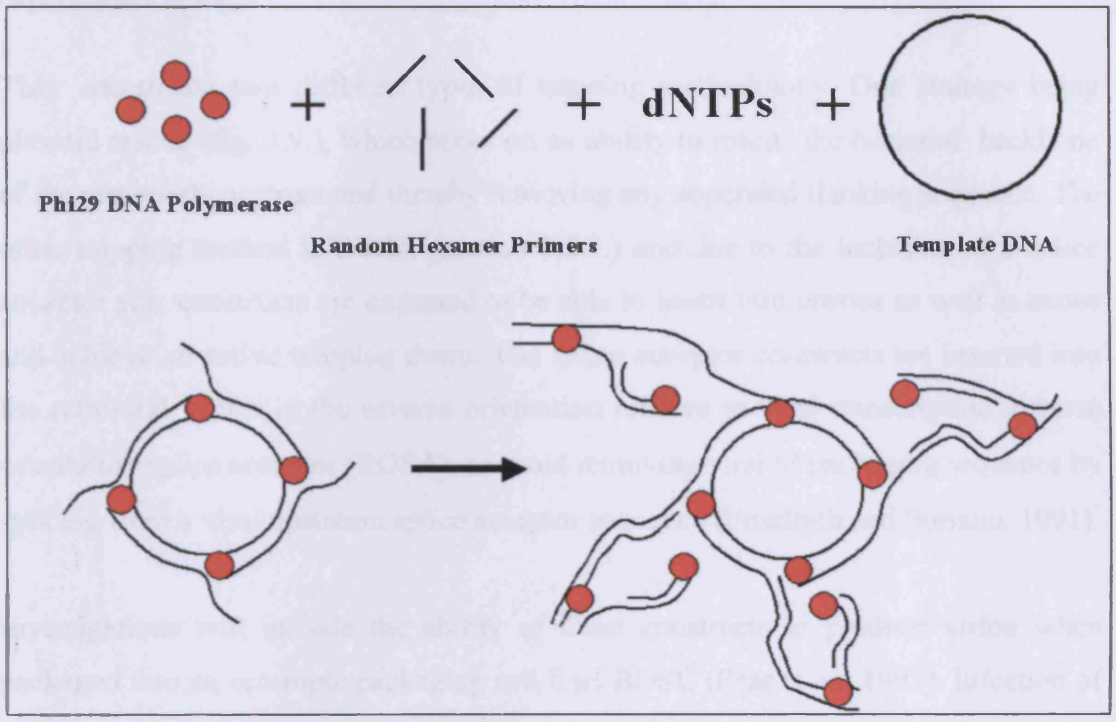
Retroviruses have short half-lives and therefore can only travel a limited distance by Brownian motion in infection medium before deactivating. With virus particles able to diffuse only a few microns before losing bioactivity, a large proportion of the initially active virus particles are inactive before they can interact with a target cell, fundamentally limiting achievable levels of transduction efficiency (Chuck et al., 1996). To circumvent these constraints, mechanical approaches, including centrifugation (Kotani et al., 1994) and flow-through transduction (Chuck and Palsson, 1996) have been developed to increase the likelihood and frequency of active virus-cell interactions by adding a convective component to the mass transport of virus. While these strategies do yield enhanced transduction efficiencies, their expense and difficult scale-up limit their usefulness for large-scale experiments. The classical approach to enhancing retrovirus transduction efficiency has been to add cationic polymers during the transduction process (Toyoshima and Vogt, 1969).

As previously stated retroviral virions are enveloped with a lipid coat from the producer cell line and therefore have a similar negative charge as the target cell. To overcome electrostatic repulsion between virion and cell, positively charged polymers such as polybrene are often used during transduction. Polybrene is a small, positively charged molecule that binds to cell surfaces and neutralises surface charge. This apparently allows the viral glycoproteins to bind more efficiently to their receptors as it reduces the repulsion between sialic acid containing molecules (Davis et al., 2002).

1.8. Rolling circle amplification (RCA)

It was felt necessary to introduce rolling circle amplification at this juncture, as the technique was employed during plasmid rescue (see section 3.11.2. for results). Due to the tiny amount of plasmid DNA expected to be rescued it was felt necessary to try to amplify, prior to electroporation. To attempt this, the TempliPhi™ DNA amplification kit (Amersham Biosciences, UK Ltd., Bucks, UK) was obtained.

Figure 1.5. Schematic representation of rolling circle amplification



The TempliPhi method (fig. 1.5.) utilises bacteriophage ϕ29 DNA polymerase enzyme to exponentially amplify single- or double-stranded circular DNA templates by rolling circle amplification (Blanco et al., 1989). Random hexamer primers anneal to the circular template DNA at multiple sites. Phi29 DNA polymerase (Blanco and Salas, 1984) extends each of these primers. When the DNA polymerase reaches a downstream extended primer, strand displacement synthesis occurs. The displaced strand is rendered single-stranded and available to be primed by more hexamer primer. The process continues, resulting in exponential, isothermal amplification. The product of the TempliPhi reaction is high molecular weight, double stranded concatemers of the circular template.

1.9. AIMS

Gene trapping can be regarded as a more efficient method of insertional mutagenesis when compared to gene targeting, as having to design and generate targeting constructs, isolating clones and breeding is very time consuming. This study looks at

four gene trapping retroviral constructs that have never been used in gene trapping experiments before.

They encompass two different types of trapping methodology. One strategy being plasmid rescue (fig. 3.9.), which relies on an ability to rescue the bacterial backbone of the retroviral construct and thereby retrieving any appended flanking sequence. The other trapping method is ROSA (section 1.5.3.) and due to the inclusion of a splice acceptor site, constructs are expected to be able to insert into introns as well as exons and achieve an active trapping event. The splice acceptor constructs are inserted into the retroviral vector in the reverse orientation relative to viral transcription, reverse orientation splice acceptor (ROSA), to avoid removing viral Ψ packaging sequence by splicing from a viral upstream splice acceptor sequence (Friedrich and Soriano, 1991).

Investigations will include the ability of these constructs to produce virion when packaged into an ecotropic packaging cell line, BOSC (Pear et al., 1993). Infection of murine pluripotent embryonic stem (ES) cells with the resulting virions. To assist in the gene trapping and subsequent characterisation these constructs contain a selection of sequence information. The splice acceptor constructs do not just contain a splice acceptor site they also contain a slice donor site, allowing 3′ endogenous sequence to remain after splicing events and subsequently analysed. One splice acceptor construct contains a T7 promoter and an Isce-1 restriction site, the other splice acceptor construct contains an internal ribosome entry site and a *lox* P stuffer region, the theory behind these additions is explained when introducing the constructs (section 3.5.).

A comparison can be made between the number of $G418^R$ ES cell lines resulting after infection from each construct. These ES cell lines will be manipulated in an appropriate manner, relating to which construct they originated from and sequence data obtained. Resulting sequence data can be compared with that held on sequence databases to ascertain the insertion location of the retroviral construct. Successful trapping events can then be examined further, ES cell lines can be injected into host blastocysts to obtain germline transmission of the mutation.

It is hoped that a performance assessment will be achieved for each of the new constructs being studied.

CHAPTER 2

MATERIALS AND METHODS

2. MATERIALS AND METHODS

For all procedures, solutions were made to the standard required for molecular biology using molecular biology grade and/or "tissue-culture-tested" reagents. All solutions were made using sterile double-distilled water and where appropriate autoclaved or filter sterilised.

2.1. TISSUE CULTURE

All tissue culture plates and plasticware was supplied by Nunc, Fisher Scientific Equipment UK Ltd., Loughborough, UK.

2.1.1. Medium Preparations:

2.1.1.i. Antibiotics:

Penicillin/streptomycin (5000 units/ml penicillin G sodium and 5000μg/ml streptomycin sulphate in 0.85% saline from Gibco BRL, Paisley, UK) used at 1X concentration.

Broad spectrum antibiotic, Gentamicin (Gibco BRL, Paisley, UK) used at a concentration of 50μg/ml.

2.1.1.ii For Maintaining Pluripotent Embryonic Stem (ES) Cells:

500ml Modified Eagles Medium (MEM) Alpha medium without ribonucleosides and deoxyribonucleosides (Gibco BRL, Paisley, UK)

5ml β Mercaptoethanol 10mM (Gibco BRL, Paisley, UK)

5ml Glutamine 200mM (unstable amino acid) (Gibco BRL, Paisley, UK)

50ml Newborn Calf Serum (NCS) (PAA, Yeovil, UK)

50ml Foetal Calf Serum (FCS) (PAA, Yeovil, UK)

50units/ml medium - ESGRO® murine Leukaemia Inhibitory Factor (mLIF) (Chemicon Europe Ltd. Hampshire, UK)

2.1.1.iii. For Maintaining STO (Fibroblast) Cells:

500ml MEM Alpha medium without ribonucleosides and deoxyribonucleosides (Gibco BRL, Paisley, UK)

50ml NCS (PAA, Yeovil, UK)

5ml Glutamine 200mM (Gibco BRL, Paisley, UK)

2.1.1.iv. For Maintaining BOSC Packaging Cells:

500ml MEM Alpha medium without ribonucleosides and deoxyribonucleosides (Gibco BRL, Paisley, UK)

50ml FCS (PAA, Yeovil, UK)

5ml Glutamine 200mM (Gibco BRL, Paisley, UK)

2.1.1.v. Freezing Mix:

10% DMSO (Gibco BRL, Paisley, UK)

90% FCS (PAA, Yeovil, UK)

2.1.2. Preparation of feeder layers:

Using STO fibroblast cells (Robertson, 1987; Ware and Axelrad, 1972):

It is important that STO cells are passaged promptly when they reach confluency to prevent the accumulation of non-contact inhibited cells in the population.

Grew up a confluent 10cm plate of STO fibroblast cells.

Aspirated the medium and replaced with 6ml of STO medium + 10μg/ml mitomycin C (Sigma-Aldrich, Poole, Dorset, UK).

Incubated at 37°C for 2-3 hours, avoiding longer exposure.

Removed mitomycin C medium and washed 3 times with PBS (phosphate buffered saline) (Gibco BRL, Paisley, UK).

Trypsinised the cells with 2ml of trypsin-EDTA (Gibco BRL, Paisley, UK)
per plate for about 2min and then collected the cells in STO medium.

Pelleted the cells by centrifugation at 1500g for 5min and resuspended in STO medium, counted the suspension using a hemocytometer.

Prepared a dilution of 10^6 feeder cells/ml of medium.

Seeded a plate which had been coated in gelatin (0.1% gelatin in water) (Stem Cell Technologies Inc. Vancouver, Canada) $12x10^3$ cells/cm^2.

The 10^6/ml feeder suspension can be stored at 4°C for up to 7 days and seeded as required.

2.1.3. Passaging and Maintenance of ES cells:

Aspirated growth medium and washed the ES cells with PBS. Aspirated the PBS. Covered cells with cold trypsin /EDTA for a few seconds and then aspirated enough trypsin /EDTA just to leave a film covering the cells. Monitored the cells under a microscope to see when they had detached. Quickly added pre-warmed growth medium to inactivate the trypsin and resuspended the cells. Pipetted up and down a few times to generate a single cell suspension. Transferred the resuspended cells into a centrifuge tube and pellet at 1500g for 5min. Aspirated off the medium and resuspended the pellet in 10ml of pre-warmed growth medium. Counted the cells using a hemocytometer and calculated the number of cells per ml. Aliquoted the appropriate number of cells into pre-gelatinised plates for passage / experimentation. Checked the cells are attaching about 30-60 mins after replating.

It is important not to allow the cultures to become too crowded or too sparse during routine culture. Either situation sets up conditions in which selection for abnormal (fast growing, aneuploid) cells can occur. When in good health ES cells grow with a population doubling time of about 18-20 hours. For routine maintenance change the medium on day 2 and subculture on day 3.

$2.5x10^4$ ES cells/cm^2 are replated so confluency achieved in 2-3 days.

2.1.4. Transfection of Packaging Cells:

2.1.4.i. Using Lipofectamine (Gibco BRL, Paisley,UK):

Day 1: Seeded a 100mm plate with $19x10^5$ BOSC cells.

Day 2: Prepared dilutions and solutions using Optimem (Gibco BRL, Paisley, UK). Solution A; 640µl Optimem, 9.6µl DNA being transfected (0.5µg /µl). Solution B; 64µl Lipofectamine, 640µl Optimem. Added solutions A and B together. Allowed to stand at room temperature for 20 minutes. Rinsed plate with 10ml Optimem. Added 5.12ml Optimem to transfection mix. Applied the 6.4ml transfection mix to the plate.

After 4 – 6 hours added 6.4ml of normal growth medium not containing antibacterial agents to the plate.

Day 3: Started retroviral supernatant production. 24 hours after the transfection mix was applied changed medium for 10ml of normal growth medium.

Day 4: Harvested 48 hour virion supernatant.

At 48 hours after transfection removed 5ml of medium and centrifuged at 1500g for 5 min. Sterilised the supernatant through a 0.45μm Nucleopore™ filter (Millipore UK Ltd., Herts, UK). Aliquoted into 1ml vials and froze at –70°C but PREFERABLY used fresh by adding directly onto preseeded ES cells.

Day 5: Harvest 72 hour virion supernatant.

At 72 hours after transfection removed 5ml of medium and treated the same as for the 48 hour supernatant.

2.1.4.ii. Using Calcium Phosphate:

2X HEPES-buffered saline (HBS), (280mM NaCl, 10mM KCl, 1.5mM $Na_2HPO_4 \cdot 2H_2O$, 12mM dextrose, 50mM HEPES).

Adjusted the pH to 7.05 with 0.5 N NaOH and adjusted the volume to 100ml with distilled H_2O. Sterilised the solution by passing it through a 0.22μm Nucleopore™ filter (Millipore UK Ltd., Herts, UK). Stored in 5ml aliquots at –20°C until required.

2M $CaCl_2$. Sterilised the solution by passing it through a 0.22μm Nucleopore™ filter (Millipore UK Ltd., Herts, UK). Stored in 1ml aliquots at –20°C until required.

0.1 X TE (pH 8.0), (1mM Tris · Cl (pH 8.0), 0.1mM EDTA (pH 8.0)). Sterilised the solution by passing it through a 0.22μm Nucleopore™ filter (Millipore UK Ltd., Herts, UK) and stored at 4°C until required.

Prepared DNA (~20μg/10^6 cells) in 0.1 X TE (pH8.0) at a concentration of 40μg/ml.

Day 1: Seeded a 60mm plate with 1 x 10^6 BOSC cells.

Day 2: Made up DNA mix, (5μg pHIT 60, 5μg pHIT 123, 10μg Construct).

For each 60mm plate to be transfected the calcium phosphate-DNA coprecipitate was prepared in a 2ml sterile eppendorf, (220μl DNA mix, 250μl 2XHBS). Gradually added 31μl 2M $CaCl_2$ with gentle agitation.

Incubated the mixture at room temperature for 20-30 mins. Removed the medium from the BOSC cells and replaced with fresh non-antibiotic BOSC medium (2.1.1.iv.). To 4.5ml of medium added 0.5ml of calcium phosphate - DNA suspension.

Day 3: Approximately 20 hours later removed the transfection mix and replaced with normal BOSC growth medium.

Day 4: Harvested virion supernatant.

At 48 hours after transfection remove the medium and centrifuge at 1500g for 5 minutes. Sterilised the supernatant through a 0.45μm Nucleopore™ filter (Millipore UK Ltd., Herts, UK).

Aliquoted into 1ml vials and froze at –70°C but PREFERABLY used fresh by adding directly onto preseeded ES cells.

Day 5: Harvest virion supernatant.

At 72 hours after transfection removed the medium and treated the same as for the 48 hour supernatant.

2.1.5. Retroviral Infection and G418 Selection of ES Cells:

Day 1: Seeded 60mm plates with $5x10^5$ ES cells/plate. It is important to also set up control plates of the same ES cells, which will not undergo retroviral infection. This allows a more accurate judgement of completed selection.

Day 2: Began infection with the virion produced during retroviral packaging. Made up a master mix; Virion supernatant (volume retrieved during packaging), Polybrene 10μg/μl (40μl/p60 plate being infected), ESGRO® mLIF 10^6units (50units/ml), Growth Medium (make up final volume to 5ml/plate).

Aliquoted 5ml of the master mix to each of the test plates to be infected with the construct. On the control plates added exactly the same except for the virion supernatant, making up the final volume to 5ml/plate with normal ES growth medium (2.1.1.ii.).

Day 3: 20 hours after the beginning of infection replaced the infecting medium with normal ES growth medium.

Day 4: Split the infected ES cells as necessary using normal ES growth medium and seeded onto STO feeders (2.1.2.).

Day 5: Began selection of transfected colonies by introducing ES growth medium with G418 (Gibco BRL, Paisley, UK) included at a concentration previously established with the ES cells being used. The batch of G418 utilised in this project is used at a concentration of 200μg/ml.

Day 6: Changed the G418 selection medium to clear away the initial abundance of cell debris due to the start of selection.

From this stage changed the G418 selection medium whenever deemed necessary.
When all the colonies on the control plate had died it was reasonable to assume that all remaining colonies on the test plates contained a functional neomycin resistance gene. At this stage the remaining colonies were picked.

2.1.6. Replica plating of clones directly on picking:

Prepared a sterile round bottomed 96 multiwell plate with aliquots of 35µl trypsin/EDTA. Gelatin coated three flat bottomed, replica, 96 multiwell plates. Added 10^4 feeders/well with 100µl of G418 selection medium.
Aspirated the medium from the plate containing the clones, washed with PBS and then covered the clones with a further 10ml of PBS.
Using a dissecting microscope with a 200µl pipette set at 5µl, sucked up the clones and ejected them one-by-one into the wells of the round bottom multiwell dish, using a lid to prevent evaporation of the trypsin/EDTA.
Using an 8-way pipette resuspended the clones by pipetting up and down with 200µl of G418 selection medium. Transferred 70µl to each of the three replicas, pre-gelatinised multiwell plates.
Incubated in 37°C, 5% CO_2 incubator.
Changed medium the next day and when a healthy culture was established (3-4 days), froze one or two of the plates. Passaged the remaining plate or used it to prepare DNA or RNA samples for analysis.

2.1.7. Culturing of G418 resistant ES cell clones:

The clones needed to be grown up for further investigation. One vial of each clone was frozen away in liquid nitrogen with feeders and freezing mix in case of any mishap or future requirement.
When the clones had become confluent in the 96 well plate they were moved individually to 6 well plates with feeders, still using G418 selection medium. All of the clones grew at different rates so they needed to be examined individually on a daily basis.

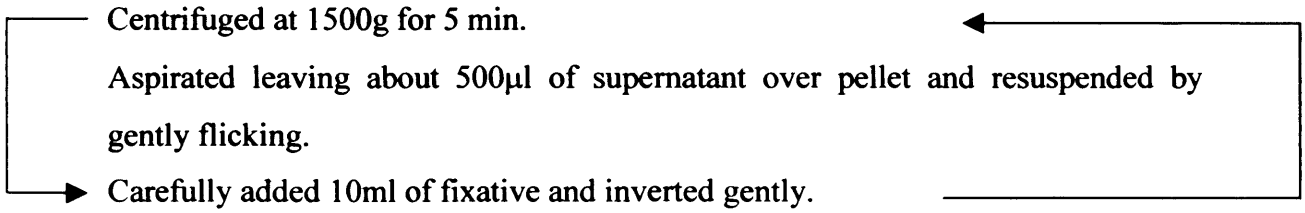
Eventually when three confluent p60 plates of each clone were established they were then harvested and prepped for DNA/RNA or stained for *LacZ*, depending on the type of clone. It must be remembered that when these ES cells are prepped they also contain a number of feeder cells so there will be a certain amount of feeder genetic material in the resultant DNA/RNA preps. It was sometimes necessary to try and remove the ES cells from the feeder layer prior to preparation, as stated in the text. On these occasions the cells were passaged a number of times until no feeder cells were visible.

2.1.8. Embryonic Stem (ES) Cell Metaphase Spreads:

Seeded cells in a 100mm dish overnight to be about 50-80% confluent the next day, ~$1x10^6$ ES cells.

Next day: Added Colcemid (Karyomax 10μg/ml, Gibco BRL, Paisley, UK) to final concentration of 0.1μg/ml and incubated for 2-3 hours at 37°C. (Colcemid inhibits spindle formation and arrests dividing cells at mitotic metaphase). After incubation harvested cells as for normal passage, centrifuged at 1500g for 5 min. Aspirated medium. Resuspended in 5ml of PBS (37°C) gently, but thoroughly, so as to achieve a single cell suspension.

Centrifuged at 1500g for 5 min. Aspirated most of the PBS (37°C), leaving about 500μl and resuspend by gently flicking the tube. Added 10ml of pre-warmed (37°C) 0.075M KCl, (freshly made from 1M stock solution), and left at room temperature for 15 minutes. (This time needs to be determined for different cell lines. The hypotonic KCl swells the cells – if cells are excessively swollen chromosomes will scatter, if insufficiently swollen they will not separate.) For the black six cell line (SHBl6.3) 15min was found to be appropriate. Centrifuged at 1500g for 5min. Aspirated leaving about 500μl and resuspended by gently flicking the tube. With a P1000 Gilson pipette slowly added 2ml of fixative, dropwise. Incubated for 40min to allow fix to penetrate.



Repeated the above steps twice. This removed cell debris and proteinaceous material. Removed all supernatant very carefully. Resuspended in less than 1ml of fixative (Methanol : Acetic Acid, 3 : 1). Dropped onto pre-cleaned microscope (soaked in 50% EtOH : 50% Acetone), using a long-form glass Pasteur pipette. The nuclei burst and released the chromosomes. A height of about 20cm is optimal for good spreads with 1 drop/slide. The aim was to achieve discrete groups of well spread non-overlapping chromosomes. Allowed to air dry. The spreads could be examined under phase contrast but for a more distinguishable picture the spreads were stained with Geimsa / PBS 3% v/v for about 30min. Washed three times with PBS and examined under bright field.

2.2. MOLECULAR BIOLOGY

2.2.1.MATERIALS

The sources of specialised reagents are listed in Table 2.1. Reagents were of molecular biology grade, unless otherwise stated.

2.2.1.i. DNA Markers:

The DNA markers used were λ DNA, λ DNA digested with *Hind* III and 1kb Plus DNA Ladder™. The fragment sizes of these markers (bp) are given in Table 2.2.

Table 2.2.1. Biological Reagents.

Reagent	Supplier
TempliPhi™ 100 Amplification Kit, Sephaglas™ BandPrep Kit	Amersham Int. plc, Little Chalfont, Bucks., UK
Tris	BDH Chemicals Ltd., Poole, UK
2mm and 1mm Electrocuvettes	Bio-Rad, Hertfordshire, UK.
Proteinase K	Boehringer-Mannheim Ltd., Lewes, East Sussex
	Clontech UK, Basingstoke, UK
Superfrost Glass slides	Corning Ltd., Bucks, UK
Dynal Oligo (dT) Magnetic Beads	Dynal, Oslo, Norway
Chloroform, EDTA, Ethidium bromide, glycerol, hydrochloric acid, isopropanol, sodium chloride, sodium citrate, sodium hydroxide.	Fisher Scientific UK , Loughborough, UK.
Ladder, Oligonucleotide primers, DH5α Competent Cells, 50x TAE, SuperScrip™ II Reverse Transcriptase, RNase H, 3' RACE System for Rapid Amplification of cDNA Ends, HEPES (N-[2-HydroxyEthyl]-Piperazine-N'-[2-EthaneSulphonic acid]), SOC Medium,	Gibco BRL, Paisley, UK.
Disposable Sterile Universal Tubes	Greiner, Stonehouse, UK
Ethanol	Hayman Ltd., Witham, Essex, UK.
0.22μM and 0.45μM Nucleopore™ filters	Millipore U.K. Ltd., Watford, Hertfordshire, U.K.
M13 sequencing primers, Custom Synthesised Primers	MWG Biotech UK., Milton Keynes, UK.

Table 2.2.1 Biological Reagents. (Continued)

Restriction Enzymes	New England Biolabs Inc. Hertfordshire, UK.
ABI Prism™ Dye Terminator Cycle Sequencing Reaction Ready Kit Version 2.1	Perkin Elmer Applied Biosystems, Warrington, UK
Agarose, dNTPs, loading dye, $MgCl_2$ (25mM), 10X Mg-free buffer, MMLV reverse transcriptase, Rnasin, 5X RT buffer, molecular mass markers (λ and φ), Taq DNA Polymerase, Wizard plus SV Minipreps kit, X-Gal, T4 DNA Ligase, 10X Ligation Buffer,	Promega Ltd., Southampton, UK.
RNeasy® Midi Kit, Plasmid Maxi Kit	Qiagen Ltd., Crawley, Sussex, UK.
Ampicillin, BSA, chloroform, isoamylalcohol, formalin, formamide, HPLC water, LB agar, LB broth, Mineral oil, Sodium acetate, Phenol.	Sigma-Aldritch, Poole, UK.
Electro Ten-Blue™ Electroporation Competent Cells	Stratagene Cloning Systems, La Jolla, CA 92037.
Better Buffer	Web Scientific Ltd., Crewe, UK.

Table 2.2.2 DNA marker fragment sizes (bp) and their respective suppliers.

DNA marker	Fragment sizes (bp)	Supplier
λ DNA	48502	Promega Ltd., Southampton, UK.
λ/HIND III	23130, 9416, 6557, 4361, 2322, 2027, 564, 125	Promega Ltd., Southampton, UK.
1kb Plus DNA Ladder™	12000, 11000, 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1650, 1000, 850, 650,500, 400, 300, 200, 100	Gibco BRL, Paisley, UK.

2.2.1.ii. Medium:

LB-broth (mixture of casein enzymatic digest (10g/l), low-sodium yeast extract (5g/l), NaCl (5g/l), inert tableting aids (2g/l)) and LB-Agar (same as broth except for the addition of bacteriological Agar (12g/l)) were prepared as described by the manufacturer, by the addition of the designated number of pellets to water. Following sterilisation, the medium was left to cool to below 50°C and where required, ampicillin was added at a concentration of 100μg/ml.

SOB medium (20M tryptone, 5M yeast extract, 0.5M NaCl, 1mM $MgCl_2$, 1mM $MgSO_4$) filter sterilised through 0.22μm Nucleopore™ filter (Millipore UK Ltd, Herts, UK).

Added to SOB medium to make SOC medium 20mM glucose solution, filter sterilised through 0.22μm Nucleopore™ filter (Millipore UK Ltd, Herts, UK).

2.2.1.iii. Antibiotics:

Stock solutions of ampicillin (50mg/ml) and kanamycinin (10mg/ml) in sterile dH_2O were passed through a 0.22μm Nucleopore™ filter (Millipore UK Ltd., Herts, UK) and stored at -20°C.

2.2.1.iv. Autoclaving Conditions:

Before any technique was performed using DNA or RNA, all equipment needed to be sterilised by autoclaving at 120°C at 15psi for 20 minutes.

2.2.1.v. Oligonucleotide Primers:

Lyophilised oligonucleotides were resuspended in sterile dH_2O to a workable concentration as specified in the text and stored at -20°C.

Primers were designed using Apple Macintosh software called Amplify (Engels, 1993). Primers selected for optimal Tm (melting temperature) at the given primer length were assessed for possible sites of self-complementarity and complementarity to the second primer (especially 3′ complementarity), mainly to reduce the risk of the formation of primer-dimer products.

Table 2.2.3. Section 2.2. Routinely used primers, sequences and suppliers.

Primer	DNA Sequence (5′→3′)	Supplier
M13 forward	GTTTTCCCAGTCACGAC TT	MWG Biotech UK., Milton Keynes, UK.
M13 reverse	CAGGAAACAGCTATAG ACCATG	MWG Biotech UK., Milton Keynes, UK.
pAGAG5215	CGAGATACAGAGCTAG TT	MWG Biotech UK., Milton Keynes, UK.
pANEO659	CCGAGTACGTGCTCGCT C	MWG Biotech UK., Milton Keynes, UK.
pAGAG4822	CAACGTCTCTTCTTGAC ATCT	MWG Biotech UK., Milton Keynes, UK.
pANEOR887	CAACGCTATGTCCTGAT AGCG	MWG Biotech UK., Milton Keynes, UK.
pANEO260	TGCAATCCGTCTTGTTC A	MWG Biotech UK., Milton Keynes, UK.
Oligo-dT	TTTTTTTTTTTTTTT	Pharmacia Biotech Ltd., St. Albans, UK.

2.2.1.vi. Reagents and Buffers:

All routine laboratory solutions were prepared using dH_2O. Sterilisation was achieved by autoclaving at 120°C at 15psi for 20 minutes where required. Heat sensitive components were passed through 0.22μm Nucleopore™ filter (Millipore UK Ltd., Herts, UK) and added separately following autoclaving.

Table 2.2.4. List of routinely employed solutions, reagents and medium.

Procedure	Solution	Components
Agarose Gel Electrophoresis	Electrophoresis loading dye	50% (v/v) Glycerol, 1% (w/v) Bromophenol blue, 1x TAE (sterilised)
mRNA Purification from Total RNA - Oligo (dT) Magnetic Beads	1X Binding buffer (Dynal)	100mM Tris-HCl, pH 7.5 500mM LiCl 10mM EDTA, pH 8.0 1% LiDS 5mM DTT
	Washing Buffer B (Dynal)	10mM Tris-HCl, pH 7.5 0.15M LiCl 1mM EDTA
RNA preparation using Qiagen rneasy midi kit	Buffer RLN	50mM Tris.Cl, pH 8.0 140mM NaCl 1.5mM $MgCl_2$ 0.5% (v/v) Nonidet® P-40 (1.06g/ml)
	Buffer RLT	Protected recipe
	Buffer RW1	Protected recipe
	Buffer RPE	Protected recipe
Polymerase Chain Reaction	10x Taq Polymerase Buffer (Promega Ltd., Southampton, UK)	50 mM KCl, 10 mM Tris-HCl (pH 9.0 at 25°C), 0.1 % Triton®x-100
Genomic DNA preparation	Proteinase K lysis buffer	10mM Tris-Cl, pH 7.6 10mM EDTA 0.6% SDS

Table 2.2.4. (Continued) List of routinely employed solutions, reagents and medium.

Mini Preparation of DNA using Promega Wizard SV Miniprep DNA Purification System	Cell Resuspension Solution	50mM Tris-HCl (pH 7.5) 10mM EDTA 100μg/ml RNase A
	Cell Lysis Solution	0.2M NaOH 1% SDS
	Neutralisation Solution	4.09M guanidine hydrochloride 0.759M potassium acetate 2.12M glacial acetic acid Final pH is approximately 4.2.
	Column Wash Solution	60mM potassium acetate 8.3mM Tris-HCl (pH 7.5) 0.04mM EDTA (pH 8.0) 60% ethanol
Gel Purification, Sephaglas™ BandPrep Kit (Amersham Int. plc, Bucks, UK)	Sephaglas BP	20% (w/v) Sephaglas BP suspended in distilled water containing 0.15% Kathon® CG/ICP Biocide.
	Gel Solubiliser	Buffered solution containing NaI.
	Wash Buffer	20mM Tris-HCl (pH 8.0), 1mM EDTA and 0.1mM NaCl solution to which 18ml of absolute ethanol must be added before use.
	Elution Buffer	10mM Tris-HCl (pH 8.0), 1mM EDTA.

Table 2.2.4. (Continued) List of routinely employed solutions, reagents and medium.

Maxi Preparation of DNA Using Qiagen Maxi Kit	Buffer P1	50mM Tris.Cl, pH 8.0 10mM EDTA, 100µg/ml Rnase A
	Buffer P2	200mM NaOH, 1% SDS
	Buffer P3	3.0M potassium acetate, pH 5.5
	Buffer QC	1.0M NaCl, 50mM MOPS, pH 7.0, 15% Isopropanol
	Buffer QBT	750mM NaCl, 50mM MOPS pH 7.0, 15% Isopropanol, 0.15% Triton® X-100
	Buffer QF	1.25M NaCl 50mM Tris.Cl, pH 8.5 15% Isopropanol

2.2.2. METHODS

2.2.2.i. Preparing Electrocompetent Bacterial Cells:

Streaked out and grew up the relevant bacteria overnight at 37°C on agar plates. Inoculated 10ml of LB with a single colony and left overnight in shaker incubator at 37°C and 225rpm. Inoculated 400ml of LB with 4ml of the overnight culture. Grew up in shaker incubator at 37°C and 225rpm for about 3 ¼ hours. Spun culture at 4°C for 10 mins, 7000g. Resuspended in 200ml ice cold 1mM HEPES. Spun down at 4°C for 10 mins, 7000g.

Resuspended in 100ml ice cold 1mM HEPES. Spun down at 4°C for 10 mins, 7000g. Resuspended in 20ml ice cold 10% Glycerol 1mM HEPES. Spun down at 4°C for 10 mins, 7000g. Resuspended in 5ml ice cold 10% Glycerol 1mM HEPES. Aliquoted cell suspension in 50µl aliquots on ice and froze immediately at -70°C for storage until required.

2.2.2.ii. Electroporation of electrocompetent bacteria DH5 and HB101:

Thawed a 50µl vial of electrocompetent bacteria on ice. Added 1µl of plasmid DNA (10pg/µl) to be introduced into the bacteria and then transferred this mixture to a pre-chilled 2mm cuvette and kept on ice. Set the Gene Pulser (Bio-Rad, Hertfordshire, UK) to 25µF and 2.5kV with the pulse controller set to 200Ω. Dried the cuvette with tissue and placed into the electroporation chamber. Pulsed once. A pulse with a time constant of 4.5 to 5 msec was obtained.

Added 300µl of LB to the cuvette and then using a sterile pasteur pipette transferred 100µl from the cuvette to each of three agar plates with the relevant antibiotic selection. Spread out using a sterile spreader, allowed to dry, inverted and incubated overnight at 37°C.

2.2.2.iii. Electroporation of ElectroTen Blue Electrocompetent Cells (Stratagene, La Jolla, CA.):

Thawed electrocompetent bacteria on ice. Added 1µl of plasmid DNA (10pg/µl)/40µl of bacteria. Transferred this mixture to a pre-chilled 1mm cuvette and kept on ice. Set

the Bio-Rad Gene Pulser machine to capacitance 25μF, applied Volts 1700V and resistance 600Ω. Dried the cuvette with tissue and placed into the electroporation chamber. Pulsed once and quickly removed the cuvette. A pulse with a time constant of 4.5 to 5msec was achieved. Immediately added 960μl of 37°C SOC medium to resuspend the cells. Transferred each cell suspension to a sterile 15ml polypropylene tube. Incubated the sample at 37°C for 90min while shaking at 225rpm. Using a sterile spreader, spread the transformation mixture onto LB-ampicillin agar plates, allowed to dry, inverted and incubated overnight at 37°C.

2.2.2.iv. Agarose Gel Electrophoresis:

DNA/RNA samples were resolved by gel electrophoresis. Agarose gels were prepared by melting the required amount of agarose to give a particular percentage strength gel, agarose : 1X TAE buffer. When the gel mix was below 50°C ethidium bromide was added to a final concentration of 0.5μg/ml. (No ethidium bromide was added to the 1X TAE running buffer as no suitable disposal available. This causes the dark and light variations of the gel as the ethidium bromide travels in the opposite direction to the DNA). The appropriate amount of DNA/RNA was mixed with loading dye and loaded onto the gel. For DNA gels, the appropriate molecular weight marker was also loaded (Table 2.2).

Electrophoresis was performed at the required voltage over varying time periods in a Pharmacia GNA-100 tank in 1XTAE buffer. Nucleic acid bands were then visualised under ultraviolet light and a picture taken.

2.2.2.v. Gel purification and DNA precipitation:

Ran the preparative digest on the 1% agarose gel and then cut out the required DNA bands avoiding excessive exposure to ultraviolet light.

Using the Sephaglas™ BandPrep kit (Amersham Int. plc, Bucks, UK) eluted the DNA from the gel slice. Transferred the gel slice to a pre-weighed 1.5ml microcentrifuge tube. Determined the weight of the actual gel slice. It weighed <250mg so the following volumes were used, however if it weighs >250mg the reagent volumes are increased proportionately as advised by the manufacturer. 250μl of gel solubiliser was

added, vortexed vigorously and incubated at 60°C for 5-10min or until the agarose slice was dissolved. Vortexed the container of Sephaglas BP to form a uniform suspension, ensured no clumps of glass remained. Added 5µl of the suspension to the dissolved gel slice and vortexed gently. Incubated for 5min at room temperature, vortexing gently every minute to resuspend the Sephaglas. Pulsed for 30sec in a microcentrifuge and carefully removed the supernatant, taking care not to disturb the Sephaglas pellet. Repeated the pulse spin for 30sec in a microcentrifuge and carefully removed any residual liquid, taking care not to disturb the pellet. Added 80µl of Wash Buffer to the pellet and pipetted up and down several times to resuspend the pellet. Pulsed for 30sec in a microcentrifuge and carefully removed the supernatant, taking care not to disturb the pellet. Repeated this step twice for a total of three washes.

Tapped the tube to partially disperse the pellet. Allowed the pellet to air-dry for at least 10min. Checked to be sure that the glass was completely dry. Added 20µl of Elution Buffer, vortexed gently to resuspend the pellet and incubated for 5min at room temperature with periodic agitation. Spun at high speed for 1min in a microcentrifuge. Carefully removed the supernatant and placed it in a clean microcentrifuge tube, taking care not to disturb the pellet. The eluted DNA was stored at –20°C until required.

2.2.2.vi. DNA Ligation:

Better to have a higher concentration of the insert than the digested plasmid so as to increase the probability of them combining to form a recombinant plasmid.

Vector : Insert MOLAR ratio between 1:5 and 1:10.

Mixed ligation components together; Xµl Vector, Yµl Insert, 2µl Ligation Buffer, 1µl Ligase, (20 – (X+Y+3) µl) Sterile H_2O.

Gently flicked the bottom of the tubes to mix the contents and then microfuged for 2-3 seconds so that all of the contents were at the bottom of the tube. Incubated overnight at 16°C.

2.2.2.vii. Single Analytical Digest:

For analytical analysis used 200-500ng of plasmid DNA in a 20µl digest.

Mixed the following components; Xµl DNA, 2µl Digest Buffer, 1µl Restriction Enzyme, (20 – (X+3)µl) Sterile H_2O. If BSA was required by the restriction enzyme being used, added to a final concentration of 100µg/ml. Gently flicked the bottom of the tube to mix the contents and then microfuged for 2-3 seconds so that all of the contents were at the bottom of the tube. Incubate at the optimum temperature for about 2-3 hours.

2.2.2.viii. Single Preparative Digest:

For preparative digests used 750-1500ng in a total volume of 50µl.

Mixed the following components; Yµl DNA, 5µl Digest Buffer, 2µl Restriction Enzyme, (50 – (Y+7)µl) Sterile H_2O. If BSA was required by the restriction enzyme being used, added to a final concentration of 100µg/ml. Gently flicked the bottom of the tube to mix the contents and then microfuged for 2-3 seconds so that all of the contents were at the bottom of the tube. Incubated at the optimum temperature for about 2-3 hours.

2.2.2.ix. Large Scale Preparation of Plasmid DNA:

DNA was isolated using Qiagen Maxi Prep Kit as described in the manufacturers protocol (Qiagen Ltd., Sussex, UK). Picked a single colony from a freshly streaked selective LB-Agar plate and inoculated a starter culture of 5ml LB medium containing the appropriate selective antibiotic. Incubated for about 8 hours at 37°C with shaking at 250rpm. Diluted the starter culture 1/500 into selective LB medium. Grew at 37°C for 12-16 hours with shaking at 250rpm. Harvested the bacterial cells by centrifugation at 6000g for 15min at 4°C. Resuspended the bacterial pellet in 10ml of Buffer P1 (50mM Tris.Cl pH8.0, 10mM EDTA, 100µg/ml RNAse A). Added 10ml of Buffer P2 (200mM NaOH, 1% SDS), mixed gently but thoroughly by inverting 4-6 times and incubated at room temperature for 5min. Added 10ml of chilled Buffer P3 (3.0M potassium acetate pH5.5), mixed immediately but gently by inverting 4-6 times and incubated on ice for 20min. Centrifuged at 20000g for 30min at 4°C. Removed

supernatant containing plasmid DNA promptly. Re-centrifuged the supernatant at 20000g for 15min at 4°C. Removed supernatant containing plasmid DNA promptly. Equilibrated a QIAGEN-tip 500 by applied 10ml Buffer QBT (750mM NaCl, 50mM MOPS pH7.0, 15% Isopropanol, 0.15% Triton®X-100) and allowed the column to empty by gravity flow. Applied the supernatant containing the plasmid DNA to the QIAGEN-tip and allowed it to enter the resin by gravity flow. Washed the QIAGEN-tip with 2 x 30ml Buffer QC (1.0M NaCl, 50mM MOPS pH7.0, 15% Isopropanol). Eluted DNA with 15ml Buffer QF (1.25M NaCl, 50mM Tris.Cl pH8.5, 15% Isopropanol). Precipitated DNA by adding 10.5ml (0.7 volumes) room temperature isopropanol to the eluted DNA. Mixed and centrifuged immediately at 15000g for 30min at 4°C. Carefully decanted supernatant. Washed DNA pellet with 5ml of room temperature 70% ethanol and centrifuged at 15000g for 10min. Carefully decanted supernatant without disturbing the pellet. Air-dried the pellet for 10min and dissolved the DNA in a suitable volume of buffer TE, pH 8.0. Stored the purified plasmid at -20°C until required.

2.2.2.x. Preparation of Plasmid DNA:

Plasmid DNA was prepared using the Promega Wizard® PlusSV Minipreps Kit (Promega Ltd., Southampton, UK). Picked a single colony from a fresh selective LB-Agar plate and inoculated 10ml of LB medium containing the same antibiotic. Incubated overnight (12-16 hours) at 37°C with shaking at 225rpm. Harvested 1ml of bacterial culture by centrifugation for 5min at 10000g. Poured off supernatant and blotted the inverted tube on a paper towel to remove excess medium. Added 250µl of Cell Resuspension Solution and completely resuspended the cell pellet by pipeting. Added 250µl of Cell Lysis Solution (0.2M NaOH, 1% SDS) and mixed by inverting the tube 4 times. Incubated at room temperature until the cell suspension cleared, between 1-5min, no longer. Added 10µl of Alkaline Protease Solution and mixed by inverting the tube 4 times. Incubated for 5min at room temperature. Added 350µl of Neutralisation Solution (4.09M guanidine hydrochloride, 0.759M potassium acetate, 2.12M glacial acetic acid, pH4.2) and immediately mixed by inverting the tube 4 times. Centrifuged the bacterial lysate at 14000g in a microcentrifuge for 10min at room temperature. Transfered the cleared lysate to the prepared Spin Column by

decanting. Avoided disturbing or transferring any of the white precipitate with the supernatant. Centrifuged the supernatant at 14000g in a microcentrifuge for 1min at room temperature. Removed the Spin Column from the tube and discard the flow through from the Collection Tube. Reinserted the Spin Column into the Collection Tube.

Added 750µl of Column Wash Solution to the Spin Column. Centrifuged at 14000g in a microcentrifuge for 1min at room temperature. Removed the Spin Column from the tube and discarded the flow through. Re-inserted the Spin Column into the Collection Tube. Repeated wash using 250µl of Column Wash Solution (60mM potassium acetate, 8.3mM Tris-HCl (pH 7.5), 0.04mMEDTA (pH 8.0), 60% ethanol). Centrifuged at 14000g in a microcentrifuge for 2min at room temperature. Transfered the Spin Column to a new, sterile 1.5ml microcentrifuge tube being careful not to transfer any of the Column Wash Solution with the Spin Column. Eluted the plasmid DNA by adding 100µl of Nuclease-Free Water to the Spin Column. Centrifuged at 14000g for 1min at room temperature in a microcentrifuge. Stored the purified plasmid at -20°C.

2.2.2.xi. Genomic DNA preparation:

Washed a p100 plate of ES cells twice with PBS. Trypsinised the cells for about 2 minutes and then inactivated with growth medium. Spun the suspension at 10000g for 5 min. Resuspended the pellet in 1ml PBS, transferred to a microcentrifuge tube and spun at 10000g for 10min. Aspirated the PBS leaving a little behind to resuspend the pellet by flicking the tube. Added 200µl of proteinase K lysis buffer and proteinase K to a final concentration of 250µg/ml. Inverted the tube gently and incubate at 55°C for at least 3-4 hours, with regular inversion. Added 1ml 70% acetone / 5% DMF at -20°C with glass pipette. Inverted vigorously to get DNA out of solution. Spun for 10min at 10000g. Washed pellet twice with 70% EtOH. Dried pellet at room temperature . Dissolved precipitated DNA in 200µl TE pH8.0 by heating to 70°C. Measured concentration on spectrophotometer $OD_{260/280}$.

2.2.2.xii. Plasmid Rescue:

Set up preparative digests (2.3.8.) of genomic DNA (gDNA) harvested from shuttle vector gene trap clones (2.3.5.).

Table 2.2.5 – Restriction enzymes (New England BioLabs, Hertfordshire, UK.) used to give particular flanking regions for each shuttle plasmid used.

PSHUT+124K (A)			PSHUT+124K SD126 (1)		
Enzymes to give both 5' and 3' Flanking Regions	**Enzymes to give 5' Flanking Region Only**	**Enzymes to give 3' Flanking Region Only**	**Enzymes to give both 5' and 3' Flanking Regions**	**Enzymes to give 5' Flanking Region Only**	**Enzymes to give 3' Flanking Region Only**
Hind III	BstE II	Xba I	Bsu36 I	Nae I	Bgl II
EcoR I	Afl II	Sal I	Aat II	Nco I	Hind III
Kpn I	Bsu36 I	Sfi I	Kpn I	Rsr II	Xho I
			Xba I		

2.2.2.xii.a. Ligation of digested gDNA:

The gDNA has to be religated at low concentrations so as to increase the likelihood of circularisation of the various fragments, instead of religation back into chromosome size pieces.

Mix the reaction components as follows; 1µl digested gDNA, 5µl T4 DNA ligase, 10µl ligase buffer, 484µl sterile H_2O. Incubated overnight at 4°C.

2.2.2.xii.b. Precipitation of ligated gDNA:

With such a dilute ligation mixture it was necessary to precipitate the DNA before transforming. DNA was precipitated overnight at -20°C, or 30 minutes at -80°C, by adding 2 volumes of 100% ethanol and 0.1 volume of 3M sodium acetate (pH 4.8).

The DNA was recovered by centrifugation at 13000g for 10 minutes. The supernatant was removed and the DNA pellet dried at room temperature. The DNA was resuspended in an appropriate volume of sterile water.

2.2.2.xii.c. Electroporation of Electro Ten-Blue Electrocompetent Bacteria (Stratagene) with Plasmid DNA:

These electrocompetent bacteria (Stratagene, La Jolla, CA.) have a very high efficiency of $1x10^{10}$ cfu/μg of pUC18 DNA. This high efficiency is needed when working with plasmid rescue as the number of re-ligated plasmids is likely to be low.

To prepare the ligated DNA prior to transfection, Stratagene supply a vial of StrataClean resin to ensure the complete removal of protein contaminants.

Vortexed the StrataClean resin to completely resuspend. Pipetted the 5μl of slurry into the tube containing the ligation reaction. Vortexed mixture for 15 seconds. Pelleted the resin by centrifugation at 2000g for 1min. Carefully removed the supernatant containing the DNA to a fresh microcentrifuge tube. Used the entire 10μl supernatant sample to electroporate into the competent cells. Thawed the electrocompetent cells on ice. Aliquoted 40μl of cells into chilled 1.5ml microcentrifuge tubes, pipeting very gently as electrocompetent cells are very fragile. Added experimental DNA to electrocompetent cells. For optimal efficiency, added 1μl of plasmid DNA (10pg/μl) / 40μl of cells. Completed the procedure as specified in section 2.2.2.iii.

Picked resistant colonies containing rescued plasmid, grew up in LB and then mini prepped (Section 2.2.2.x.) prior to setting up sequencing PCR reaction (Section 2.2.2.xvi.)

2.2.2.xii.d. Rolling Circle Amplification:

Due to the limited amount of circularised plasmid retrieved from the ligation reaction, amplification of that retrieved was attempted using TempliPhi™ DNA amplification kit (Amersham Biosciences, UK Ltd., Bucks, UK).

The reaction mixture consisted of 1μl template DNA (precipitated ligation reaction 2.2.2.xii), 5μl Sample Buffer (random hexamers), 5μl Reaction Buffer (salts and deoxynucleotides), 0.2μl Enzyme mix (ϕ29 DNA polymerasse in 50% glycerol).

Incubated at 30°C for 8 hours. Inactivated polymerase by heating to 65°C for 10 minutes. Proceeded with electroporation (2.2.2.xii.c.).

2.2.2.xiii. Preparation of Cytoplasmic RNA:

Cytoplasmic RNA was isolated using Qiagen RNeasy midi kit (Qiagen Ltd., Sussex, UK), instead of total RNA to try and minimise DNA contamination.

Used $5 \times 10^6 - 1 \times 10^8$ cells grown in a monolayer. Lysed cells directly on culture dish by adding 1ml of cold (4°C) Buffer RLN [50mM Tris.Cl pH 8.0, 140mM NaCl, 1.5mM $MgCl_2$, 0.5% (v/v) Nonidet® P-40 (1.06g/ml)]. Detached cells gently using a cell scraper and transferred to a centrifuge tube. Incubated on ice for 5min. Centrifuged lysate at 4°C for 5min at 500g. Transferred supernatant to an RNase-free 15ml centrifuge tube and discarded the pellet. Added 4ml Buffer RLT. Mixed thoroughly by shaking. Added 2.8ml 100% ethanol and mix thoroughly by shaking. Applied the sample to RNeasy midi column placed in a 15ml centrifuge tube. Centrifuged at 3000g for 5min. Discarded the flow through. Added 15ml Buffer RW1 to the RNeasy column. Centrifuged at 3000g for 5min. Discarded the flow through. Added 2.5ml Buffer RPE to the RNeasy column. Centrifuged at 3000g for 2min. Discarded the flow through. Added another 2.5ml Buffer RPE to the RNeasy column. Centrifuged at 3000g for 5min to dry the RNeasy silica-gel membrane. Eluted by transferring the RNeasy column to a new 15ml collection tube. Pipetted 200µl of RNase-free water directly onto the spin-column membrane. Let it stand for 1min and then centrifuged for 3min at 3000g. Stored RNA at -70°C until required.

2.2.2.xiv. Messenger RNA Extraction:

Used 75µg total RNA and adjusted the volume of total RNA sample to 100µl using RNase free water. Heated the sample to 65°C for 2 min to disrupt secondary structures. Remove 200µl of resuspended Dynabeads Oligo $(dT)_{25}$ into a 1.5ml microcentrifuge tube and place the vial onto the magnetic stand (Promega Ltd., Southampton, UK). After 30 seconds pipetted off the supernatant and washed the Dynabeads Oligo $(dT)_{25}$ once. Removed the tube from the magnet, resuspended the beads in 100µl Binding Buffer (100mM Tris-HCl (pH7.5), 500mM LiCl, 10mM

EDTA (pH8.0), 1% LiDS, 5mM DTT) and placed the tube back on the magnet. After 30 seconds pipetted off the supernatant and removed the tube from the magnet. N.B. Do not leave the Dynabeads Oligo $(dT)_{25}$ unsuspended for a long period as drying of the beads may reduce their efficiency.

Added 100µl Binding Buffer to the Dynabeads Oligo $(dT)_{25}$ and the added the total RNA to the Dynabeads Oligo $(dT)_{25}$ in Binding Buffer.

Mixed thoroughly and annealed by rotating on a mixer for 3-5 min at room temperature. Placed the tube on the magnet for at least 30sec or until the solution is clear and removed the supernatant. Removed the tube and washed twice with 200µl Washing Buffer B (10mM Tris-HCl (pH7.5), 0.15M LiCl, 1mM EDTA) using the magnetic stand. Removed all the supernatant between each washing step. The complete removal of supernatant is extremely important when working with small volumes.

Eluted in 10µl RNase free water by heating to 80-90°C for 2 min. Placed the tube immediately on the magnet and transferred the eluted mRNA to a new RNase free tube.

2.2.2.xv. PCR Amplification:

PCR was performed using a reaction mixture containing the target DNA, 5U *Thermus aquaticus* (*Taq*) polymerase, 0.4µM of specific primer, 0.2mM dNTPs, 2.5mM $MgCl_2$ and 1X *Taq* buffer. Preparative mixtures (100µl) were contained in 500µl thin walled PCR tubes and overlaid with mineral oil. Denaturing, annealing and chain extensions were performed on a Biometra thermocycler (Biometra, Thistle Scientific, Glasgow, UK) programmed according to required conditions. Following amplification, the aqueous layer was transferred to a fresh tube for analysis.

2.2.2.xvi. DNA Sequencing using ABI 3100 System:

The DNA to be analysed (approximately 100ng PCR product) was combined with ABI BigDye™ V2.1 (1µl), Better Buffer (5µl) (Web Scientific Ltd., Crewe, UK) and 1.6µM M13 forward/reverse primer (1.5µl). A final volume of 15µl was achieved by the addition of sterile water. The reaction mix was placed in a Biometra thermocycler

using the programme of 96°C for 3 minutes (initial denaturation), followed by 25 repeated cycles of 96°C for 15 seconds (denaturation), 50°C for 10 seconds (annealing) and 60°C for 2 minutes (extension). Labelled sequencing products were precipitated by the addition of 2.5M EDTA (2µl) and 95% ethanol (60µl). The reaction mix was centrifuged at 2000g for 30 minutes (4°C), then the supernatant removed and the pellet washed with ice-cold 70% ethanol (250µl). The mix was spun for a further 15 minutes at 2000g (4°C), the supernatant removed and the pellet air-dried.

Prior to sequence analysis, the pellet was dissolved in 15µl sequencing grade formamide and denatured at 95°C for 2 minutes before placing on ice for 1 minute. Sequence analysis was performed by capillary electrophoresis by the sequencing technician, Mr. Ezra Linley.

2.2.2.xvii. Quantification of Nucleic Acid Concentration by Spectrophotometry:

Nucleic acids absorb at a wavelength of 260nm. Given that an absorbance of 1 OD is equivalent to 50µg/ml of double stranded DNA or 40µg/ml single stranded DNA and RNA, it was possible to quantify DNA and RNA concentrations. Furthermore, determining absorbencies at 230nm and 280nm provides a means of examining sample purity. Pure preparations of DNA and RNA will have A_{260}/A_{280} ratio of approximately 1.8 and 2.0 respectively. Any protein or phenol contaminants will lower these values. Likewise a A_{260}/A_{280} ratio of less than 2 is considered to be indicative of ethanol or salt contamination (Sambrook and Gething, 1989).

Optical density readings were performed on a Camspec M330 spectrophotometer (Camspec, Cambridge, UK).

2.2.2.xviii. PCR Screening of Bacterial Colonies:

PCR was used to amplify cDNA inserts from bacterial colonies harbouring plasmid DNA. A single recombinant colony was inoculated into 1ml of LB medium containing the appropriate antibiotic selection. The culture was incubated with constant agitation (37°C for 3 h at 300rpm). A 500µl aliquot was placed in a sterile microcentrifuge tube and the cells were pelleted by centrifugation (10 000g for 1min

at RT). The medium was removed and the cells were resuspended in 50µl of sterile dH_2O. The cells were lysed (95°C for 5 min) and centrifuged (10 000g for 1 min at RT) to pallet any debris. The supernatant (which contained the plasmid DNA template) was removed to a clean microcentrifuge tube and stored (-20°C) for future use.

Amplification reactions containing 2µl template, 0.2mM dNTP's 0.4µM M13 forward/reverse primers, 1x reaction buffer, 2.5mM $MgCl_2$ and 5U of *Taq* DNA polymerase were aliquoted into sterile 0.5ml microcentrifuge tubes. The reactions were made up to a final volume of 20µl with sterile dH_2O. The tubes were briefly centrifuged (10 000g for 10 seconds at RT) and a drop of mineral oil was laid on top to minimise evaporation. Amplifications were performed as detailed previously (Section 2.3.15.).

2.2.2.xix. Purification of PCR Products Using QIAquick PCR Purification Kit:

PCR products were purified using QIAquick PCR purification kit (Qiagen Ltd, Crawley, UK), as described in the manufacturers protocol.

Added 5x volume of buffer PB to 1x PCR product and mixed thoroughly. The solution was transferred to a QIAquick spin column and spun at 10000g for 1 minute. To wash, added 750µl buffer PE and spun at 10000g for 1 minute, discarded the flow through and spun for a further minute to dry the QIAquick spin column membrane. Moved the column to a clean 1.5ml microcentrifuge tube and eluted the DNA in 50µl TE pH8 by adding it to the centre of the QIAquick spin column membrane and spinning for 1 minute at 10000g. Purified cDNA was stored at -20°C.

2.3. CREATING TRANSGENICS

NB: Even though no transgenics were ever created in this project it was felt that it was appropriate to include this section of procedures as a great deal of time and effort was dedicated to this endeavour.

All animals were maintained in accordance with Animal Scientific Procedures Act 1986.

2.3.1. Collecting blastocysts:

On day 3.5 post coitum (pc) the pregnant female is killed by cervical dislocation. Abdominal skin is sterilised by swabbing with 70% ethanol.
Grasped the skin with blunt forceps and lifted, using a pair of dressing scissors opened the abdominal cavity to display the contents. Displaced the guts to one side, located and removed the two uterine horns. Trimmed away the associated adipose tissue and mesenteries using a pair of fine scissors. Filled a 1ml disposable syringe with Optimem media, BSA (4mg/ml) and 2x penicillin-streptomycin, attached a 25-gauge hypodermic needle. Introduced the needle into one end of the uterine horn and recovered the contents by flushing a stream of medium through the uterine lumen into a p35 tissue culture dish (Nunc, Fisher Scientific Equipment UK Ltd., Loughborough, UK). Discarded the one horn and repeated with the remaining horn. The resulting blastocysts were then transferred using a drawn out Pasteur pipette held in a mouth-controlled tube to small drops of pre-equilibrated media under inert liquid paraffin oil and returned to a 37°C, CO_2 incubator until required.

2.3.2. Injecting selected ES cell clones into host blastocysts:

2.3.2.i. Preparation of pipettes:

Before injecting pipettes were pulled on an automatic electro-magnetic electrode puller (Model 773, micropipette puller, Campden Instruments, Sileby, UK). The glass capillary tubing used to construct the pipettes was thin walled (1.0mm external diameter, 0.8mm internal diameter) borosilicate glass without a fibre (Harvard Apparatus, Edenbridge, Kent). Once a number of pipettes have were pulled they were fashioned into either injection or holding pipettes.

2.3.2.i.a. Injection pipettes; have an internal diameter only marginally larger than the stem cells they are designed to inject and a sharp point which allows the needle end to penetrate cleanly into the blastocoel cavity. This was achieved by snapping the glass capillary by hand under a stereo dissecting microscope, at a magnification of x40, using a sharp scalpel blade and a spongy transparent silicon rubber sheet. As thin walled borosilicate glass was used the pulled glass usually breaks with a sharp bevelled end when touched with the edge of the scalpel blade. Using a De Fonbrune microforge (Beaudouin, Rue Ratud, Paris) a 30° bend was introduced and the orientation of the bend is important as it must be ensured that the bevel is facing the experimenter when in place in the injection chamber.

2.3.2.i.b. Holding pipettes; are blunt flame-polished pipettes through which suction is applied to immobilise the blastocyst. An appropriate diameter is about 80 -100μm. Secondary structure was necessary to enable the micro-instruments to be moved freely in the injection chamber (using the joystick manipulator arm) in the plane of focus without the necessity of re-adjusting their position. Using a De Fonbrune microforge (Beaudouin, Rue Ratud, Paris) 30° bends were introduced.

2.3.2.ii. Injecting:

Using a magnification of x200 selected the cells for injection from the population seeded on the floor of the injection chamber. Collected the cells into the injection pipette by gentle suction. Selected about 12-15 cells and positioned them near the tip of the injection pipette.

Orientated the blastocyst so that the inner cell mass (ICM) was at the side being held by the pipette. Did not apply too much suction as this would cause the blastocyst to collapse. When the injection pipette was positioned correctly pushed the pipette into the blastocoel cavity. The speed of the movement was important and had consequences on the success of the injection. If entry was too slow and hesitant, the pressure in the blastocoel would be released and the embryo would collapse before the pipette had penetrated. If the entry movement was very fast and uncontrolled, the pipette would enter the blastocoel and it is easy to accidentally damage the end on the opposing holding pipette, or damage the ICM component of the embryo.

When the pipette was clearly within the blastocoel the cells were expelled from the injection pipette by gently increasing the pressure.

When the cells had been expelled, slowly withdrew the pipette. Retained the blastocyst by suction on the holding pipette, changed to lower power objective and moved the embryo to a remote location in the injection chamber. Released the blastocyst by reducing the suction.

Repeated this procedure on successive blastocysts, placing successful embryos in one group and segregating failed embryos in another group.

Transferred the successfully injected embryos back to the culture drops and incubated at 37°C. Re-expansion of the blastocysts normally occurs within 1-3 hours.

2.3.3. Surgical transfer of embryos to pseudo-pregnant females:

Embryos were transferred to 2.5 day p.c. pseudo-pregnant foster mothers 1-6 hours after injection. If they are left too long after injection the re-expanding embryos begin to hatch from the zona pellucida which makes them more difficult to handle. A minimum of five blastocysts per uterine horn were transferred. If fewer were transferred, embryonic growth would be too rapid, resulting in a large foetus which could exceed the nutritive capacity of the placenta and result in pre-natal death.

Weighed the mouse and anaesthetised by intra-peritoneal injection of Avertin anaesthetic (0.02ml/g body weight). Placed the mouse dorsal side uppermost and swabbed the back with a liberal quantity of 70% ethanol. Grasped the skin on the midline with blunt forceps at the midpoint of the back and raised the skin away from the body wall. Holding the dressing scissors vertically cut the skin with a single action and produced a lateral incision about 4-5mm long. Used the incision in the skin as a "window" by sliding it ventrally over the surface of the underlying body wall. Both uterine horns could be accessed through this single incision. Moved the hole to a position about one third of the distance between the dorsal and ventral surfaces. It was possible to see through the body wall and distinguish the pink ovary and cream ovarian fat pad against the dark contents of the gut. When the general position had been located used fine forceps to grasp the slippery body wall. Made a small incision in the body wall using fine iris scissors. It was important to avoid cutting the blood vessels which run dorso-ventrally. Holding the cut edge of the body

wall with fine forceps, inserted blunt forceps into the abdominal cavity and grasped the ovarian fat pad. It was important to handle the reproductive tract by the associated fat only. By gently pulling the fat through the opening it was followed by the ovary, oviducts and finally the upper part of the uterus. Under the microscope loaded the transfer pipette (drawn-out Pasteur pipette) with embryos. Included two marker air bubbles above and below the position of the embryos in the capillary. Under the stereo dissecting microscope (x10 magnification) focused on the exposed uterine horn. Gently grasped the top of the uterus with blunt forceps and selected an area which is relatively devoid of blood vessels. Inserted a 25-gauge hypodermic needle through the uterine wall into the lumen. Whilst looking at the hole picked up the loaded transfer pipette and inserted the end about 3mm into the uterine lumen. Gently blew the embryos into the uterus using the air bubbles to monitor the transfer. After transfer, checked that the embryos have been transferred by expelling the remainder of the fluid in the pipette back into the culture drop and examined under the microscope. Grasped the edge of the incision with fine forceps and gently pushed the uterus, fat pad and ovary back into the peritoneal cavity using the blunt forceps. Repeated the process on the other side. Closed the incision by clipping the skin with one or two Michel clips. When the foster mother(s) regained consciousness transferred them to clean cages in groups of two to four. They should litter 17-18 days later.

2.4. PHENOTYPE SCREENING

Table 2.4.1. Section 2.4. routinely used primers, sequences and suppliers.

Primer	DNA Sequence (5'→3')	Supplier
M13 forward	GTTTTCCCAGTCACGAC TT	MWG Biotech UK., Milton Keynes, UK.
M13 reverse	CAGGAAACAGCTATAG ACCATG	MWG Biotech UK., Milton Keynes, UK.
Universal Adapter Primer (UAP)	CUA CUA CUA CUA GGC CAC GCG TCG ACT AGT AC	Invitrogen, Paisley, UK.
Adapter Primer (AP)	GGC CAC GCG TCG ACT AGT ACT TTT TTT TTT TTT TTT T	Invitrogen, Paisley, UK.
Nested 3'RACE UDG primer	CAU CAU CAU CAU TTC CTC GTG CTT TAC GGT ATC	MWG Biotech UK., Milton Keynes, UK.
Gene Specific Primer (Isce-1)	CAU CAU CAU CAU TAG GGA TAA CAG GGT TAA T	MWG Biotech UK., Milton Keynes, UK.
Neo Specific Primer (3'RACE UDG)	CAU CAU CAU CAU ATA TTG CTG AAG AGC TTG GCG GC	MWG Biotech UK., Milton Keynes, UK.
gDNANEO R	GCA CGA GGA AGC GGT C	MWG Biotech UK., Milton Keynes, UK.
gDNANEO F	TGC GCA GCT GTG CTC GAC G	MWG Biotech UK., Milton Keynes, UK.
Template switch (TS)	AAG CAG TGG TAA CAA CGC AGA GTA CGC GGG	MWG Biotech UK., Milton Keynes, UK.
Oligo-dT	TTTTTTTTTTTTTT	Pharmacia Biotech Ltd., St. Albans, UK.

2.4.1. T7 Enrichment step for pCRR+SD126 T7 I-Sce-X Selected Clones:

2.4.1.i. <u>First strand cDNA synthesis</u>

Mixed in a PCR reaction tube 0.01 - 5μg tRNA in 9μl DEPC H_2O with 1μl (0.1 – 0.5μg/μl) oligo dT primer and heated to 70°C for 3 mins.

Cooled to room temperature and added the following; 4μl 5X First strand buffer (supplied with Superscript II enzyme), 1μl (0.1 – 0.5μg/μl), template switch (TS)

primer, 2µl 0.1M DTT, 1µl RNAsin (Promega Ltd. Southampton, UK), 2µl 10mM dNTP, 2µl Superscript II (200U/µl, Invitrogen, Paisley, UK). Heated to 42°C for 90min in thermal cycler.

2.4.1.ii. Second strand cDNA synthesis

To the reaction tube added 106µl DEPC H_2O, 15µl Advantage PCR buffer (supplied with Advantage cDNA polymerase), 3µl 10mM dNTP mix, 1µl RNase H (2U/µl Invitrogen, Paisley, UK), 3µl Advantage cDNA Polymerase mix (Clontech, Basingstoke, UK).

Ran in thermal cycler (37°C for 5 min to digest mRNA, 94°C for 2 min to denature, 65°C for 1 min for specific priming, 75°C for 30 min for extension). Stopped reaction with 7.5µl 1M NaOH solution containing 2mM EDTA, incubated at 65°C for 10min to inactivate enzymes. The reaction was stored at -20°C.

2.4.1.iii. Double stranded cDNA cleanup

Phenol-Chloroform-Isoamyl isolation and ethanol precipitation.

Added 1µl Linear Acrylamide (0.1µg/µl), (Ambion, AMS Biotechnology(UK) Ltd., Abingdon, UK) and 150µl Phenol:Chloroform:Isoamyl alcohol (25:24:1) (Boehringer Mannheim), mixed well by pipetting.

Transferred the slurry solution to phase lock gel tube (Eppendorf) and spun at 14000g for 5min at room temperature. Transferred the aqueous phase to an RNase/DNase free 1.7ml tube and added 70µl of 7.5M ammonium acetate and then 1ml 100% ethanol. Mixed well and centrifuged immediately at 14000g for 20 min at room temperature to prevent co-precipitation of oligos. A white pellet suggested successful precipitation. Washed the pellet with 800µl 100% ethanol and spun down at 14000g for 8 min. Repeated the washing step once. Air dried the pellet and resuspended the double stranded cDNA in 70µl DEPC H_2O.

2.4.1.iv Gel-filtration

Prepared micro bio-spin chromatography column (Bio-gel P-6 column in Tris, Bio-Rad). Inverted the column several times to resuspend the settled gel and remove any bubbles. Snapped off the tip and placed the column in a collection tube. Removed the cap and let the buffer drain out.

Discarded the drained buffer and applied 700µl DEPC treated H_2O to the column. Discarded the flow through and repeated the step 3 times.
Spun column at 700g for 2 mins at room temperature. When opening the column aspirated off any gel from the underside of the cap and washed with DEPC H_2O to prevent contamination. The provided collection tubes were not RNase-free, so ensured RNase-free tubes were used instead to collect the sample. Loaded 70µl sample to the centre of the column and spun at 700g for 4 min. Dried collected sample and resuspended in 8µl DEPC H_2O.

2.4.1.v. In Vitro T7 Enrichment Step

To the 8µl double stranded cDNA sample added [2µl of each 75mM NTP (A, G, C and UTP), 2µl reaction buffer, 2µl enzyme mix (RNase inhibitor and T7 phage RNA polymerase)]
Incubated the mix for 6 hours at 37°C

2.4.1.vi. Amplified RNA purification

Added 1µl of TRIzol to the *in vitro* transcription tube and mixed well by pipetting up and down. Added 200µl chloroform/1ml chloroform. Mixed by inverting the tubes for 15 seconds and incubated at room temp for 2-3 min. Centrifuged at 12000g for 15 min at 4°C. Transferred the aqueous phase to a new RNase free tube and added 500µl of isopropanol/1ml TRIzol. Incubated sample at room temp for 10 min and then centrifuged at 14000rpm for 15min. Washed pellet twice with 1ml 70% ethanol in DEPC treated water. Air dried pellet and quickly resuspended in 20µl DEPC-treated water. Checked RNA concentration and quality by measuring OD_{260} and $OD_{260/280}$ (Section 2.2.2.xvii.).

2.4.2. First Strand cDNA synthesis and 3′ RACE of Splice Acceptor Series Selected Clones:

2.4.2.i. First strand cDNA Synthesis –

Used 1 to 5µg of total RNA OR 50ng of mRNA, made the sample volume up to 11µl using RNase free water (Ambion, AMS Biotechnology (UK) Ltd., Abingdon, UK) in

a 0.5ml microcentrifuge tube. Added 1µl of 10µM Adapter Primer (Invitrogen Ltd., Paisley, UK), mixed gently and collected the reaction by brief centrifugation. Heated at 70°C for 10min and then chilled on ice for at least 1min. Collected the contents by brief centrifugation and added the following, (2µl 10X PCR buffer, 2µl 25mM $MgCl_2$, 1µl 10mM dNTP mix, 2µl 0.1M DTT). Mixed gently and collected the reaction by brief centrifugation.

Equilibrated the mixture to 42°C for 5min. Added 1µl of Superscript II RT and incubated the tube at 42°C for 50min. Terminated the reaction by incubating at 70°C for 15min. Chilled on ice. Collected the reaction by brief centrifugation. Added 1µl of RNase H, mixed and incubated for 20min at 37°C. The cDNA can be stored at -20°C but it was found that better results were obtained if the cDNA was immediately transferred to the 3′ RACE PCR reaction.

2.4.2.ii 3′ RACE PCR Reaction (Frohman et al., 1988)-

Made up the following 3′ RACE PCR mix per reaction; 10µM Neo Specific Primer 1µl (3′RACE UDG), 10µM Universal Adapter Primer 1µl (UAP), Taq DNA Polymerase (2 to 5 units/µl), 10X PCR buffer 5µl, 25mM $MgCl_2$ 3µl, 10mM dNTP mix 1µl, Sterile dH_2O 36.5µl.

Added 2µl of the cDNA synthesis reaction (2.4.2.i.) to the tube. Mixed gently and collected by brief centrifugation. Covered the reactions with a thin layer of mineral oil to minimise evaporation. The reactions were place in Biometra thermocycler using the programme of 95°C for 3 minutes (initial denaturation), followed by 30 repeated cycles of 95°C for 1 minute (denaturation), 66°C for 1 minute (annealing) and 72°C for 3 minutes (extension). There was a final 10 minutes at 72°C to complete extension.

Analysed the amplified sample using agarose gel electrophoresis with the appropriate molecular size standards and ethidium bromide staining (2.2.2.iv). If no definite product was visible, proceeded with Nested PCR amplification (2.4.2.iii.).

2.4.2.iii. Nested 3′ RACE PCR Reaction:

The nested amplified reaction was conducted if there was no definite, discrete band when the original 3′ RACE reaction was analysed on an agarose gel (2.2.2.iv.). If

there was a smear of bands then using a capillary pipet removed a small (~ 5µl) agarose plug from the areas of the gel containing the bands. Prepared an amplification mix as in 2.4.2.ii, except used the Nested 3′ RACE UDG primer in place of the original 3′ RACE UDG primer. Performed the PCR reaction as in 2.4.2.ii. and analysed the amplified sample using agarose gel electrophoresis with the appropriate molecular size standards and ethidium bromide staining (2.2.2.iv).

2.4.3. Cloning 3′RACE Products Using Uracil DNA Glycosylase (UDG):

The relevant 3′ RACE band products were cut out of the agarose gel and purified using Sephaglas Band Prep Kit (Amersham Int. plc, Bucks, UK) (2.2.2.v.)
To a 0.5ml microcentrifuge tube on ice, add the following components;
10-50ng isolated 3′ RACE PCR products, 25 ng/µl pAMP1 vector DNA (Invitrogen, Paisley, UK.), 1X Annealing Buffer (20mM Tris-HCl (pH8.4), 50mM KCl, 1.5mM $MgCl_2$), 1U/µl Uracil DNA Glycosylase (Invitrogen, Paisley, UK). Mixed and then incubate at 37°C for 30min. After incubation placed the reaction tube on ice. After annealing, a portion of the reaction mix (1 - 5µl) was used for transformation into ElectroTen Blue Electrocompetent Cells (Stratagene) (2.2.2.xii.c).

2.4.4. Beta Galactosidase Staining of Eukaryotic Cells in vitro:

Washed confluent p60 plate 3 times with PBS being careful not to dislodge cells from the surface. Aspirated PBS, added 3ml of 0.2% Glutaraldehyde / PBS. Fixed the cells for 5 min at room temperature.
After 5 min aspirated the fix solution and washed the cells 3 times with PBS. Essential to remove all the glutaraldehyde as this could inhibit the staining reaction. Covered the cells with 4ml of freshly prepared X-Gal solution (1mg/ml X-Gal (20mg/ml in dimethylformamide), 5mM Potassium ferricyanide (0.5M made in sterile dH_2O), 5mM Potassium ferrocyanide (0.5M made in sterile dH_2O), 2mM $MgCl_2$, made up volume using PBS). Left cells at 37°C overnight in a humidified incubator, the plate would dry out quickly if incubated in a normal oven.

2.4.5. Sequence analysis:

Obtained DNA sequence data as specified in 2.2.2.xvi. and the nucleotide sequences were subject to comparison of all database entries:

2.4.5.i. BioEdit Sequence Alignment Editor

The computer programme BioEdit sequence alignment editor (version 4.8.10) was used to retrieve and open a DNA sequence. The sequence was then saved in FASTA format for further investigation.

2.4.5.ii. NCBI Database Entrez Nucleotide

Used VecScreen on this site to discover if sequence information retrieved from 3' RACE products was in fact vector sequence, before proceeding further.

2.4.5.iii. The CAP EST Assembler at IFOM

Used this site (http://bio.ifom-firc.it/ASSEMBLY/assemble.html) to create a consensus sequence of multiple sequence data obtained for a single clone.

2.4.5.iv. NCBI Database

Once a consensus sequence minus vector sequence was obtained the DNA sequences were compared to all database entries, searching for nucleotide similarities (BlastN searches), using the Blast search analysis programme.

2.4.6. High Throughput 3' RACE Screen for Gene Trapped Clones:

When 96 well plate of clones was generally confluent, aspirated medium and washed cells with PBS. Kept plate cool on ice, added 20µl/well lysis buffer (0.5% Nonidet P40, 20mM Vanadyl-Ribonucleoside Complexes, 1mM DTT, 10mM Tris-Cl pH8.6, 1.5mM $MgCl_2$, 0.14M NaCl).

Left for about 5min with constant agitation. Spun the plate at 500g for 5min @ 4°C. Transferred supernatant to a fresh 96 well PCR plate and added 20µl binding buffer (0.2M Tris-Cl pH 8.0, 25mM EDTA pH 8.0, 0.3M NaCl, 2% SDS, 5mM DTT, Filter sterilised 0.2um) (1:1, lysis buffer : binding buffer), mix.

Added oligo dT_{25} beads – (Removed 20µl/well of beads, ~ 2ml / 96well, resuspended in 10µl/well, 1ml binding buffer).

Aliquoted 10µl/well. Incubated at room temperature ~ 10min with agitation. Moved to magnetic plate, removed supernatant (~50µl), resuspended beads in 10µl RNase free water. Heated at 85°C ~2min, then transferred directly to magnetic plate. Transferred supernatant to fresh PCR plate (~10µl). Set up RT Superscript II reaction. Aliquoted 9µl of RT Superscript II reaction mix (200µl 10x buffer, 200µl 25mM $MgCl_2$, 100µl 10mM dNTP, 200µl 0.1M DTT, 100µl 10uM oligo dT, 100µl Superscript II) per well then transferred the plate to PCR machine to run (50min @ 42°C, 15min @ 70°C)

Removed plate from PCR machine and put on ice. Spun down briefly, added 1µl RNaseH/well. Incubated 20min @ 37°C. Aliquoted out 2µl/well of cDNA to a fresh 96 well PCR plate. Added 48µl/well of 3' RACE master mix (500µl 10x PCR Buffer, 300µl 25mM $MgCl_2$, 100µl 10mM dNTP, 100µl 10µM Adapter primer, 100µl Neomycin primer 10µM, 50µl Taq Polymerase (Promega), 3650µl dH_2O).

Transferred plate to PCR machine and run on 3' RACE programme; 3min @ 95°C (initial denaturation), then 35 cycles of 1min @ 95°C (denaturation), 1min @ 58°C (annealing), 1min @ 72°C (extension). A final 10min @ 72°C to complete extension. Analysed the amplified sample using agarose gel electrophoresis with the appropriate molecular size standards and ethidium bromide staining (2.2.2.iv).

2.4.7. Genomic DNA PCR:

Set up PCR reaction as specified in Section 2.2.2.xv. using the primers gDNANEOF and gDNANEOR (MWG Biotech UK., Milton Keynes, UK.) (Table 2.4.1.). Ran the reaction on a programme of 3minutes at 95°C (initial denaturation), then 30 cycles of 1 minute at 95°C (denaturation), 1.5 minutes at 65°C (annealing), 1.5 minutes at 72°C (extension). A final 10minutes at 72°C to complete extension. Analysed the amplified sample using agarose gel electrophoresis with the appropriate molecular size standards and ethidium bromide staining (2.2.2.iv).

CHAPTER 3

RESULTS

3. RESULTS

Due to the way this project has progressed it was felt that it would be clearer to present the results as a single chapter, thereby allowing the flow of progression to be appreciated.

3.1. RETROVIRAL VECTORS

As mentioned in the introduction retroviral vectors have become a standard tool for gene transfer technology. Compared with other gene transfer systems, retroviral vectors have several advantages, including high efficiency of gene transfer and insertion of transferred genes into host cell chromosomes. When compared to plasmid vectors, retroviral vectors are advantageous primarily because they insert a single copy with the ends of the vector intact. Plasmid-based vectors are susceptible to nucleases during electroporation and recombination into the genome. Deletions from the ends of the vector would reduce the efficiency of cloning and make primer selection more difficult.

Construction of retroviral vectors involves placement of gene(s) to be transferred between viral long terminal repeats (LTRs), which contain elements required to drive gene expression, reverse transcription and integration into the host chromosome (Armentano et al., 1987).

The vectors used in this project are based on the oncogenic Moloney murine leukaemia virus (MoMLV) and essentially all of the viral protein encoding regions can be deleted with retention of a transmissible vector. Oncogenic retroviruses require a dividing host cell to allow the pre-integration complexes to enter the nuclei. However, lentiviruses such as human immunodeficiency virus type 1 (HIV-1) are less restricted by cell proliferation. Unlike MoMLV, which have to wait until the nuclear membrane dissolves during cell division, HIV-1 can enter the nuclei of nondividing cells through the nuclear pore (Lewis and Emerman, 1994).

The ability of retroviruses to stably integrate their own genome into the host genome of many cells has led to the production of vector systems capable of long-term gene transfer.(Bender et al., 1987) The biology of retroviruses has made possible the development of packaging cell lines which provide in *trans* all the viral proteins required for viral particle formation. The design of different types of packaging cells

has evolved to reduce the possibility of helper virus production. (Miller and Buttimore, 1986)

3.2. PACKAGING CELLS

Retroviral vectors promote the efficient transfer of genes into a variety of cell types from many animal species. An important contribution to their utility was the development of retrovirus packaging cells, which allow the production of retroviral vectors in the absence of replication-competent virus (RCV). A critical element in producing the components to carry out retroviral-mediated gene transfer, is the cell that generates the retroviral particles carrying the gene to be transferred. These cells are called "packaging cells" because they "package" the retroviral vector into the delivery vehicle, carrying the exogenous gene(s) to be delivered to the cells (Miller, 1990).

After the formation of a retrovirus virion, no further viral protein synthesis is required for the events leading to retrovirus integration into the genome of an infected cell. All of the viral protein coding regions can be removed from a replication competent virus without markedly affecting the ability of viral RNA to be encapsidated, reverse transcribed and integrated. This is the principle behind the retroviral vector system (Mann et al., 1983).

The retroviral vector consists of DNA sequences intended for transfer flanked by the signals present at the ends of the retroviral genome and the packaging cells are designed to produce all of the retroviral proteins and promote "packaging" of retroviral vector RNA into virion. Retroviral vectors produced by using packaging cells are expected to infect cells but be unable to replicate further.

3.3. REVERSE ORIENTATION SPLICE ACCEPTOR (ROSA)

A couple of the retrovirus vectors used in this project are based on the ROSAβ-geo design (Friedrich and Soriano, 1991). This means that the splice acceptor (SA) β-geo cassette is inserted in reverse orientation relative to viral transcription. This is necessary to avoid removing the viral ψ packaging sequence by splicing from an upstream splice acceptor sequence. The resulting retroviral vector is termed ROSAβ-

geo, which due to the reporter gene β-geo, also exhibits kanamycin resistance along with neomycin resistance and β-gal activity as well.

3.4. INTRODUCING THE PACKAGING CELL LINE USED IN THIS PROJECT

A retroviral vector contains the *cis*-acting elements required for replication as a virus, but is missing some of the viral genes, which are replaced by foreign coding sequences. In order for a retrovirus vector to replicate as a virus, it is necessary to provide in *trans* the missing viral genes. BOSC cells are designed to provide all viral proteins but not to package or transmit the RNAs encoding these functions. Virion produced by packaging cells can infect but cannot replicate further.

The generation of high titre, helper free retroviruses by transient transfection was achieved by using the highly transfectable 293T cell line into which are stably introduced constructs that express retroviral packaging functions. The resulting ecotropic virus packaging cell line BOSC 23 produces infectious retrovirus at $> 10^6$ infectious units /ml of supernatant within 72 hours after calcium phosphate mediated transfection (Pear et al., 1993).

The BOSC cells contain the *gag*, *pol* and *env* retroviral genes and are therefore capable of producing retrovirus from the experimental constructs on their own. However, there is probably only a single copy of each gene per BOSC cell and therefore greater titres can be achieved by co-transfecting pHIT plasmids which provide further copies of the *gag*, *pol* and *env* genes.

The retroviral vectors being used in this project were co-transfected with packaging vectors pHIT 60 and pHIT 123. These vectors encode MoMLV *gag/gag-pol* and the MoMLV env gene respectively (see Appendix I for plasmid diagrams). The pHIT 60 plasmid contains the murine leukemia virus gag/pol cassette under the control of the human cytomegalovirus immediate early promoter, whereas pHIT 123 contains the human cytomegalovirus, that is to say driven murine leukemia virus ecotropic envelope.

Virion supernatant was harvested 48 hours and 72 hours post-transfection to ensure the expression of the various genes and therefore optimal retrieval of virus particles.

3.5. INTRODUCING THE RETROVIRAL VECTORS USED IN THIS PROJECT

At the outset, this project consisted of studying four retroviral vectors, two shuttle and two splice acceptor vectors.

3.5.1. Shuttle Series

This series of retroviral vector is rather special in that they contain a plasmid backbone within the LTRs of the construct. This enables plasmid rescue when the provirus is incorporated into genomic DNA (gDNA).

It is possible to retrieve shuttle vectors from genomic DNA (gDNA) with flanking regions of the endogenous trapped gene attached. It is necessary to digest the gDNA at sites that are not present in the vector (section 2.2.2.xii). This means the fragment that contains the provirus and plasmid backbone will be flanked by 5' and 3' regions at the site of insertion. Alternatively, the gDNA can be cut with an enzyme that cuts within the vector as long as the site is beyond the origin and ampicillin, then any rescued plasmid will contain only the 5' flank. This avoids the confusion of having the 5' flank fused to the 3' flank in the rescued plasmid, but has the drawback of only providing information of one flanking region.

3.5.1.a. Proviral pSHUT+124K SD126-1

This retroviral vector would be expected to insert near to an exon due to the lack of a splice acceptor site. Activation would require either in-frame exon integration or integration in 5′ UTR of an actively expressed gene. With there being no polyA site after the *neo* cassette, a polyA signal from an endogenous gene would be required to generate stable *neo* mRNA and in turn, G418 resistant clones. Suggesting that only insertions in genes should generate G418 resistant clones and background intergenic insertions should be lost.

Due to these requirements for activation it is expected that this construct would appear inefficient with the number of G418 resistant colonies being low (Niwa et al., 1993; Salminen et al., 1998; Yoshida et al., 1995; Zambrowicz et al., 1998).

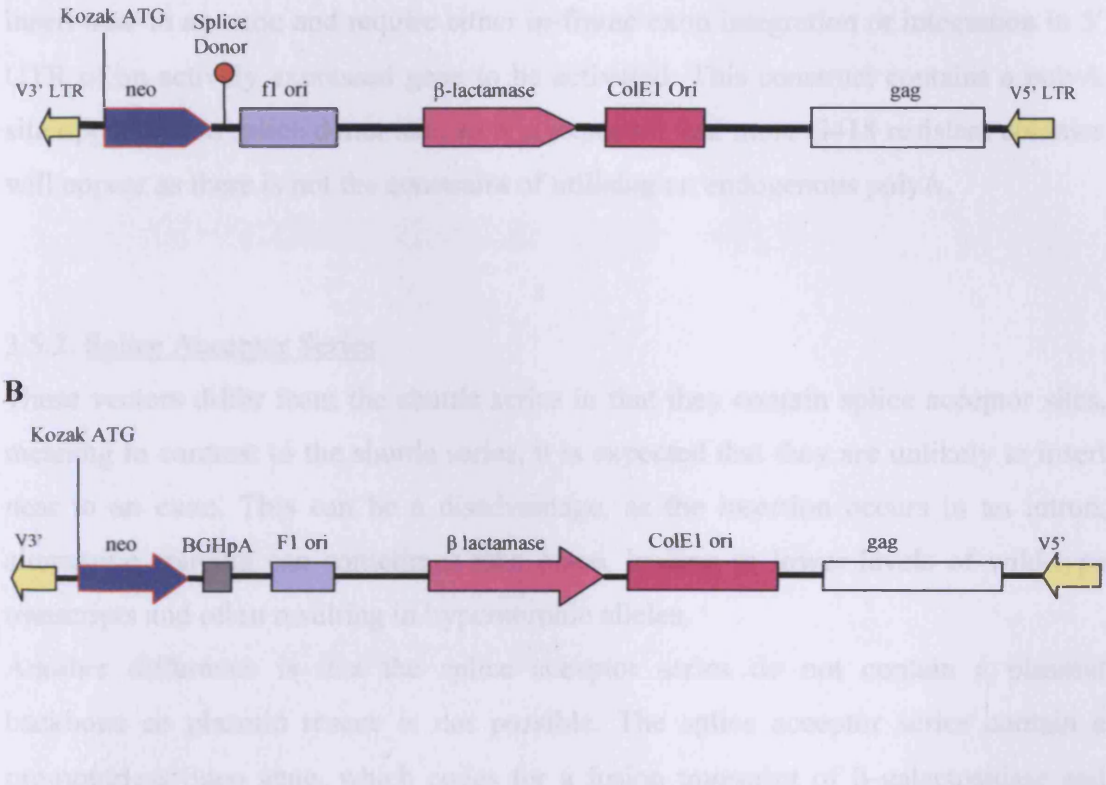


Figure 3.1. Schematic representation of the shuttle retroviral vectors used in this study (supplied by John Dixon, CRC Institute, Cambridge).
(**A**) Proviral pSHUT+124K SD126-1 and (**B**) Proviral pSHUT+124K-A.
Abbreviations: Neo, *E. coli* neomycin phosphotransferase gene; Splice Donor, mouse *Pax-2* splice donor site; F1 ori, filamentous phage origin of replication which produces ssDNA; β lactamase, *E. coli* TEM-7 beta lactamase gene; ColE1 Ori, bacterial origin of replication; gag, Moloney murine leukemia virus gag gene; BGH pA, transcription termination and polyadenylation region from the bovine growth hormone gene; Koazak ATG, sequence is contiguous with ATG start codon, greatly increases the efficiency of translation by slowing down the rate of scanning by ribosome and improving chance of recognising start of translation at ATG start codon.

3.5.1.b. Proviral pSHUT+124K-A

As with the above retroviral vector (3.5.1.a.), this construct would also be expected to insert near to an exon and require either in-frame exon integration or integration in 5′ UTR of an actively expressed gene to be activated. This construct contains a polyA site opposed to a splice donor site, so it is expected that more G418 resistant colonies will appear as there is not the constraint of utilising an endogenous polyA.

3.5.2. Splice Acceptor Series

These vectors differ from the shuttle series in that they contain splice acceptor sites, meaning in contrast to the shuttle series, it is expected that they are unlikely to insert near to an exon. This can be a disadvantage, as the insertion occurs in an intron, alternative splicing can sometimes take place, leading to lower levels of wild-type transcripts and often resulting in hypomorphic alleles.

Another difference is that the splice acceptor series do not contain a plasmid backbone so plasmid rescue is not possible. The splice acceptor series contain a promoterless βgeo gene, which codes for a fusion transcript of β-galactosidase and *neomycin phosphotransferase*. As expression of this gene relies on the activity of the trapped promoter, all G418 resistant clones represent gene trap events.

Drug selection identifies cells that have integrated the vector, in the correct orientation, into an intron of an expressed gene such that the reporter can be spliced in-frame into the gene transcript. This transcript encodes a fusion protein possessing both neomycin phosphotransferase and β-galactosidase (*lacZ*/β-gal) activities. (Friedrich and Soriano, 1991) ES cell clones can be selected using G418 and the trapped gene can be studied by staining for β-galactosidase activity.

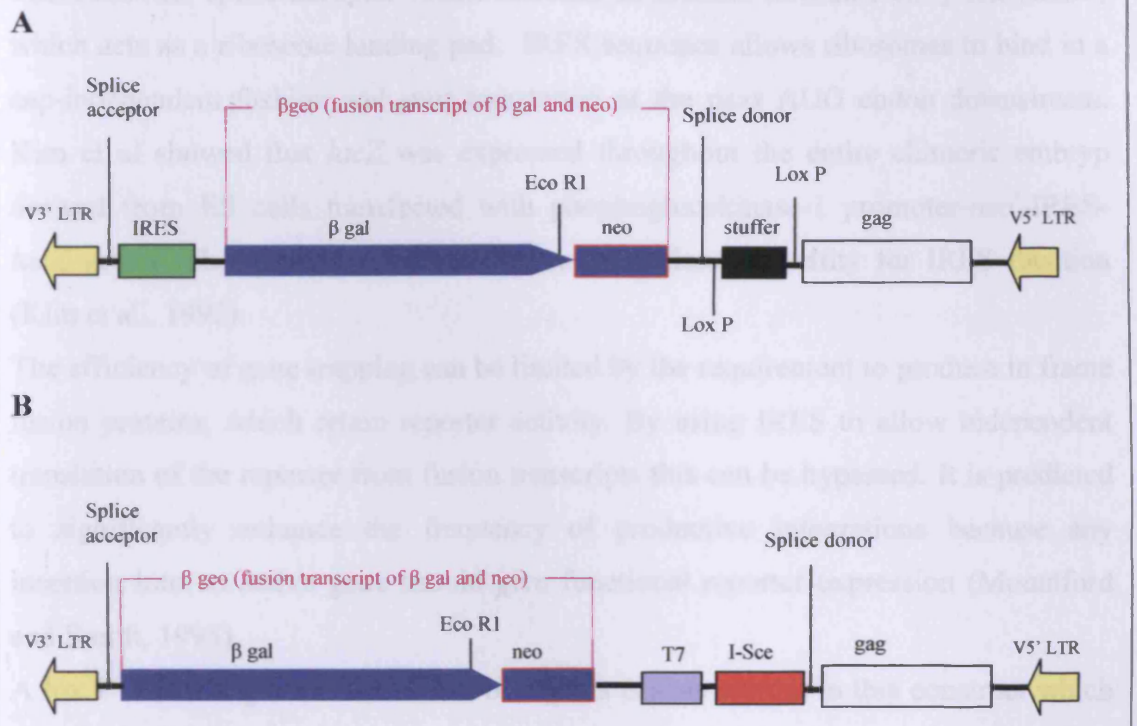


Figure 3.2. Schematic representation of the splice acceptor retroviral vectors used in this study (supplied by John Dixon, CRC Institute, Cambridge.
(**A**) Proviral pCSI+SD126 -2 and (**B**) Proviral pCRR+SD126 T7 I-Sce-X.
Abbreviations: Neo, *E. coli* neomycin phosphotransferase gene; Splice Acceptor, adenovirus major late transcript splice acceptor sequence; Splice Donor, mouse *Pax-2* splice donor site; IRES, encephalomyocarditis virus internal ribosome entry site; βgeo, a fusion gene between the β-galactosidase (β-gal) and neomycin phosphotransferase (neo) genes; Lox P, lox-cre site specific recombination system of bacteriophage P1; gag, Moloney Murine Leukemia Virus gag gene; T7, bacteriophage T7 promoter; I-Sce, intron encoded endonuclease from the mitochondria of *Saccharomyces cerevisiae*.

3.5.2.a. Proviral pCSI+SD126 L4-2

This retroviral splice acceptor vector contains an internal ribosome entry site (IRES) which acts as a ribosome landing pad. IRES sequence allows ribosomes to bind in a cap-independent fashion and start translation at the next AUG codon downstream. Kim et al showed that *lacZ* was expressed throughout the entire chimeric embryp derived from ES cells transfected with phophoglucokinase-1 promoter-*neo*^{r}-IRES-*lacZ* vector. The result suggests that there is no tissue specifity for IRES function (Kim et al., 1992).

The efficiency of gene trapping can be limited by the requirement to produce in frame fusion proteins, which retain reporter activity. By using IRES to allow independent translation of the reporter from fusion transcripts this can be bypassed. It is predicted to significantly enhance the frequency of productive integrations because any insertion into an active gene should give functional reporter expression (Mountford and Smith, 1995).

A lox P stuffer region of about 330 base pairs is also present in this construct which when trapped lines are crossed through a cre expressing strain should be lost, aiding the identification of heterozygous and homozygous progeny.

3.5.2.b. Proviral pCRR+SD126 T7 I-Sce-X

This retroviral splice acceptor vector also contains a T7 promoter just upstream of the splice donor site, which should allow an enrichment step (section 2.4.1.) to be carried out after cDNA synthesis. In theory the T7 RNA polymerase should copy more RNA from the promoter through the splice donor and trapped transcript to the poly A and appended anchor sequences. By performing 3′ RACE on this enriched template it would be more likely that the correct region be amplified.

This construct also contains an I-Sce site included in the T7 oligo. It is expected that this will enable easy retargeting of the same locus. This could be useful to remove an exon and create a definite null or even replace *lacZ* with GFP.

The theory is that I-Sce is an extremely rare restriction enzyme that cuts an 18 base pair site, which is more than likely absent from the mouse genome, unless the trap vector has been integrated. To retarget this locus, a normal targeting or expression

vector could be electroporated along with the I-Sce1 enzyme. This enzyme causes a double strand break, which triggers repair enzymes and this localised recruitment of repair enzymes greatly increases homologous recombination at the trapped locus. This would allow recombinase-mediated, post-insertional modifications of the gene trap locus.

This "knock-in" strategy permits additional modifications to be made to a trapped locus, such as allowing the promoter elements of the trapped gene to drive the expression of a knocked-in transgene for use in rescue or cell-labelling experiments. In addition, this approach can also be used to generate an allelic series by recombining in different mutant cDNAs of the trapped gene.

Another example (Westerman and Leboulch, 1996) is inserting simian virus 40 large T (SV40T) antigen into a locus to target the lineages that express the trapped gene for immortalization.

3.6. ANNOTATED SEQUENCE OF SHUTLLE AND SPLICE ACCEPTOR CONSTRUCT

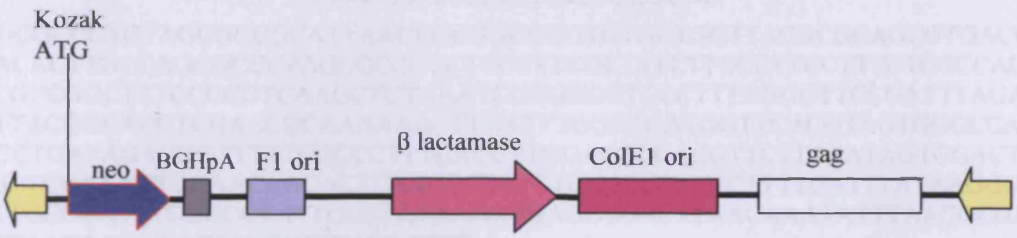
It was felt it would be prudent to try and confirm sequence information for one of each type of vector being studied.

Proviral pSHUT+124K (A) and proviral pCRR+SD126 T7 I-Sce (X) were selected.

Sequencing reactions were conducted commercially (Oswell Research Products Ltd, University of Southampton, UK) and contiguous sequence prepared using CAP EST Assembler at IFOM (section 2.4.5.iii.).

See the following, figure 3.3a annotated sequence of Proviral pSHUT+124K(A) and figure 3.3b annotated sequence of Proviral pCRR+SD126 T7 I-Sce(X)

Figure 3.3.a. - Annotated sequence of proviral pSHUT+124K (A)



NB. Arrows indicate the direction of transcription

5'MoMLV LTR 1-248bp [Shinnick, 1981 #335]

5'AATGAAAGACCCCCGCTGACGGGTAGTCAATCACTCAGAGGAGACCCTCCCAAGGAAC
AGCGAGACCACAAGTCGGATGCAACTGCAAGAGGGTTTATTGGATACACGGGTACCCGGG
CGACTCAGTCAATCGGAGGACTGGCGCCCCGAGTGAGGGGTTGTGCTAGCTTGCCAAACC
TACAGGTGGGGTCTTTCATTCCCCCCTTTTTCTGGAGACTAAATAAAATCTTTTATTTTATC
GATAAGC

Kozak ATG 249-257bp [Kozak, 1987 #230]

Neomycin Phosphotransferase 258-1051bp [Beck, 1982 #336]

CCACCATGGGAATTGAACAAGACGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGA
GGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCG
GCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAAT
GAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCA
GCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCG
GGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATG
CAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAGCGAAACA
TCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGA
CGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCC
CGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGA
AAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAG
GACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGC
TTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCT
TGACGAGTTCTTCTGA

STOP

TCTAGAGAGCTCAGCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG
CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAA

BGHpA 1162-1258bp

ATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGG
GGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAG

STOP

CAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGTCGACGAA
TTCGAGCTCGGCCGACTTGGCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCC
GTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAC
ACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCA
ACAGTTGCGCAGCCTGAATGGCGAATG

Phage F1 Origin 1528-1985bp

GACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACC
GCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCAC
GTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGA
GCTTTACGGCACCTCGACCGCAAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCAT
CGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACT
CTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGG
ATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAATATTTAACGCGA
ATTTTAACAAAATATTAACGTTTACAATTT

CGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCA
CTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACC
CGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACC
GTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGA
AAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGA
CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAAT
ACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGA
AAAAGGAAGAGT

Beta lactamase 2424-3285bp [Kalnins, 1983 #326]

ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTT
TTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGA
GTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAA
GAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTCATACACTATTATCCCGTAT
TGACGCCGGGCAAGAGCAACTCGGTCGCCGGGCGCGGTATTCTCAGAATGACTTGGTTGA
GTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAG
TGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGG
ACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGT
TGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGT
AGCAATGCCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC
CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTA
TCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGG
GGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGA
TTAAGCATTGGTAA STOP

Col. E1 Origin 3286-4227bp

CTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAA
AAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTT
TCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTT
TTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTT
GCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGAT
ACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA
CCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCT
GAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGAT
ACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGG
TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA
CGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGT
GATGCTCGTCAGGGGGGCGGAGCCTATCGAAAAACGCCAGCAACGCGGCCTTTTTACGGT
TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGC
GCAGCGAGTCAGTGAGCGAGGAAGCGGAAGA

GCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCAC
GACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTC
ACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGT
GAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTATTTAG
GTGACACTATAGAATACTCAAGCTTATGCATGCGGCCGCATCTAGCGAATTCG

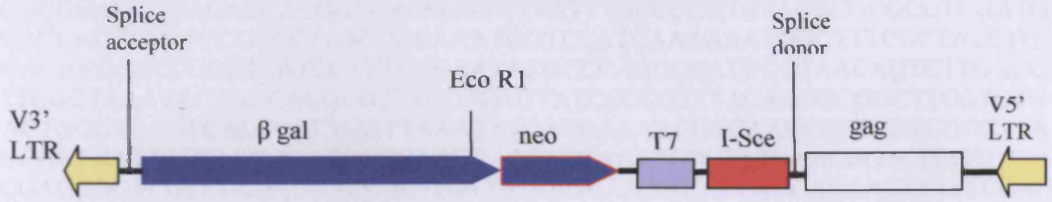
Truncated gag gene 4525-5404bp [Leis, 1988 #327] with Psi+ extended viral packaging signal 4984–5404bp [Alford, 1991 #334]

GCGCCTAGAGAAGGAGTGAGGGCTGGATAAAGGGAGGATCGAGGCGGGGTCGAACGAGG
AGGTTCAAGGGGGAGAGACGGGGCGGATGGAGGAAGAGGAGGCGGAGGCTTAGGGTGT
ACAAAGGGCTTGACCCAGGGAGGGGGGTCAAAAGCCAAGGCTTCCCAGGTCACGATGTA
GGGGACCTGGTCTGGGTGTCCATGCGGGCCAGGTGAAAAGACCTTGATCTTAACCTGGGT
GATGAGGTCTCGGTTAAAGGTGCCGTCTCGCGGCCATCCGACGTTAAAGGTTGGCCATTCT
GCAGAGCAGAAGGTAACCCAACGTCTCTTCTTGACATCTACCGACTGGTTGTGAGCGATC
CGCTCGACATCTTTCCAGTGACCTAAGGTCAAACTTAAGGGAGTGGTAACAGTCTGGCCC
ATATTCTCAGACAAATACAGAAACACAGTCAGACAGAGACAACACAGAACGATGCTGCA
GCAGACAAGACGCGCGGCGCGGCTTCGGTCCCAAACCGAAAGCAAAAATTCAGACGGAG
GCGGGAACTGTTTTAGGTTCTCGTCTCCTACCAGAACCACATATCCCTCCTCTAAGGGGGG
TGCACCAAAGAGTCCAAAACGATCGGGATTTTTGGACTCAGGTCGGGCCACAAAAACGGC
CCCCGAAGTCCCTGGGACGTCTCCCAGGGTTGCGGCCGGGTGTTCCGAACTCGTCAGTTCC
ACCACGGGTCCGCCAGATACAGAGCTAGTTAGCTAACTAGTACCGACGCAGGCGCATAAA
ATCAGTCATAGACACTAGACAATCGGACAGACACAGATAAGTTGCTGGCCAGCTTACCTC
CCGGTGGTGGGTCGGTGGTCCCTGGGCAGGGGTCTCCCG

MoMLV V5′ LTR 5405 – 5662bp [Shinnick, 1981 #335]

ATCCCGGACGAGCCCCCAAATGAAAGACCCCCGCTGACGGGTAGTCAATCACTCAGAGGA
GACCCTCCCAAGGAACAGCGAGACCACAAGTCGGATGCAACTGCAAGAGGGTTTATTGGA
TACACGGGTACCCGGGCGACTCAGTCAATCGGAGGACTGGCGCCCCGAGTGAGGGGTTGT
GCTAGATAACTTCGTATAGTATACATTATACGAAGTTATCTAGCTTGCCAAACCTACAGGT
GGGGTCTTTCATT 3′

Figure 3.3.b. – Annotated sequence of Proviral pCRR+SD126T7 I-Sce (X)



NB. Arrows indicate the direction of transcription

MoMLV V3′ LTR 1-341bp [Shinnick, 1981 #335]

5′AATGAAAGACCCCCGCTGACGGGTAGTCAATCACTCAGAGGAGACCCTCCCAAGGAAC
AGCGAGACCACAAGTCGGATGCAACTGCAAGAGGGTTTATTGGATACACGGGTACCCGGG
CGACTCAGTCAATCGGAGGACTGGCGCCCCGAGTGAGGGGTTGTGCTAGCTTGCCAAACC
TACAGGTGGGGTCTTTCATTCCCCCCTTTTTCTGGAGACTAAATAAAATCTTTTATTTTATC
TATGGCTCGTACTCTATAGGCTTCAGCTGGTGATATTGTTGAGTCAAAACTAGAGCCTGGA
CCACTGATATCCTGTCTTTAACAAATTGGACTAATCGAT

CTCGACCTCGAGACTAGTGGATCCCCCGGGCTGCAGATCTGTAGGGCGCAGTAGTCCAGG
GTTTTCCTTGATGATGTCATACTTATCCCTGTCCCTTTTTTTTCCAC

Splice Acceptor
449-450bp

AGCTCGCGGTTGAGGACAAACTCTTCGCGGTCT
TTCCAGTGGGGATCGACGGTATCGATAAGCTTGATGATCTGTGAC

Beta Geo with ATG 527-4409bp [Soriano, 1991 #325]

ATGGCGGATCCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTA
ATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCG
ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGC
ACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGT
CGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTAACCTAT
CCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCA
CATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCG
TTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTCGGTTACGGCCAGGACAGTC
GTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCGCCTCGCGGT
GATGGTGCTGCGTTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAG
CGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCAT
GTTGCCACTCGCTTTAATGATGATTTCAGCCGCGCTGTACTGGAGGCTGAAGTTCAGATGT
GCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACGCAGG
TCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATGCCG
ATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGA
ATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAAG
CCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACGGCA
AGCCGTTGCTGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCAT
GGATGAGCAGACGATGGTGCAGGATATCCTGCTGATGAAGCAGAACAACTTTAACGCCGT
GCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGCTACGGCCTG
TATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACC
GATGATCCGCGCTGGCTACCGGCGATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGAT
CGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAAT

CACGACGCGCTGTATCGCTGGATCAAATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAAG
GCGGCGGAGCCGACACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGGATG
AAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTG
GAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAACAGTCTTGGCG
GTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTG
GGACTGGGTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTGGTCGGCTTA
CGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTCTTT
GCCGACCGCACGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTCCAG
TTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATA
ACGAGCTCCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGC
CTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGAACTACCGCAGCCGG
AGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGT
CAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAGTGTGA
CGCTCCCCGCCGCGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCAT
CGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTGG
ATTGGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTG
GATAACGACATTGGCGTAAGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGAACGC
TGGAAGGCGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCAGATACA
CTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTA
TTTATCAGCCGGAAAACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGAT
GTTGAAGTGGCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCG
CAGGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACTATCCCGACCG
CCTTACTGCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATACCCCGTAC
GTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATTATGGCCCACAC
CAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCGCTACAGTCAACAGCAACTGATGGAA
ACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTC
CATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGCGGAATTCCAGCTG
AGCGCCGGTCGCTACCATTACCAGTTGGTCTGGTGTCAGGGGATCCCCCGGGCTGCAGCC
AATATGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTG
GAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTG
TTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCC
TGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTT
GCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAG
TGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGC
TGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGC
GAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGA
TCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCG
CATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCAT
GGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCG
CTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGC
TGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATC
GCCTTCTTGACGAGTTCTTCTGA

STOP

→

TAATACGACTCACTATAGGGCGATCGAG

T7 Promoter Site
4410-4426bp

TTACGCTAGGGATAACAGGGTAATATAGTCTA

I-Sce1 Restriction Site
4438-4461bp

Splice Donor
4476-4477bp

GAACTAGTGGATCCAGCTTATCGAGCTTGCCAACTACAGCAGGTAAGCATATTTATGAAA
AATGCTGATTGTGTTAGCTACTTGTGTCAGTGTTGTGATAGTCGAC

Truncated GAG gene 4576-5455bp[Leis, 1988 #327] with Psi+ extended viral packaging signal 5035-5455bp [Alford, 1991 #334]

←

GGCGCCTAGAGAAGGAGTGAGGGCTGGATAAAGGGAGGATCGAGGCGGGGTCGAACGAG
GAGGTTCAAGGGGGAGAGACGGGGCGGATGGAGGAAGAGGAGGCGGAGGCTTAGGGTG
TACAAAGGGCTTGACCCAGGGAGGGGGGTCAAAAGCCAAGGCTTCCCAGGTCACGATGT
AGGGGACCTGGTCTGGGTGTCCATGCGGGCCAGGTGAAAAGACCTTGATCTTAACCTGGG
TGATGAGGTCTCGGTTAAAGGTGCCGTCTCGCGGCCATCCGACGTTAAAGGTTGGCCATTC
TGCAGAGCAGAAGGTAACCCAACGTCTCTTCTTGACATCTACCGACTGGTTGTGAGCGATC
CGCTCGACATCTTTCCAGTGACCTAAGGTCAAACTTAAGGGAGTGGTAACAGTCTGGCCC
ATATTCTCAGACAAATACAGAAACACAGTCAGACAGAGACAACACAGAACGATGCTGCA
GCAGACAAGACGCGCGGCGCGGCTTCGGTCCCAAACCGAAAGCAAAAATTCAGACGGA
GGCGGGAACTGTTTTAGGTTCTCGTCTCCTACCAGAACCACATATCCCTCCTCTAAGGGGG
GTGCACCAAAGAGTCCAAAACGATCGGGATTTTTGGACTCAGGTCGGGCCACAAAAACGG
CCCCCGAAGTCCCTGGGACGTCTCCCAGGGTTGCGGCCGGGTGTTCCGAACTCGTCAGTTC
CACCACGGGTCCGCCAGATACAGAGCTAGTTAGCTAACTAGTACCGACGCAGGCGCATAA
AATCAGTCATAGACACTAGACAATCGGACAGACACAGATAAGTTGCTGGCCAGCTTACCT
CCCGGTGGTGGGTCGGTGGTCCCTGGGCAGGGGTCTCC CG

MoMLV V5′ LTR 5456-5699bp[Shinnick, 1981 #335]

←

ATCCCGGACGAGCCCCCAAATGAAAGACCCCCGCTGACGGGTAGTCAATCACTCAGAGGA
GACCCTCCCAAGGAACAGCGAGACCACAAGTCGGATGCAACTGCAAGAGGGTTTATTGGA
TACACGGGTACCCGGGCGACTCAGTCAATCGGAGGACTGGCGCCCCGAGTGAGGGGTTGT
GCTAGCTAGCTTGCCAAACCTACAGGTGGGGTCTTTCATT 3′

3.7. TRANSFECTION OF RETROVIRAL VECTORS INTO ECOTROPIC BOSC PACKAGING CELLS

All retroviral vectors were transfected into BOSC packaging cells using both lipofectamine (2.1.4.i.) and calcium phosphate (2.1.4.ii.) methods.

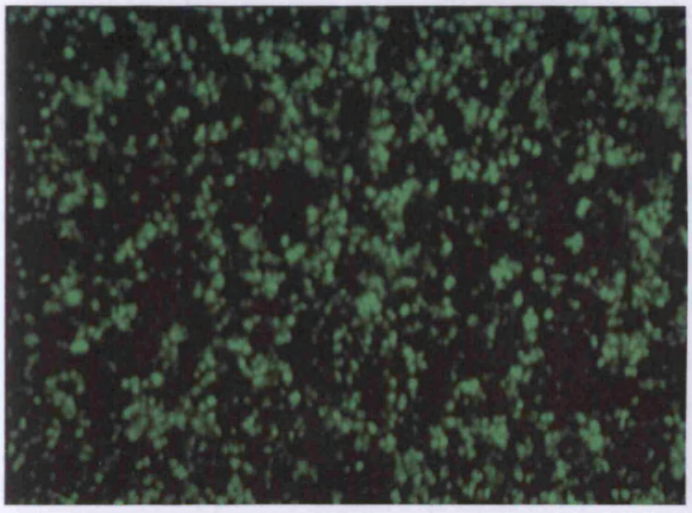
3.7.1. pCX-GFP Plasmid

Originally the method being used was lipofectamine and the retroviral vectors were being co-transfected along with pHIT 60, pHIT 123 and the pCX-GFP plasmids. Instead of transfecting four plasmids at once and therefore reducing the total amount of experimental vector DNA that could be transfected on each occasion, it was decided to introduce the pCX-GFP plasmid separately. This GFP reaction could act as a positive control with the same conditions being used for all transfection reactions.

Began using the GFP positive control when first starting to use the calcium phosphate method ($CaPO_4$ 1) and obtained a carpet of green fluorescent BOSC cells showing a successful transfection (see figure 3.4.).

To ascertain a quantitative value for transfection success, focused on an area of cells and then turned on the UV light, by contrasting the green with the non-green it was possible to estimate a percentage value for the number of cells which had been successfully transfected. It can be said, that by eye the number of cells showing green and therefore successful transfection, was in the region of 70-80%.

Figure 3.4. Fluorescent image (5X magnification) showing the successful transfection ($CaPO_4$ 1) of BOSC cells with pCX-GFP.



3.7.2. Harvested Virion

Another consideration and subsequent amendment was to use the harvested virion stocks immediately rather than storing them at -80°C. This meant that the ES cells to be used for infection needed to be prepared the day before. It was necessary to split the cells the day prior to infection, as the cells need to be actively dividing to allow the pre-integration complexes to enter the nuclei and integrate into the genome. Also, trypsinisation would damage all the membrane bound proteins, including the receptor which the Moloney hijacks, gp70, so it was essential to allow time for recovery before attempting to infect the ES cells.

3.8. EMBRYONIC STEM CELLS

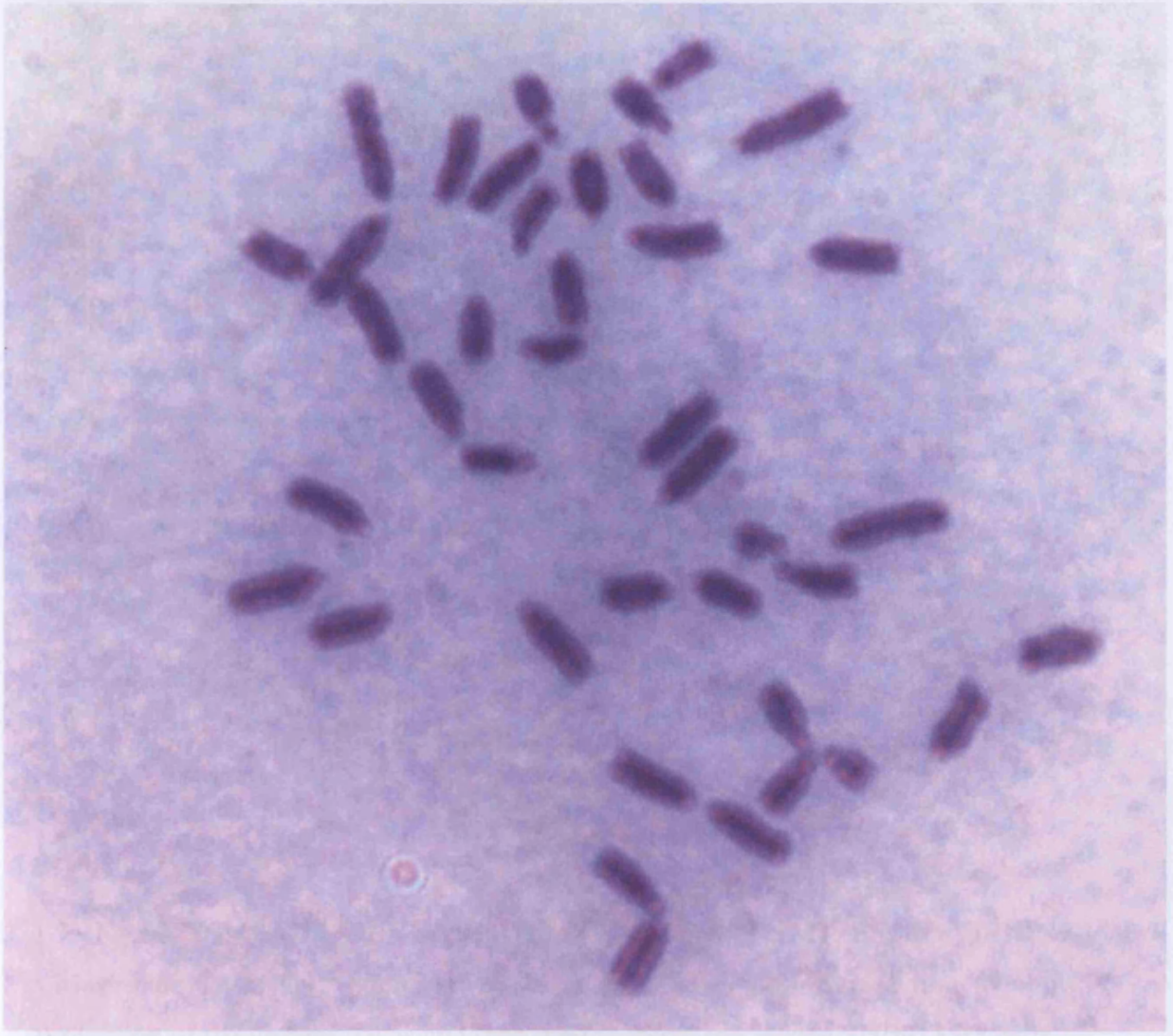
Murine embryonic stem (ES) cells are undifferentiated, pluripotent stem cells derived from blastocysts.(Evans and Hunter, 2002) These cells undergo unlimited self-renewal in the presence of the cytokine leukaemia inhibitiry factor (LIF) while retaining multilineage differentiation capacity and may be kept entirely karyotypically normal (Evans and Kaufman, 1981; Martin, 1981). They retain the ability to participate in normal embryonic development and following reintroduction to the blastocyst, they generate chimaeric animals that are mosaic in all their tissues. Mosaicism extends to the germ cell lineage and ES cells can contribute fully functional gametes (Bradley et al., 1984; Robertson et al., 1986).

3.8.1. -Karyotyping

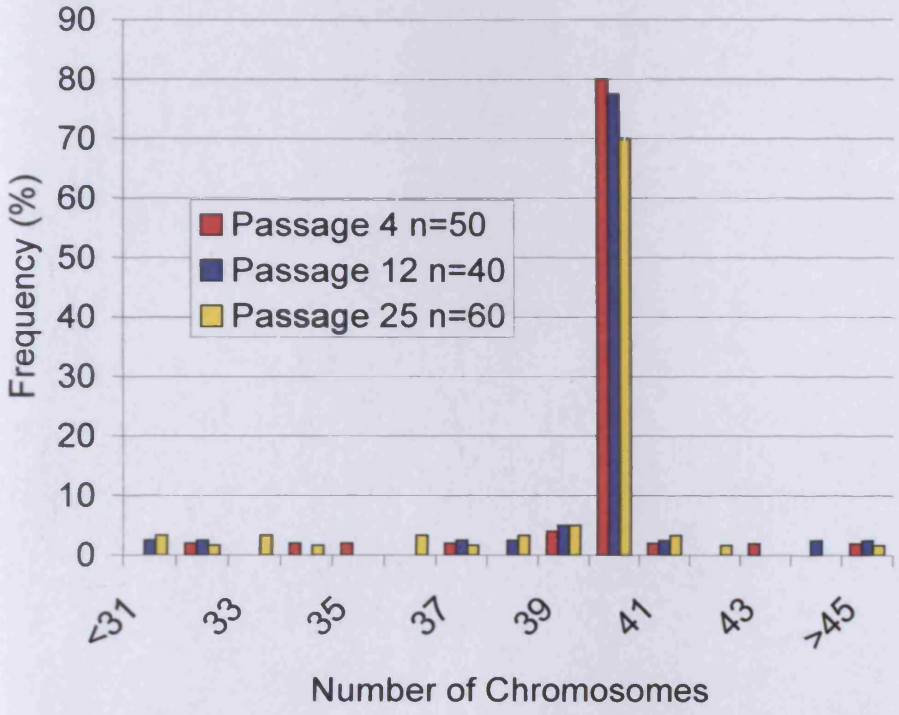
The ES cell line used in this project was derived from C57BL/6J blastocysts, are male and have maintained a normal karyotype in culture.

The karyotype (2.1.8.) of these cells was studied at passage 4, 12 and 25 by examining metaphase spreads (Fig. 3.5.), a slight decrease was detected in the frequency of spreads with the normal number of chromosomes, but the modal number was still 40 (fig. 3.6.).

Figure 3.5. Metaphase spread of the SHBl6.3 cell line (isolated by Susan Hunter, Cardiff University, Wales), which was found to have a euploid chromosome complement with a strong modal number of 40.



Graph 3.1. Distribution of metaphase spreads with different chromosome numbers at passage 4, 12 and 25 of SHBl6.3 ES cell line



3.9. INFECTION OF EMBRYONIC STEM CELLS WITH HARVESTED VIRION

Virion supernatant was harvested at 48 and 72 hours post transfection to allow expression of the various genes and then used to infect pre-seeded SHBl6.3 ES cells.

3.9.1. – Changing Medium

During selection, medium was being changed on an almost daily basis. This was later amended to every 3 or 4 days. When selection begins there is a lot of cell death and therefore a lot of cell debris so at this stage medium needs to be changed daily but after this period the cells appear to prefer being left in conditioned medium. Selection lasts about 12-14 days and as there are no feeders present it is mainly necessary to

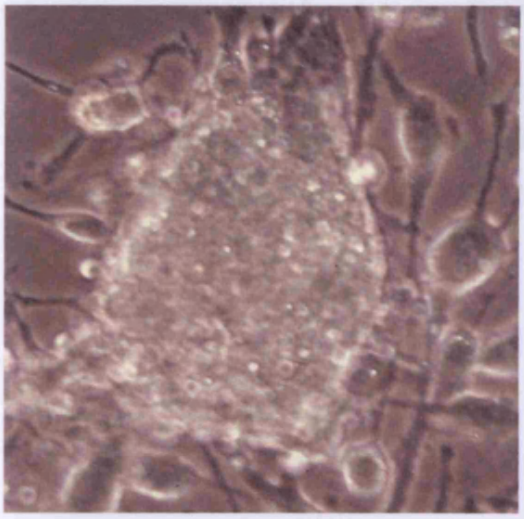
change the medium to prevent cell differentiation, by adding LIF with the medium. Stopped such frequent medium changes during the second calcium phosphate transfection ($CaPO_4$ 2).

3.9.2. – Use of Feeder Layer

Originally the ES cells were prepared with feeders for infection but this was something that was later amended. It was realised that mitotically inactivated feeder layers will mop up the retroviral particles as they can still be infected but without a division the virus cannot integrate. First amended this procedure with the third calcium phosphate transfection ($CaPO_4$ 3).

However, it was also realised that feeders needed to be added to the ES cells once the resistant colonies had been selected and picked for expansion. At this stage it seemed essential feeders be added as there were so few ES cells being transferred to the multiwell plate, the feeders seemed to promote cell attachment and viability. Amended this procedure during the sixth calcium phosphate transfection ($CaPO_4$ 6).
It seemed this was the crucial amendment to the procedure, as it was at this point ($CaPO_4$ 6) that the first G418 resistant ES cell colonies (Fig. 3.6.) survived picking and were able to expand.

Figure 3.6. G418 resistant ES cell colony on feeders (40X magnification)



3.10. COMPARING SUCCESSFUL INFECTIONS OF SHUTTLE AND SPLICE ACCEPTOR SERIES

After amending certain factors as mentioned in sections 3.7. and 3.8., transfection and subsequent infection began to give results with calcium phosphate transfection number 6 ($CaPO_4$ 6). There were four attempts at lipofectamine transfection prior to trying the calcium phosphate method. Unfortunately, both $CaPO_4$ 8 and $CaPO_4$ 10 had to be disposed of due to infection.

Table 3.1. Showing G418 resistant ES cell clones obtained for the four retroviral vectors used.

Transfection type and number of ES cells seeded for infection/construct/ time point.	Constructs transfected into BOSC				Clones obtained 48Hr				Clones obtained 72Hr			
	X	2	1	A	X	2	1	A	X	2	1	A
$CaPO_4$ 6 ($1x10^6$)	√	-	√	√	6	-	1	4	2	-	0	7
$CaPO_4$ 7 ($1x10^6$)	√	-	√	√	25	-	0	23	134	-	1	14
$CaPO_4$ 9 ($1x10^6$)	√	√	√	√	5	23	1	7	7	42	2	9
$CaPO_4$ 11 ($1x10^6$)	√	√	-	√	4	74	-	18	4	48	-	18
$CaPO_4$ 8-11 ($1x10^6$)	√	√	-	-	26	38	-	-	18	33	-	-
LIPO 11-3 ($19x10^5$)	√	√	-	√	83	99	-	26	INF	INF	-	INF

Abbreviations: **LIPO**, Lipofectamine transfection; **$CaPO_4$**, Calcium phosphate transfection; **INF**, infected cells disposed of;

X, pCRR+SD126T7 I-Sce; **2**, pCSI+SD126 L4; **1**, pSHUT+124K SD126;

A, pSHUT+124K.

It can be seen from table 3.1. that the splice acceptor series gave rise to a far greater number of resistant ES cell colonies than the shuttle series.
These results can be compared as an infection assay as opposed to an actual particle assay. By dividing the total number of $G418^R$ colonies obtained for each construct, by the number of successful infections, shows the number of ES cells infected with each retroviral construct per infection. Only the successful infections were included, as otherwise the result would be affected unfairly, as on some occasions ES cells had to be disposed of due to bacterial/fungal infections.

It must also be noted that only the calcium phosphate transfections are being counted and compared as the lipofectamine transfection had a different number of ES cells seeded and a different amount of DNA transfected, so the lipofectamine result stands alone. It is assumed that as the same amount of DNA was transfected for each calcium phosphate transfection and the same number of cells were seeded it is fair to compare the number of $G418^R$ colonies obtained for each construct (table 3.2.).

The shuttle series lacks splice acceptor sequence and therefore would be activated mostly from insertion into exons, unlike the splice acceptor series which could be activated from intronic insertions also. This means the splice acceptor series have a much greater number of productive sites for infection compared to the shuttle series.
The one splice acceptor vector requires in-frame fusion and at least low level gene expression of the trapped locus in undifferentiated ES cells for the reporter gene to be translated and subsequent survival of G418 selection. The other splice acceptor vector has an IRES upstream of the β-geo reporter thereby allowing cap-independent translation of β-geo from fusion transcripts. Another factor to consider is that as all of these retroviral gene trap vectors require a cellular promoter for activation, the maximum number of genomic targets equals the number of expressed genes.

It has subsequently been noted that the use of a positive control retroviral vector may have been useful in comparing successful infections and the numbers of $G418^R$ colonies obtained. Perhaps the ROSAβGEO retroviral vector could have been used, without any modifications to then give a baseline figure for $G418^R$ colonies.

Table 3.2. Shows the number of G418 resistant ES cell colonies obtained for each retroviral construct

RETROVIRAL CONSTRUCT	TOTAL NUMBER OF $G418^R$ ES CELL COLONIES PICKED	AVERAGE NUMBER OF $G418^R$ ES CELL COLONIES/ INFECTION
PSHUT+124K SD126-1	5	<1
pSHUT+124K-A	100	12.5
PCRR+SD126 T7 I-Sce -X	231	23.1
pCSI+SD126 L4 -2	258	43

The above table 3.2. shows the average number of $G418^R$ colonies obtained after $CaPO_4$ transfection of the retroviral constructs into BOSC packaging cells and subsequent infection of the ES cells. These figures can be compared to an expected return of about $5x10^3$ $G418^R$ colonies per 10^6 ES cells infected. This figure is assumed after comparing the number of expected active genes to the size of the genome and the fact that only about a third of the 10^6 ES cells can be expected to be infected and integration occur.

It can be seen that the number of $G418^R$ colonies achieved in this project is extremely lacking when compared to the expected numbers. This may be due to some extent to a lack of experience but it must also be mentioned that the numbers quoted are actually for the number of $G418^R$ colonies picked as opposed to the number that appeared. Perhaps speed was an issue as the longer the plates are kept out for the picking of the colonies the more likely it is that a number are lost due to inhospitable conditions.

Due to resistant clones being obtained, a reverse transcriptase assay was never attempted. With hindsight however, it does appear that a viral assay would have been beneficial to allow a value to be applied to the number of $G418^R$ colonies obtained compared to the amount of active virus present during infection. This would also have been useful to compare with the suggested positive control above, ROSAβGEO, to see if these novel retroviral vectors allowed packaging to the same standard as established retroviral gene trap constructs.

3.10.1. -pSHUT+124K SD126-1

This shuttle series vector gave less than one G418 resistant ($G418^R$) colony per infection (table 3.2.). Even though a low number was expected (3.5.1.a.), this result was not acceptable.

A reverse transcriptase assay was considered but as $G418^R$ colonies appeared for the other retroviral vectors, this eliminated the concern that packaging may have failed.
After conducting many restriction digests on pSHUT+124K SD126-1 and obtaining sequence data, it became apparent that the *neomycin phosphotransferase* (*neo*) cassette was in the wrong position. Instead of being located at the 5′ end, upstream of the splice donor it was found to be 3′ of the bacterial origin of replication (see figure 3.7.). This could explain the lack of $G418^R$ colonies as it is likely that the *neo* cassette would be spliced out after integration, due to the splice donor now located upstream of the neo cassette.
Due to time constraints, no attempt was made to correct this vector and efforts for plasmid rescue were focused on the remaining shuttle vector, pSHUT+124K-A.

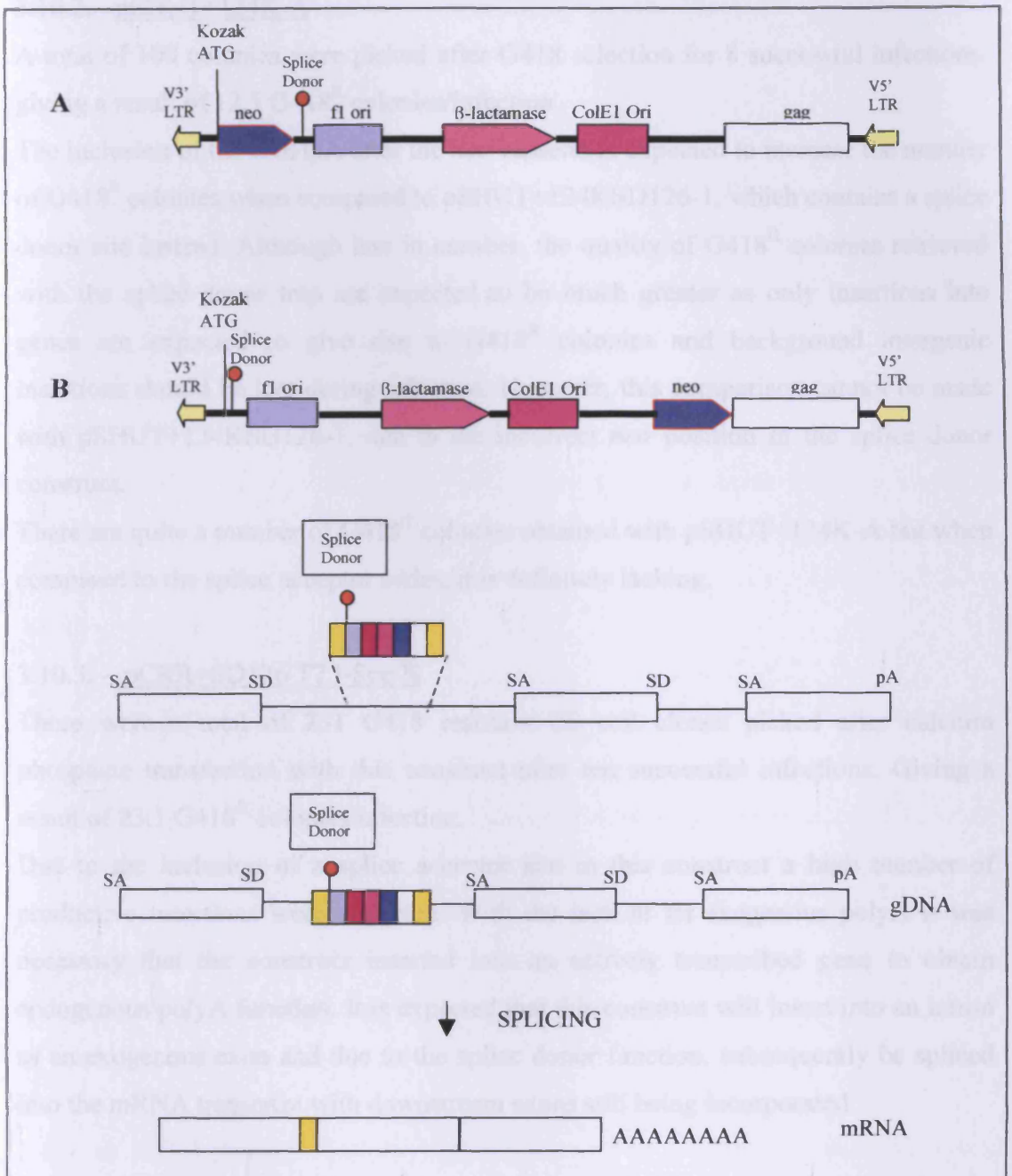


Figure 3.7. Shows the incorrect location of the *neo* cassette in the plasmid rescue construct pSHUT+124K SD126-1 and the possible splicing event after integration.
(**A**) pSHUT+124K SD126-1 with the *neo* cassette in the correct location.
(**B**) pSHUT+124K SD126-1 with the *neo* cassette in the wrong position, 3′ of the bacterial origin of replication. The diagram indicates how it is possible for the *neo* cassette to be spliced out after integration into genomic DNA.

3.10.2. - pSHUT+124K-A

A total of 100 colonies were picked after G418 selection for 8 successful infections, giving a result of 12.5 $G418^R$ colonies/infection.

The inclusion of the BGHpA after the *neo* cassette is expected to increase the number of $G418^R$ colonies when compared to pSHUT+124KSD126-1, which contains a splice donor site instead. Although less in number, the quality of $G418^R$ colonies retrieved with the splice donor trap are expected to be much greater as only insertions into genes are expected to give rise to $G418^R$ colonies and background intergenic insertions should be lost during selection. However, this comparison cannot be made with pSHUT+124KSD126-1, due to the incorrect *neo* position in the splice donor construct.

There are quite a number of $G418^R$ colonies obtained with pSHUT+124K-A but when compared to the splice acceptor series, it is definitely lacking.

3.10.3. – pCRR+SD126 T7 I-Sce-X

There were a total of 231 G418 resistant ES cell clones picked after calcium phosphate transfection with this construct after ten successful infections. Giving a result of 23.1 $G418^R$ colonies/infection.

Due to the inclusion of a splice acceptor site in this construct a high number of productive insertions were expected. With the lack of an exogenous polyA it was necessary that the construct inserted into an actively transcribed gene to obtain endogenous polyA function. It is expected that this construct will insert into an intron as an exogenous exon and due to the splice donor function, subsequently be spliced into the mRNA transcript with downstream exons still being incorporated.

3.10.4. – pCSI+SD126 L4-2

This construct gave 258 G418 resistant clones after six successful infections, giving an average of 43 $G418^R$ colonies/infection. It can be seen that this was the most successful construct for obtaining $G418^R$ colonies.

The efficiency of gene trapping can be limited by the requirement for the production of in-frame fusion proteins that retain reporter activity, the IRES was used to allow independent translation of the reporter from fusion transcripts. This is predicted to significantly enhance the frequency of productive integration as any insertion into an

active gene should give functional reporter expression. The IRES appears to have performed as expected.

Statistical confirmation of the IRES function using the Mann-Whitney test:

To confirm the IRES function statistically, the numbers of clones obtained from calcium phosphate transfections of both pCRR+SD126T7 I-Sce and pCSI+SD126 L4 (which contains IRES) were compared in a non-parametric, Mann Whitney test.

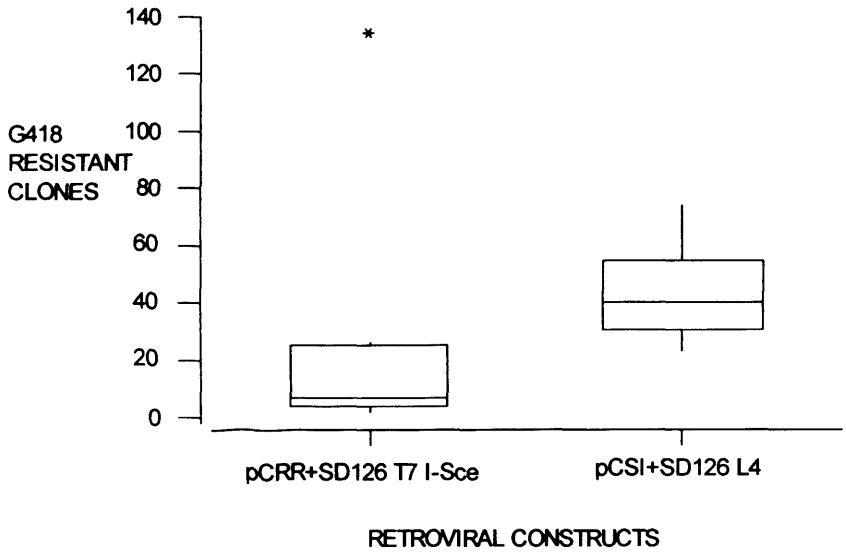
The Mann-Whitney test was used to compare two independent samples. Unlike the *t* test, the Mann-Whitney test is valid no matter what the form of the population distributions therefore, it is a distribution-free type of test. Also, as the Mann-Whitney does not focus on any particular parameter such as a mean or a median, it is a non-parametric test.

The resulting data was displayed as a box plot graph (see graph 3.2.). The interior line of the box depicts the median value. The lower quartile is found below the median and forms the bottom line of the box. The upper quartile is above the median value and forms the top line of the box. The minimum and maximum values can be read directly from the graph. If there are no outliers these numbers correspond to the ends of the whiskers of the graph and are connected to the upper and lower quartile lines.
Box plot provides a visual summary that shows the median value, spread, range and outliers.

It can be seen from graph 3.2. that pCSI+SD126 L4 gives a much increased number of $G418^R$ colonies when compared to pCRR+SD126T7 I-Sce and there is an outlie point of 134 for pCRR+SD126 T7 I-Sce.

The resulting p value was 0.0196 which confirms that these two constructs give significantly different results.

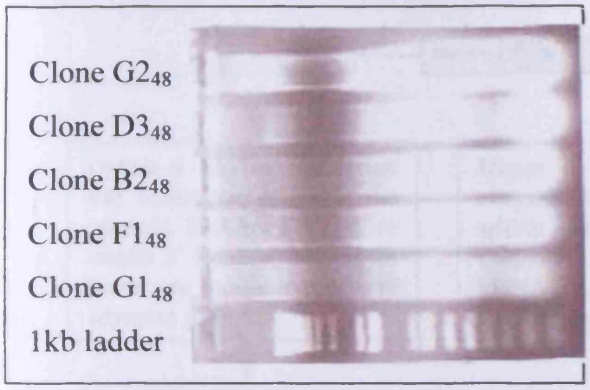
Graph 3.2. – Box plot comparing the number of G418^R colonies obtained for the two splice acceptor vectors



3.11. PLASMID RESCUE SERIES RESULTS

Genomic DNA was prepared (2.2.2.xi.) from all plasmid rescue $G418^R$ ES cell colonies. The resulting DNA was electrophoresed on a 1% agarose gel (2.2.2.iv) to ensure that it was successful (fig. 3.8.).

Figure 3.8. 1% agarose gel showing resulting isolation of genomic DNA from five $CaPO_4$11 clones.



NB: It has to be noted that subsequent investigations have led to the discovery that the genomic DNA shown to the left is not of sufficient quality for plasmid rescue procedures. The DNA needs to be treated with RNase to remove the obvious 18s and 28s ribosomal bands that can be seen as bright spots in the lower region of the gel. The genomic DNA is of considerable size and will be visible at the uppermost region of the gel.

It can be seen from table 3.3, that only a very small number of resistant bacterial colonies were actually recovered when attempting plasmid rescue.

Due to the lack of rescued plasmids it was felt necessary to check that the gene trap construct was actually present in these resistant colonies. Neomycin phosphotransferase offers G418 resistance in mammalian cells but also confers kanamycin resistance in bacteria. The 5′ flanking rescued plasmids, recovered after electroporation were spread out onto kanamycin (40μg/ml) agar plates.

Three out of four of the 5′ flanking rescued plasmids grew on kanamycin agar plates, showing that the *neo* cassette must be present. Only rescued plasmids containing the 5′ flanking region can be tested in this way as this is where the *neo* cassette would be located (fig. 3.9.).

Negative controls of no DNA and also undigested gDNA were added to the electroporation mix and used to confirm that there must be a plasmid of some description conferring antibiotic resistance to the plasmid rescue bacterial plates. No resistant colonies arose from the negative control plates at any time. A positive control was also set up, which consisted of pUC18 plasmid electroporated in exactly the same fashion as the plasmid rescued samples. Many bacterial colonies appeared for this positive control on all occasions.

Figure 3.9. Schematic outline of the plasmid rescue procedure

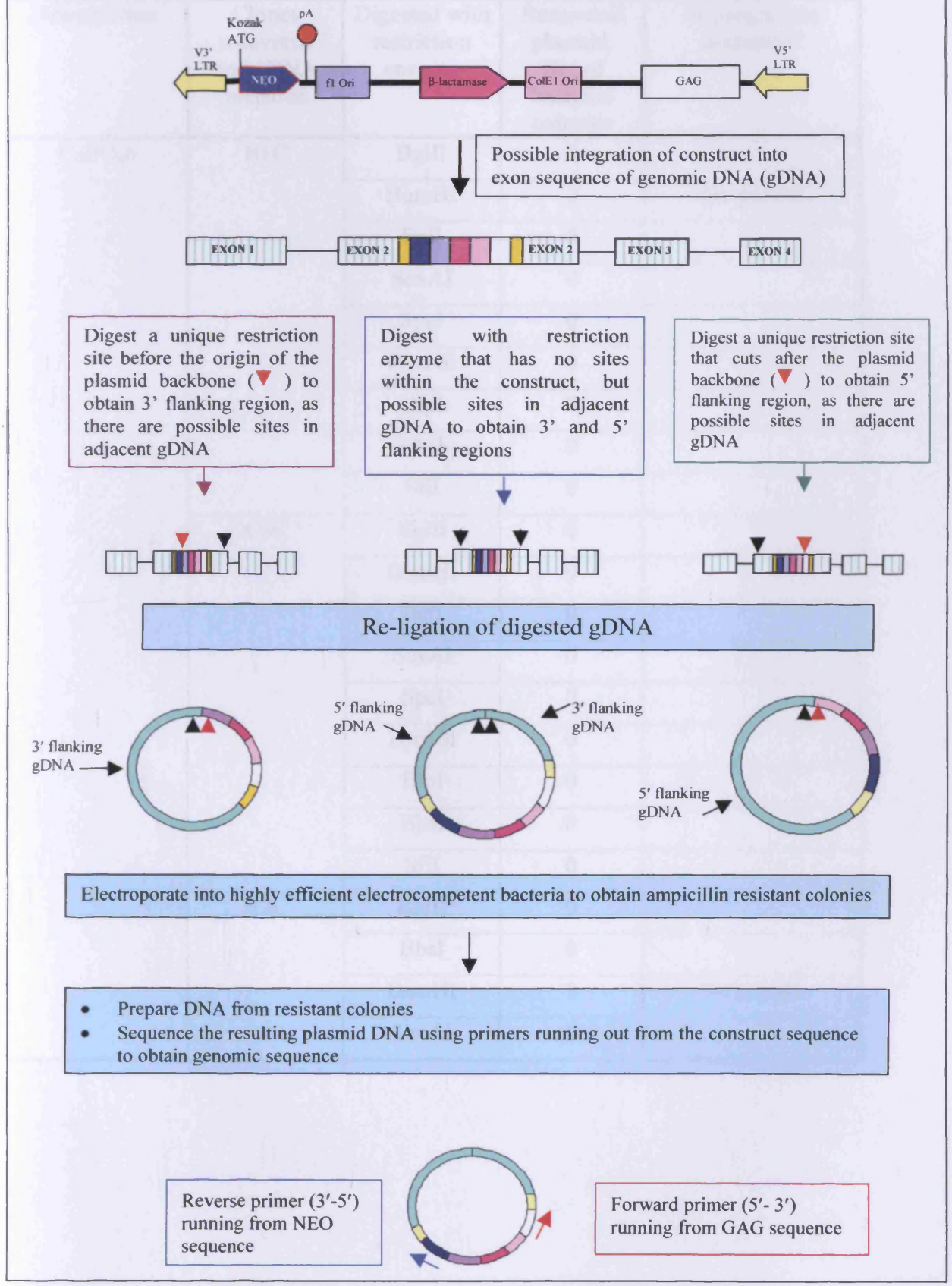


Table 3.3. Outcome of restriction digests for different shuttle clones

Transfection	Clones recovered and gDNA prepared	Digested with restriction enzyme	Recovered plasmid. No. of bacterial colonies	Sequence data obatained?
$CaPO_46$	B1C	BglII	0	
		BamHI	2	No growth
		BclI	0	
		SexAI	0	
		SpeI	0	
		Bsu36I	0	
		BlpI	0	
		BbsI	0	
		SfiI	0	
	C3C	BglII	0	
		BamHI	0	
		BclI	0	
		SexAI	0	
		SpeI	0	
		Bsu36I	0	
		BlpI	0	
		BbsI	0	
		SfiI	0	
	H2C	BglII	0	
		BbsI	0	
		BamHI	1	No growth
		SexAI	0	

Table 3.3. Outcome of restriction digests for different shuttle clones (continued)

Transfection	Clones recovered and gDNA prepared	Digested with restriction enzyme	Recovered Bacterial Colonies		Sequence data obtained?
			+ RCA	- RCA	
$CaPO_4$9	1A5	BstEII	1	0	Bad sequence
		BglII	11	0	All grew, bad sequence
		Bsu36I	0	0	
		NaeI	0	0	
	1D1	HindIII	0	0	
		BstEII	5	0	Bad sequence
		XbaI	0	0	
		Bsu36I	2	0	Bad sequence
	1A1	HindIII	0	0	
		BstEII	5	0	Bad sequence
		XbaI	0	0	
		Bsu36I	8	0	All grew, bad sequence
	2E1	BglII	0	1	No growth
		BbsI	0	0	
		BamHI	0	7	1 colony grew, bad sequence
		SexAI	0	0	
	2C1	BglII	0	0	
		BbsI	0	0	
		BamHI	0	3	Bad sequence
		SpeI	0	3	Bad sequence
		SexAI	0	0	
	2B2	BstEII	5	0	Bad sequence
		Bsu36I	0	0	
		NaeI	0	0	
		BglII	0	0	
	1G1		0	0	
	3D1	Bsu36I	0	0	
		NaeI	0	0	
		BglII	0	0	

Table 3.3. Outcome of restriction digests for different shuttle clones (continued)

Transfection	Clones recovered and gDNA prepared	Digested with restriction enzyme	Recovered Bacterial Colonies		Sequence data obtained?
			+ RCA	- RCA	
$CaPO_4$11	G2(48)	XbaI	2	3	Bad sequence
	H2(48)	XbaI	0	0	
	D2(48)	XbaI	3	1	2 grew RCA+, bad sequence
	A3(48)	XbaI	0	0	
	B3(48)	XbaI	7	5	4 RCA+ and 3RCA-grew, bad sequence
	G1(48)	XbaI	0	0	
	F1(48)	XbaI	0	0	
	B2(48)	XbaI	1	0	Bad sequence
	D3(48)	XbaI	0	0	
	C2(72)	XbaI	0	0	
	C2(48)	XbaI	2	0	Bad sequence
	E2(72)	XbaI	0	0	

3.11.1. -Use of Glycogen when Precipitating DNA from 500µl Ligations

It was felt that due to the small yield of rescued plasmid expected and the large volume of ligation mix required perhaps the DNA needs a carrier to assist in precipitation.

Glycogen is a highly purified polysaccharide from mussels and can be used as a carrier for nucleic acid precipitation. It is an inert material compatible with most molecular biology procedures, including PCR, DNA sequencing and competent cell transformation which are the major concerns in this project.

It can be seen when comparing (table 3.3.) the number of rescued plasmids from $CaPO_46$, where glycogen was not used, to the number recovered from $CaPO_49$, where it had been used, there has been a definite improvement. The numbers of resistant bacterial colonies had improved and also some of the resultant liquid medium cultures had grown.

3.11.2. – Rolling Circle Amplification (RCA)

An attempt was made to amplify any rescued plasmid by rolling circle amplification (2.2.2.xii.d.) using TempliPhi™ DNA amplification kit (Amersham Biosciences, UK Ltd., Bucks, UK). As it was likely that only a tiny amount of plasmid would be recovered from the ligation reaction it was felt prudent to try and amplify what there was prior to electroporation.

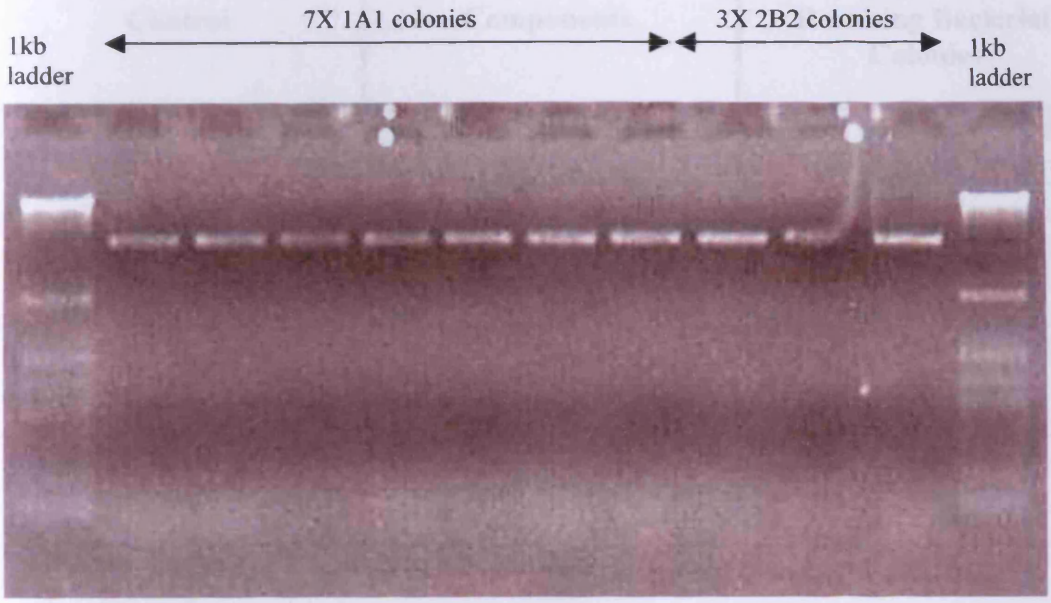
A positive control was set up using pUC19 plasmid with the RCA mix and only a few (~15) bacterial colonies appeared after electroporating into DH5α electrocompetent cells.

A total of ten bacterial colonies resulted after electroporating two RCA treated plasmid rescued clones ($CaPO_49$ - 1A1 and 2B2) into DH5α electrocompetent cells, 7 colonies for 1A1 and 3 for 2B2, (table 3.3.). This result is not really a considerable improvement on the untreated samples.

When the RCA treated and untreated samples are compared in $CaPO_411$ (table 3.3.) it can be seen that there is slight increase in the number of bacterial colonies retrieved with RCA treatment but not really noteworthy.

Plasmid DNA was isolated (2.2.2.x.) from the resulting ten bacterial colonies and electrophoresed on a 1% agarose gel (fig. 3.10.). The resulting DNA all appeared the same size on the gel (about 4kb), which was not expected, as two different clones were used.

Figure 3.10. 1% agarose gel showing the results of rolling circle amplification for two $CaPO_4$11 clones, 1A1 and 2B2.



Sequencing four of these products was attempted (fig. 3.9.) using both forward (gag) and reverse (neo) primers, resulting in unreadable sequence. There was a considerable amount of background, which would normally suggest that single bacterial colonies had not been picked adequately resulting in other plasmid DNA being present and therefore causing background when sequenced. However, this is extremely unlikely in this case as there were so few bacterial colonies present it would have been almost impossible for the accidental picking of two bacterial colonies together.

When next attempting RCA, another positive control, pUC19, was used as previously stated but also a negative control was prepared using 1µl of water instead of DNA with the normal RCA mix (table 3.4.).
On this occasion a more efficient electrocompetent cell was employed, ElectroTen Blue, Stratagene, La Jolla, CA 92037.

Table 3.4. Showing controls used when electroporating RCA products into ElectroTen Blue electrocompetent cells

Control	Components	Resulting Bacterial Colonies
RCA positive	(RCA mix + pUC19 DNA) + electrocompetent cells	Lots of descent sized colonies
RCA negative	(RCA mix + water) + electrocompetent cells	No colonies
Electroporation positive	PUC18 DNA + electrocompetent cells	Lots of descent sized colonies
Electroporation negative	Water + electrocompetent cells	No colonies

Seven RCA treated plasmid rescue clones, the positive RCA control, the negative RCA control, pUC18 positive electroporation control and a negative (no plasmid) electroporation control were all electroporated into ElectroTen Blue electrocompetent cells.

From all 7 of the RCA treated plasmid rescue clones there was a total of 18 bacterial colonies recovered, these were picked and expanded in liquid medium.

The resulting plasmid DNA from the 18 RCA treated plasmid rescue bacterial colonies was amplified (section 2.2.2.xv.) using both forward (gag) and reverse (neo) primers to try and obtain a single, discrete band product which could then be purified (section 2.2.2.xix) sequenced. This step was to try and eliminate background sequence prior to the sequencing reaction (section 2.2.2.xvi.). Unfortunately, this did not appear to improve the result, as the sequence data received was still unreadable.

Redesigned primers in both the gag and neo regions hoping to increase complementarity and thereby improve the quality of PCR products. Again no great improvement was seen with the resulting sequence data.

When comparing the positive controls for both the DH5α and the ElectroTen Blue electroporations it was seen that the ElectroTen Blue cells gave a much more efficient return of resistant bacterial colonies, so this was the eletrocompetent cell of choice in further plasmid rescue attempts.

3.11.3. – Size of Resistant Colonies

On a number of occasions picked bacterial colonies never expanded when transferred to liquid medium. The colonies were very small at the time of picking and had taken longer than normal to appear, about 24 hours on some occasions.

The fact that there were no larger colonies present on the ampicillin agar plate eliminates the theory that these were satellite colonies. Larger ampicillin resistant transformants would be required to excrete *beta lactamase* into the medium, thereby depleting the antibiotic in that area and allowing the formation of ampicillin sensitive satellite colonies over an extended period of time.

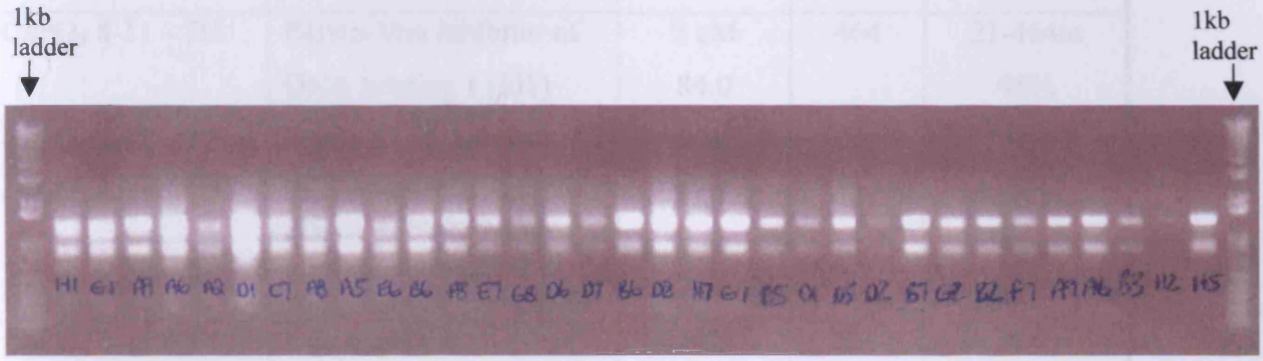
When these tiny colonies were transferred to LB broth, many did not expand, even when left for over 24 hours. Perhaps over a period of time on the ampicillin agar plate a couple managed to survive due to the reduced activity of the ampicillin, but when introduced to the ampicillin rich liquid medium they could not perpetuate survival.

Progress was made with the plasmid rescue methodology, in as far as isolating resistant bacterial colonies, possibly containing the rescued plasmid. However, this could never be confirmed as no readable sequence data was ever forthcoming. After many sequencing attempts, plasmid DNA isolated from the resistant bacterial colonies was sent to a commercial sequencing company (Oswel Research Products Ltd., University of Southampton, UK), but again sequencing failed. This sequencing problem needed to be investigated further, but due to time constraints, a need was felt to prioritise as deadlines were looming, so attention was focused on the splice acceptor series.

3.12. SPLICE ACCEPTOR SERIES RESULTS

All splice acceptor G418^R ES cell colonies were expanded to a minimum of 5x10^6 cells and then cytoplasmic RNA was prepared (section 2.2.2.xiii.). The resulting RNA was electrophoresed on a 1% agarose gel (section 2.2.2.iv) to ensure a successful preparation was achieved.

Figure 3.11. 1% agarose gel showing the successful isolation of RNA for a number of splice acceptor series clones from $CaPO_4$ 8-11



After checking the RNA (fig. 3.11.), proceeded with 3′ RACE (2.4.2.) reactions. The clones that gave a product during 3′ RACE, bacterial colonies after cloning into pAMP1 (2.4.3.) and subsequent sequence information can be seen in table 3.5.

On each occasion, a minimum of six bacterial colonies was picked for each clone. Each bacterial colony was prepped and then sequenced with the appropriate primers. After conducting a VecScreen (NCBI Database Entrez Nucleotide) on the resulting sequence information, a CAP EST Assembler at IFOM was run to obtain a contiguous sequence for each clone.

These contiguous sequences obtained from the 3′ RACE products are listed in Appendix III.

Table 3.5. Positive results of 3′ RACE and subsequent comparison with NCBI GenBank data using Blastn homology program showing similarity to non-redundant gene sequences

Transfection-clone	Sequence database search after VecScreen	Insertion into genome	No. of nt in BLAST search	% identity
$CaPO_4$ 9 – $1A6_{72}$	Blastn-Mm vaccinia related kinase 3	7 B2	79	12-79nt 100%
$CaPO_4$ 9 – $1F8_{72}$	Blastn-Mm (similar to) phophoglycerate kinase I	12 A1.1	67	20-67nt 97%
$CaPO_4$ 8-11 – H1	Blastn-Mm Inhibitor of DNA binding 1 (Id1)	2 cM 84.0	464	21-464nt 95%
$CaPO_4$ 8-11 – D1	Plasmid sequence	NA	72	NA
$CaPO_4$ 8-11 – E7	Blastn-Mm RIKEN cDNA 1500001M02 gene, mRNA (cDNA clone IMAGE: 4218748)	17 E1.3	447	183-436nt 92%
$CaPO_4$ 8-11 – B5	Bad sequence	NA	121	NA
$CaPO_4$ 8-11 – H7	Bad sequence	NA	186	NA
$CaPO_4$ 8-11 – $F6_{48}$	Blastn-Mm ribosomal protein L30 (Rpl30)	15 cM 14.7	618	1-434 99%
$CaPO_4$ 11 – $D6_{72}$	Bad sequence	NA	76	NA
$CaPO_4$ 11 – $E12_{48}$	Bad sequence	NA	82	NA
$CaPO_4$ 11 – $E11_{48}$	Blastn-Mm ribosomal protein L28	7 cM 4.0	452	119-241nt 96%
$CaPO_4$ 11 – $F11_{48}$	Bad sequence	NA	63	NA
$CaPO_4$ 11 – $B2_{48}$	Blastn-Mm Cytochrome c oxidase subunit Vb	1 B	807	16-450nt 99%
$CaPO_4$ 11 – $B6_{72}$	Bad sequence	NA	61	NA
$CaPO_4$ 11 – $D1_{48}$	Blastn-Mm Histone H2A.1 (AY158913)	13	395	119-378nt 99%

Table 3.5. Positive results of 3′ RACE and subsequent comparison with NCBI GenBank data using Blastn homology program showing similarity to non-redundant gene sequences (continued)

Transfection-clone	Sequence database search after VecScreen	Insertion into genome	No. of nt in BLAST search	% identity
$CaPO_4$ 11 – $H3_{72}$	Bad Sequence	NA	61	NA
$CaPO_4$ 11 – $D4_{48}$	Blastn-Mm Zinc finger protein 369, mRNA (cDNA clone MGC:41715 IMAGE:1365300)	13	283	71-238nt 97%
$CaPO_4$ 11 – $H6_{72}$	Bad sequence	NA	145	NA
$CaPO_4$ 11 – $C7_{72}$	Bad sequence	NA	82	NA
$CaPO_4$ 8-11 – B6	Blastn-Mm ribosomal protein L12 (Rpl12)	2B	581	51-556nt 98%
$CaPO_4$ 11 – $D8_{72}$	Bad sequence	NA	42	NA
$CaPO_4$ 11 – $H8_{72}$	Blastn- Mm proteasome (prosome, macropain) subunit, beta type 6	11 cM 39.0	190	35-190nt 97%
$CaPO_4$ 11 – $E5_{72}$	Blastn-Mus musculus ribosomal protein L30	15 cM 14.7	719	55-498nt 99%
$CaPO_4$ 11 – $F7_{72}$	Bad sequence	NA	59	NA
$CaPO_4$ 11 – $D4_{72}$	Blastn-Mm HRAS like suppressor 3	19A	191	19-188nt 98%
$CaPO_4$ 11 – $G6_{72}$	Blastn-Mm peptidyl-prolyl isomerase A	11 cM 1.0	485	15-485nt 99%
$CaPO_4$ 11 – $G4_{72}$	Bad sequence	NA	142	NA
$CaPO_4$ 11 – $B7_{72}$	Blastn-Mm clone IMAGE:3968454 mRNA	8	151	17-151nt 97%
$CaPO_4$ 11 – $C6_{72}$	Bad sequence	NA	81	NA

3.12.1 – Use of Feeder Layer

Originally STO feeders (section 2.1.2.) were being used to support the growth of $G418^R$ colonies. However, it became apparent that over this long period of selection the STO cells were not surviving, they were curling up and coming away from the plate, subsequently taking any $G418^R$ colonies with them.

It was decided that instead of using STO feeders, STO-NEO feeders would be employed as they should survive the G418 selection.

The use of STO-NEO feeders led to further considerations when attempting 3′ RACE, due to the feeders containing the *neomycin phosphotransferase* gene. The presence of this gene would lead to a product band from the feeders when 3′ RACE products were analysed on an agarose gel (section 2.2.2.iv). It was therefore essential that feeders were removed from the $G418^R$ ES cells prior to RNA extraction.

Removal of feeders entails passaging the cells a couple of times onto gelatinised plates and leaving the cells to settle for about 20 minutes. The feeders bed down before the ES cells, so after twenty minutes the medium can be removed still containing ES cells but minus a great many feeder cells. After a couple of passages most of the feeders were removed.

3.12.2. -PCR Screen

After receiving sequencing data back from 3′ RACE products containing nothing but plasmid sequence it was felt prudent to screen the bacterial colonies prior to sequencing (section 2.2.2.xviii). M13 primers were used to amplify the multiple cloning site of pAMP1 (see Appendix II for vector diagram), which is 400bp in size. Single band products greater than 400bp in size were selected and sent for sequence analysis.

3.12.3. -mRNA isolation

An attempt was made to increase the return of 3′ RACE products by isolating messenger RNA (section 2.2.2.xiv). As mRNA comprises about 1% to 4% of total RNA it was felt that the purified sample would give rise to a higher number of 3′ RACE PCR products.

There were a couple of successful products using mRNA but when again compared to total RNA it was realised that it was other factors that had caused the positive results, not the purifying of the sample. It was an adjustment of the PCR cycle programme and the reduction of the $MgCl_2$ concentration used in the 3′ RACE PCR reaction.
With the considerable effort, time and expense to isolate mRNA, the limited advantage was not felt worthwhile.

3.12.4. - T7 Enrichment

As mentioned in 3.5.2.b. the pCRR+SD126 T7 I-Sce construct contains a T7 promoter which can be utilised once RNA has been harvested from $G418^R$ colonies and cDNA prepared. In theory T7 RNA polymerase should copy more RNA from the T7 promoter through the splice donor and trapped transcript to the polyA and appended anchor sequences. By performing 3′ RACE on this enriched template it would be more likely that the correct region be amplified.

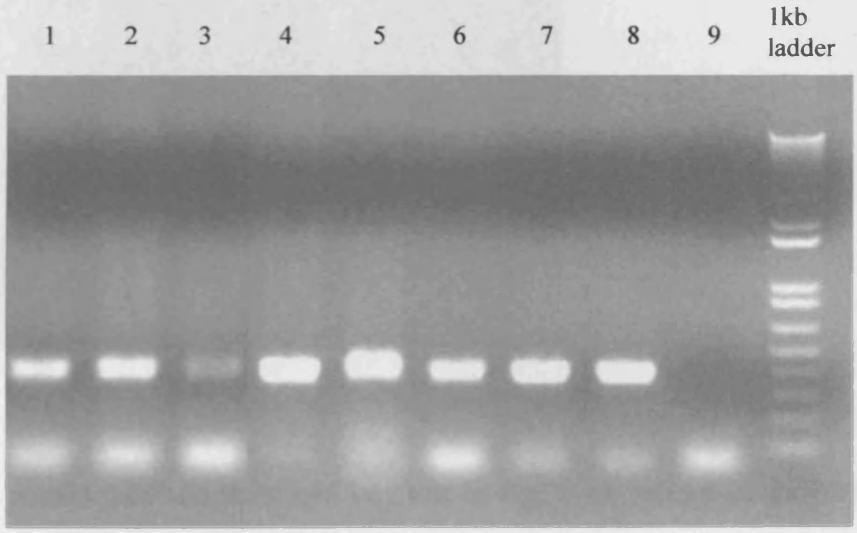
The T7 transcription was attempted as detailed in 2.4.1. After incubating for 6 hours and then precipitating RNA, the amplified RNA concentration and quality was read by measuring OD_{260} and $OD_{260/280}$ (section 2.2.2.xvii). Also, as unreliable readings were obtained from the spectrophotometer a small portion of the resulting mix was electrophoresed on an agarose gel to see if any RNA visible. It was expected that the antisenese RNA transcribed from the cDNA template would represent the size and complexity of the synthesised cDNA. A long smeared band was anticipated, but unfortunately not forthcoming.

Due to the small volume of amplified RNA loaded on the gel it was considered that it may not have been concentrated enough to be apparent, so proceeded with 3′ RACE (as detailed in 2.4.2.), with one exception. As the T7 promoter is located downstream of the *neo* cassette, the *neo* is lost during T7 enrichment, so it is essential that an alternative gene specific primer is utilised. The remaining gene specific region is the I-SceI site so a primer was designed specific for this area and named "Gene specific primer (I-SceI)" (table 2.4.1.).

After a number of attempts, this procedure gave no products after 3′ RACE PCR. Due to results beginning to come through with 3′ RACE PCR of the splice acceptor constructs this T7 enrichment step was put to one side whilst efforts were concentrated on obtaining sequence data and injecting ES cell clones into host blatsocysts.

3.12.5. -Genomic DNA PCR

To test whether retroviral sequence was preserved in transduced ES cell lines, genomic DNA was prepared (section 2.2.2.xi.) from a number of clones and then subjected to genomic PCR (section 2.4.7.) using specific primers for the retroviral sequence called gDNANEOF and gDNANEOR (table 2.4.1.) which prime from the β-geo gene.



KEY:
1 - $CaPO_4$11 $E12_{48}$
2 - $CaPO_4$11 $D6_{72}$
3 - $CaPO_4$11 $G4_{72}$
4 - $CaPO_4$11 $E5_{72}$
5 - $CaPO_4$11 $B6_{72}$
6 - $CaPO_4$11 $C7_{72}$
7 - $CaPO_4$11 $H8_{72}$
8 - positive control
9 - negative control

Figure 3.12. 1% agarose gel showing the results of genomic PCR to identify those clones with definite retroviral insertion

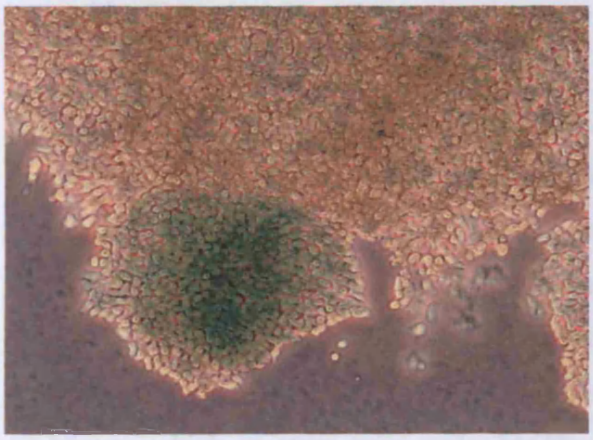
It can be seen from the above agarose gel picture (fig. 3.12.) that all seven samples show definite retroviral insertion with an obvious 470bp β-geo product band.

3.12.6 -Beta Galactosidase staining

A number of splice acceptor trapped clones were stained as described in section 2.4.4. As expression of the β-geo relies on the activity of the trapped promoter (section 3.5.2.), all $G418^R$ clones represent gene trap events and should be β-gal-positive. However, βgal activity was undetectable by X-gal staining in about 50% of colonies, reflecting the higher sensitivity of *neo* than β-gal activity.

It was also apparent that some clones gave lighter patterns of staining than others and not all cells in a clone stained blue. This observation is in agreement with previous work showing heterogeneity of staining in β-gal colonies (MacGregor et al., 1987).

Figure 3.13. Beta galactosidase staining (5X magnification) of one splice acceptor trapped ES cell line ($CaPO_4$11 $D6_{72}$)



A considerable contrast can be seen in fig. 3.13. where an individual colony shows much greater staining than the adjacent colony.

Of course, the inclusion of the *beta galactosidase* gene and the subsequent ability of staining really can be appreciated once germline chimeras have been created. The progeny can then be studied in detail to identify particular areas of expression with this staining technique.

3.12.7. -High Throughput 3'RACE

In order to obtain a greater number of positive products from 3'RACE, a high throughput option was considered (section 2.4.6.). A new robot (Multiprobe II HT EX) had been delivered to the school and so the temptation was there to try and increase throughput. It was necessary to try and plan the procedures in the laboratory before attempting anything with the robot.

A known positive clone $CaPO_4$11- $D1_{48}$ was grown up in a 96 multiwell plate and treated as detailed in section 2.4.6. Only the first three rows of cells were harvested as the expense of all the reagents restricted the scale of this experiment. Out of 24 wells there was one product band on the first attempt. In subsequent attempts of this nature there were never more than 3 out of 24 positives. Considering these are all meant to be positives it was not a great return and definitely not when considering the investment in both time and expense.

However, whilst trying to optimise protocols for use with the robot, there was a definite improvement in the laboratory technique where manual procedures were streamlined. When watching the robot pipette and then comparing the speed with a manual operator, personally, I think the manual technique is faster. There were a number of occasions where I witnessed the robot malfunction due to one of the pipettes becoming stuck and the tips not dispensing. These of course are technical problems, which in time will undoubtedly be improved. Due to time constraints, manual procedures were put back into action, returned to attempting 3'RACE as previously detailed (section 2.4.2.).

3.13. INVESTIGATING THE SPLICE ACCEPTOR CONSTRUCTS INSERTION SITES

After obtaining the preliminary results detailed in table 3.5. the insertion events of the different clones were investigated more thoroughly, at the nucleotide level. The recovered clone sequence was compared to the retroviral construct and cloning vector sequences to try and ascertain the exact insertion location. The homologous sequences were then highlighted in the reference sequence of the identified trapped gene.

Figure 3.13.

Sequence either side of the multiple cloning site in pAMP1 cloning vector

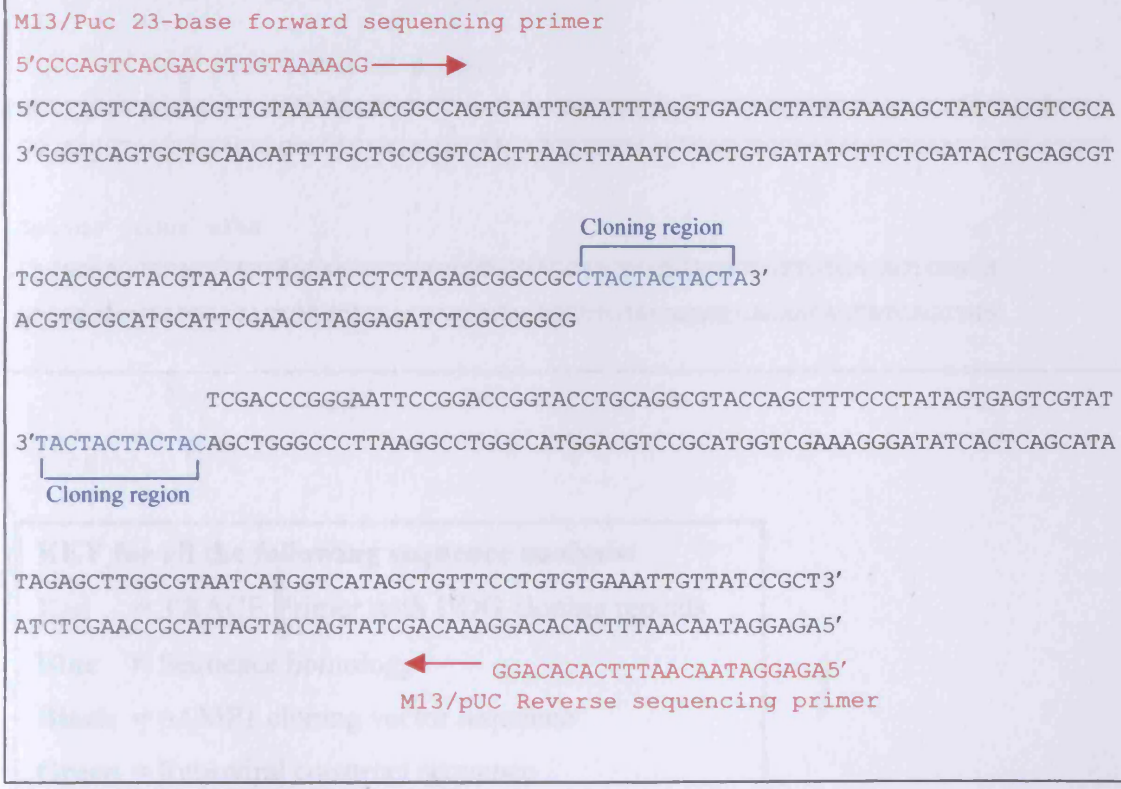


Figure 3.14.

Sequence surrounding the 3′RACE (NEO) primer site of the retroviral splice acceptor construct

```
CCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTT
GGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGGTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAA

                                                    UDG Cloning repeats
                                                    ┌──────────┐
                                                    CAUCAUCAUCAUATATTGC
CTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGC
GACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCGCAACCGATGGGCACTATAACG

3′RACE Primer Sequence
TGAAGAGCTTGGCGGC ──►
TGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATC
ACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAG

                STOP Codon of β-geo
GCCTTCTATCGCCTTCTTGACGAGTTCTTCTGATCTAGAACTAGTGGATCCAGCTTATCGAGCTTGCCAACTACAG
CGGAAGCTAGCGGAAGAACTGCTCAAGAAGACTAGATCTTGATCACCTAGGTCGAATAGCTCGAACGGTTGATGTC

Splice Donor site
CAGGTAAGCATATTTATGAAAAATGCTGATTGTGTTAGCTACTTGTGTCAGTGTTGTGATAGTCGAC3′
GTCCATTCGTATAAATACTTTTTACGACTAACACAATCGATGAACACAGTCACAACACTATCAGCTG5′
```

KEY for all the following sequence analysis:

Red = 3′RACE Primer with UDG cloning repeats

Blue = Sequence homology

Black = pAMP1 cloning vector sequence

Green = Retroviral construct sequence

Pink = No Match

3.13.1. – Clone $CaPO_4$9 – $1A6_{72}$ showed homology to Mus Musculus Vaccinia Related Kinase 3 (Vrk3)

5'CGCCCAATACGCAANCCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGNTGGCACGACAGGTTTCCCGAC
TGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGNTCACTCATTAGGCACCCCAGGCTTTACACTTTA
TGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGNAACAGCTATGACCATGATTA
CGCCAAGCTCTAATACGACTCACTATAGGGAAAGCTGGTACGCCTGCAGGTACCGGTCCGGAATTCCCGGGTCGAC
ATCATCATCATATATTGCTGAAGAGCTTGGCGGCGGCGGCAGCAGCCACTCCACTCCCTATGGCATTTCTGTGATG
Homologous to Vrk3
GCATAATAAACTGTTTTTAATCAAAANNAAAAAAAAAAGTACTAGTCGACGCGTGGCCTAGTAGTAGTAGGCGG3'

It can be seen from the above sequence that the *neo* primer (red) runs directly into the "trapped" gene sequence (blue). There is no retroviral construct sequence between the primer and the novel gene sequence, which was very unexpected. However, it can be seen that there are 6 bp of homology from the *neo* primer with the Vrk3 gene. This could be explained as an artefact product from the 3'RACE PCR reaction.

It can be seen from the below diagrams that the sequence homology is located in the final exon, number 14, of the Vaccinia Related Kinase 3 (Vrk3) gene.

EXON 14 SEQUENCE (291bp)

5'AGCTTCCGACTTGCAGCTTGAAGTAGAACATGAAGTAGTGTGACTGGAGGCCTGTTTGA
ACTCATAGCTCCTAAAAGAATCCCTTGAATGTGCATTCTCACCGCTCCCTTAGGACATAT
GAATCAGCACTTGTGTTGGGGAACCTGAGTCATGTCATGTAATGTGAAACTCCTCCCTGTC
TCAGCTCTGGCAGCTGTGGATGGAGGTAAGTGGATGCTGGCGGCGGCGGCGGCAGCAGCC
ACTCCACTCCCTATGGCATTTCTGTGATGGCATAATAAACTGTTTTTAATC3'

3' DOWNSTREAM SEQUENCE

aaagtctggagagttcagtgttcacagaagttctggaagaatcttcctcc....

3.13.2. – Clone $CaPO_4$ 8-11 – H1 showed homology to Mus Musculus Inhibitor of DNA binding 1 (Id1)

5′CCCGGGGNNNNCCCCNGCATNTTNCNCCAGNTCTAAACGCCCCCTAAAGGAAAGCTGGTAGCCTGCAGGTACGGT
CCGGAATTCCCGGGTCGACATCATCATCATATATTGCTGAAGAGCTTGGCGGCCGAGGCGGCATTTGTTCCAGCCG
ACGATCGCATCTTGTGTCGCTTAGGCGGCGCACTGAGGGCCCAGATGGCCTCCAGCCCTTCAGGAGGCAAGAGGAA
AAAAGTGCTCTCGGTTCCCCAGGGGATCTCTGGGAAAGACACTACCGCAAGCCCCCGGACTCTTGGCGGATCGGTC
CAGTGGGTAAAGGGTTTGATCAACAGAGCCTCACCCTCTCCACCTTTCAGCCTCCAGAGACTTTGGGGAGGGGGTT
AATCAACCCCGCGTGTTTCTGTTTTATTGAAAAAGCAGACATTTTTTTTTTAAATGGTCACATTTCGTGCTTCTCG
GATTTCTGAGGAAATATTTTGTATTGTATATTACAATGATCCCTGGCTGAAAATATTGTTTTACAATAGTTCTATG
GGGGTGAGTTTTTTGTTGTTATTAAACAAACCCTTTAGATAAAAAAAAAAAAAAAAAAAAGTACTAGTCGACGCGT
GGCCTAGTAGTAGTAGGCGGCCGCTCTAGAGGATCCAAGCTTACGTACGCGTGCATGCGACGTCATAGCTCTTCTA
TAGTNCACCTAAATTCAATTCNCTGGC3′

Again, the above sequence shows homologous novel sequence running directly from the *neo* primer into the "trapped" gene. On this occasion, no homology between the *neo* primer and the novel sequence can be seen, in fact there are two nucleotides of non-homologous sequence between the *neo* primer and the beginning of homology with Id1. Looking at the area of primer binding, immediately before homology begins with the ID1 gene, it does not appear particularly GC rich, unlike the Vrk3 scenario. Sequence homology with the Id1 gene is located at the 3′ end of the single exon and running into 3′ downstream sequence.

SINGLE EXON (1151 bp)

5′TTGTACAACCTTTCTTCAACTTCTTGTTCTCTTCCCACACTCTGTTCTCAGCCTCCTCCGCTCCCCTCCGCCTGT
TCTCAGGATCATGAAGGTCGCCAGTGGCAGTGCCGCAGCCGCTGCAGGCCCTAGCTGTTCGCTGAAGGCGGGCAGG
ACAGCGGGCGAGGTGGTACTTGGTCTGTCGGAGCAAAGCGTGGCCATCTCGCGCTGCGCTGGGACGCGCCTGCCCG
CCTTGCTGGACGAGCAGCAGGTGAACGTCCTGCTCTACGACATGAACGGCTGCTACTCACGCCTCAAGGAGCTGGT
GCCCACCCTGCCCCAGAACCGCAAAGTGAGCAAGGTGGAGATCCTGCAGCATGTAATCGACTACATCAGGGACCTG
CAGCTGGAGCTGAACTCGGAGTCTGAAGTCGGGACCACCGGAGGCCGGGGACTGCCTGTCCGCGCCCCGCTCAGCA
CCCTGAACGGCGAGATCAGTGCCTTGGCGGCCGAGGTGAGGTCCGAGGCAGAGTATTACATTATTCTTCAGTGGGA
AACCGAGGCCACGGGAGGCGGGTGTCCCCCTTCCTTACTTTTCAGGCGCATAGCTATTTAGGGGCGACTAATAGGA
AAAAGCTCGCATTTTCATCGTGCCTCCTGGAGTAGAGAAATGGGAACGCCTCTCCCCTCCTTGTCCTTTCCAGTGG
GTCTCATCCCTTATCTCGCTCTGGTGTTCACAGGCGGCATGTGTTCCAGCCGACGATCGCATCTTGTGTCGCTGAG
GCGGCGCACTGAGGGACCAGATGGACTCCAGCCCTTCAGGAGGCAAGAGGAAAAAAGTGCTCTCGGTTCCCCAGGG
GATCTCTGGGAAAGACACTACCGCAGCCACCGGACTCTTGGCGGATCGGTCCAGTGGGTAGAGGGTTTGATCAACA
GAGCCTCACCCTCTCCACCTTTCAGCCTCCAGAGACTTTGGGGAGGGGGTTAATCAACCCCGCGTGTTTCTGTTTT
ATTGAAAAAGCAGACATTTTTTTTTAAATGGTCACATTTCGTGCTTCTCGGATTTCTGAGGAAATATTTTGTATTG
TATATTACAATGATCACTGGCTGAAAATATTGTTTTACAATAGTTCTATGGGGGTGAGTTTTTTGTTGTTATTAAA
CAAACACTTTAG 3′

3′ DOWNSTREAM SEQUENCE

ataacgttgctgtgacttcttggatgggggaggggcgggacaggcatgac.....

3.13.3. – Clone $CaPO_4$ 8-11 – $F6_{48}$ showed homology to Mus Musculus Ribosomal protein L30 (NM_009083)

5′CGCTTTTNGCCANGCNTCCAATACNACTNNCTNTANGGCAAAGANTCGGGACACCNGCAGGTACCGGTCCGGAAT
TCCCGGGTCGACATCATCATCATATATTGCTGTAAGTAGCTTGGCGGCCTGGTGTTGGTCTGGGGGTTGTCCCGCA
CCTAAGGCAGGAAGATGGTGGCCGCAAAGAAGACGAAAAAGTCTCTGGAGTCGATCAACTCTAGGCTCCAACTTGT
TATGAAAAGTGGGAAGTACGTGCTGGGCTACAAGCAGACTCTGAAGATGATCAGACAAGGCAAAGCGAAGTTGGTT
ATCCTTGCCAACAACTGTCCAGCTTTGAGGAAATCTGAAATAGAGTACTATGCCATGTTGGCTAAAACTGGGGTCC
ATCACTACAGTGGCAATAATATTGAACTGGGCACAGCGTGTGGAAAATACTACAGAGTATGCACACTGGCTATCAT
TGACCCAGGTGATTCTGATATTATTAGAAGCATGCCAGAACAGACTGGTGAGAAGTAAACAAGAAAGTTTTCCTTT
AATAAAACTTTGCCAGAGCTCCTTTTGAAAAAAAAAAAAAAAAATTAAAACCTTAAAAAAAAAAAAAAAAAAGTNCT
ATTCNACNCGNGGCCTANTANTATTAGGCGGCCNCTCTANAGGATCCAAGCTNACGTACCCNNGCGTGCGAGGTNT
TAGCTNTTTTATNGNGNCCCCTAANTNAATTCACTGGCCGNGGTTTNAAACGNGNTGGGGNNAAAAA3′

Homologous sequence runs directly from the *neo* primer again, but there is no direct overlap with the primer. However, it can be seen from the below diagram that homology begins at a rather "GC" rich area in the exon sequence which probably influenced primer binding.

Sequence homology with the Rpl30 gene is located at the 3′ end of the single exon.

5′ UPSTREAM SEQUENCE

.....gcctaggcaaaccaatgagatcaataaaatgttatatttaaaataacaag

SINGLE EXON (462 bp)

5′CTTTCTCGCTCCCTGGCCATCTTGGCGGCTGGTGTTGGTTGGGGGTTGTCCCACACCTAA
GGCAGGAAGATAGTGGCCGCAAAGAAGACGAAAAAGTCTCTGGAGTCGATCAACTCTAGGCT
CCAACTTGTTATGAAAAGTGGGAAGTACGTGCTGGGCTACAAACAGACTCTGAAGATGATCA
GACAAGGCAAAGCGAAGTTGGTTATCCTTGCCAACAACTGTCCAGCTTTGAGGAAATCTGAA
ATAGAGTACTATGCCATGTTGGCTAAAACTGGGGTCCATCACTACAGTGGCAATAATATTGA
ACTGGGCACAGCGTGTGGAAAATACTACAGAGTATGCACACTGGCTATCATTGACCCAGGTG
ATTCTGATATTATTAGAAGCATGCCAGAACAGACTGGTGAAAAGTAAACAAGAAAGTTTTCC
TTTAATAAAACTTTGCCAGAGCTCCTTTTG 3′

3′ DOWNSTREAM SEQUENCE

aaaaaaaaataaataaataaaataaaataaaataaaataacaaagttcag.....

3.13.4. – Clone $CaPO_4$ 11 – $B2_{48}$ showed homology to Mus Musculus Cytochrome c oxidase subunit Vb

5'NGNTNTNCCAGNCTCCCAATACGACTCACTATAGGGAAAGCTNGTACGCCTGCAGGTACCNGGTCCGGAATTCCC
GGGTCGACATCATCATCATATATTGCTGAAGAGCTTGGCGGCGCAGGCCCTGAGGGCCCACGGCCCCCGTGGCGCG
Homology to Cox5b
GCCGTGACCCGCTCCATGGCTTCTGGAGGTGGTGTCCCCACTGATGAGGAGCAGGCTACTGGGCTGGAGAGGGAGA
TCATGATAGCAGCACAGAAGGGACTGGACCCATACAATATGCTACCTCCAAAGGCAGCTTCAGGCACCAAGGAAGA
CCCTAATCTAGTCCCGTCCATCAGCAACAAGAGAATAGTGGGCTGCATCTGTGAAGAGGACAACTGTACTGTCATC
TGGTTTTGGCTGCACAAAGGCGAGAGTCAGCGATGCCCCAACTGTGGAACCCATTACAAGCTGGTGCCCCACCAAA
TGGCCCACTGAGCCCCTGTGTTATCTTTTCAGAATGTAAAGAAAAACTTCTCTCTAATAAAGATTAGCCATTGCAC
TGGCTCCTCCCATAAAAAAAAAAAAAAAAGTACTANTCGACTNCTAGGCCTACGCGTGGCCTAGCTNGTNGNCNCGC
GGGCGCTNTACANGATCCNAGCTTACGTACGACGTGCATGCTACTTCATAGCNTNTTCNATAGTGNCACCTANNGT
CAATTCACTGNACNGCNGTTTNACNNGGNANNGACTGGGNAAAACCCTGNCGTTACCCATNTTAATCGACTTGCAT
TNATCCCCCTNTNGNCAATTGAAGNNAGTACCGAAAAAGNCCGCTNCGATNGACNTTNCCAAANTGATGNANNNTG
AATGGCNACNNGGNACCGCCCTNTAACNNCCNTNAACNCGGGGGNNNNGGNGGNTNNCCCANNNNTNCCN3'

It can be seen from the above diagram that there are 8 bp in the *neo* primer that are homologous to the Cox5b gene, this could definitely lead to artefact product during 3' RACE PCR. Sequence homology runs from exon 1 of the Cox5b gene, right through to exon 4 as shown below.

EXON 1

5'CATCTTGCTCAGCCTGTTCCCGGAAGTGCATCTGCTTGTCTCGGGCGAGATGGCTTCAAG
GTTACTTCGCGGAGTGGGCGCTTTGGCGGCGCAGGCCCTGAGGGCCCACGGCCCCCGTGG
CGCGGCCGTGACCCGCTCCATGGCTTCTGGAG3'

EXON 2

5'GTGGTGTCCCCACTGATGAGGAGCAGGCTACTGGGCTGGAGAGGGAGATCATGATAGCAG
CACAGAAGGGACTG 3'

EXON 3

5'GACCCATACAATATGCTACCTCCAAAGGCAGCTTCAGGCACCAAGGAAGACCCTAATCTA
GTCCCGTCCATCAGCAACAAGAGAATAGTGGGCTGCATCT3'

EXON 4

5'GTGAAGAGGACAACTGTACTGTCATCTGGTTTTGGCTGCACAAAGGCGAGAGTCAGCGAT
GCCCCAACTGTGGAACCCATTACAAGCTGGTGCCCCACCAAATGGCCCACTGAGCCCCTG
TGTTATCTTTTCAGAATGTAAAGAAAAACTTCTCTCTAATAAAGATTAGCCATTGCACTG
GCTCCTCCC3'

3' DOWNSTREAM SEQUENCE

atagctggctagtcttatgtctatagtctttttttttttaatatatatat.....

3.13.5. – Clone $CaPO_4$ 11 – $D1_{48}$ showed homology to Mus Musculus Histone H2A.1 (AY158913)

5'GTATNGCCNTGATTNCNCCNAGCTTTAATNCGNCTCCCTATAGGGAAANCTGGTACGCCTGCAGGTACCGGTCCG
GAATTCCCGGGTCGACATCATCATCATATATTGCTGAAGAGCTTGGCGGCCGTGCTGGAGTACCTGACGGCCGAGA

Homology to H2A.1

TCCTGGAGCTGGCGGGCAACGCGGCCCGCGACAACAAGAAAACGCGCATCATCCCGCGCCACCTGCAGCTGGCCAT
CCGCAACGACGAGGAGCTCAACAAGCTGCTGGGCCGCGTGACCATCGCGCAGGGCGGCGTCCTGCCCAACATCCAG
GCCGTGCTGCTGCCCAAGAAGACCGAGAGCCACCACAAGGCCAAGGGAAAATAAGACCAGCCGTTCACCCACCCGA
AAAAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCCTAGTAGTAGTAGGCGGCCGCTCTAGAGGATCCAAGCTTAC
GTACGCGTGCATGCGACGTCATAGGTATTCTGATAGTGTCACCCTAANTCCATTC3'

From the above diagram it can be seen that, again there is homology between the *neo* primer and the "trapped" gene, 7 bp. It looks as though this sequence information is not from the "trapped" gene but an artefact sequence amplified during 3'RACE PCR.

Sequence homology with the H2A.1 gene is located at the 3' end of the single exon.

5' UPSTREAM SEQUENCE

.....gtataccgtttttgcttcttactgcggtggttatctacagctgagttatg

SINGLE EXON (390 bp)

5'TCTGGACGCGGCAAGCAGGGCGGCAAGGCTCGCGCCAAGGCCAAGACTCGCTCCTCCCGG
GCCGGCCTGCAGTTCCCCGTGGGCCGCGTGCACCGGCTGCTCCGCAAGGGCAACTACTCG
GAGCGCGTGGGCGCCGGCGCCCCGGTGTACCTGGCGGCCGTGCTGGAGTACCTGACGGCC
GAGATCCTGGAGCTGGCGGGCAACGCGGCCCGCGACAACAAGAAGACGCGCATCATCCCG
CGCCACCTGCAGCTGGCCATCCGCAACGACGAGGAGCTCAACAAGCTGCTGGGCCGCGTG
ACCATCGCGCAGGGCGGCGTCCTGCCCAACATCCAGGCCGTGCTGCTGCCCAAGAAGACC
GAGAGCCACCACAAGGCCAAGGGAAAATAAGACCAGCCGTTCACCCACCCG3'

3' DOWNSTREAM SEQUENCE

aaccaaacattacgaatcaccaaaggctcttttcagagccactcactttc.....

3.13.6. – Clone $CaPO_4$ 11 – $H8_{72}$ showed homology to Mus Musculus Proteasome (prosome, macropain) subunit, beta type 6 (Psmb6)

5′TNGTATGTTGTGTGGAATTGTGAGCGGATACNATTTCCCACAGGNAACAGNCTATGGCCCATGATTACGCCAAGC
TCTAATACGACTCACTATAGGGAAAGCTGGTACGCCTGCAGGTACCGGTCCGGAATTCCCGGGTCGACATCATCAT
CATATATTGCTGAAGAGCTTGGCGGCCATTCAGGTGTCAGGGGTAGAGCGGCAGGTGCTTTTGGGAGACCAAATCC

Homology to Psmb6

CCAAGTTCACCATTGCCACGTTGCCACCTCCCTGAGGCCTGGTGTTCTAGGGAGTAGTAAAGGATGGTTCTTACCG
ACGCAGAAGCTAATAAACAGTTTGTTAAAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAGGTCCTAGTCGACGCGT
GGCCTATTAGTAGTAGGCGGCCGCTCTAGAGGATCCAAGCTTACGTACGCGTGCATGCGACGTCATAGCTCTTCCT
ATAGTGNCACCTAATTCAATTCACTGGCCGTCGTTTTACAACGNCGGGACTGGGAAACCCCTGGCGTTACCCCAC3′

In this example there are 9 bp of homology between the *neo* primer and the "trapped" gene. It appears that this sequence information is again the product of non-specific binding by the *neo* primer. An artefact product has been amplified and analysed.

It can be seen from the diagram below that homology is at the 3′ end of the final exon, number 6.

FINAL EXON NUMBER 6 (213 bp)

5′CTCTCGCTTTGGCCATGGAACGCGACGGCTCCAGTGGAGGGGTGATCCGCTTGGCAGCCA
TTCAGGAGTCAGGGGTAGAGCGGCAGGTGCTTTTGGGAGACCAAATCCCCAAGTTCACCA
TTGCCACGTTGCCACCTCCCTGAGGCCTGGTGTTCTAGGGAGTAGTAAAGGATGGTCCAT
ACCGACGCAGAAGCTAATAAACAGTTTGTTAAA 3′

3′ DOWNSTREAM SEQUENCE

gagacgtttctgtgtgaaggttcttgacttcaacttcaacttgccaaaac.....

3.13.7. – Clone $CaPO_4$ 11 – $D4_{72}$ showed homology to Mus Musculus HRAS like suppressor 3 (Hrasls3)

5′CACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGAC
CATGATTACGCCAAGCTNTAATACGACTCACTATAGGGAAAGCTGGTACGCCTGCAGGTACCGGTCCGGAATTCCC
GGGTCGACATCATCATCATATATTGCTGAAGAGCTTGGCGGCCCTGGGCCTCGTTGGAGTCATGCTCTCCAGAAAC

Homology to Hrasls3

AAGAAACAGAAGCAATGAGCTGAATGACTGCCCAGTTTTTGGGCTCTTCTTTTGCTAGAGGGTTCGGAGTTTGATT
TATAGATTCTATTGCTTTATAATTAGGTTTATTTTCACAACATACAATAAACCACAAGAAAGGAAAAAAAAAAAAA
AAAAAGTACTAGTCGACGCGTGGCCTAGTAGTAGTAGGCGGCCGCTCTAGAGGATCCAAGCTTACGTACGCGTGCA
TGCGACGTCATAGCTCTTCTATAGTGTCACCTAAATTCAATTCAGTGGCCGTCGTTTTACAACGTCGTGACTGGG3′

It can be seen that in this example there are 9 bp of homology between the *neo* primer and the "trapped" gene. An artefact product has been amplified and analysed.

Sequence homology is located at the 3′ end of the final exon, number 5.

FINAL EXON NUMBER 5 (236 bp)

5′GTCAGAGATGCGGTCAAGGCGGTAGGCATCGCTGGAGTGGGCTTGGCGGCCTTGGGCCTC
GTTGGAGTCATGCTCTCCAGAAACAAGAAACAGAAGCAATGAGCTGAATGACTGCCCAGT
TTTTGGGCTCTTCTTTTGCTAGAGGGTTTGGAGTTTGATTTATAGATTCTATTGCTTTAT
AATTAGGTTTATTTTCACAACATACAATAAACCACAAGAAAGGAATTTTTGTGAGG 3′

3′ DOWNSTREAM SEQUENCE

atactgcagtagctacctggtgtgaagtctctggagggaagctcggggct.....

3.13.8. – Clone $CaPO_4$ 11 – $G6_{72}$ showed homology to Mus Musculus Peptidyl-prolyl Isomerase A (PPIase)

5′ACTATAGGGAAAGCTGGTACGCCTGCAGGTACCGGTCCGGAATTCCCGGGTCGACATCATCATCATATATTGCTG
AAGAGCTTGGCGGCAGGTCCATCTACGGAGAGAAGTTTGAGGATGAGAACTTCATCCTAAAGCATACAGGTCCTGG

Homology to PPIase

CATCTTGTCCATGGCAAATGCTGGACCAAACACAAACGGTTCCCAGTTTTTTATCTGCACTGCCAAGACTGAATGG
CTGGATGGCAGGCATGTGGTCTTTGGGAAGGTGAAAGAAGGCATGAACATTGTGGAAGCCATGGAGCGTTTTGGGT
CCAGGAATGGCAAGACCAGCAAGAAGATCACCATTTCCGACTGTGGACAGCTCTAATTTCTTTTGACTTGCGGGCA
TTTTACCCATCAAACCATTCCTTCTGTAGCTCAGGAGAGCGTCCCTACCCCATCTGCTCGCAATGTCCTGTAATCT
CTGCTCTCACTGAAGTTCTTTGGGTTCCATATTTTCCTCATTCCCCTTCAAGTCTAGCTGGATTGCAAAGTTAAGT
TTATGATTATGAATAAAAACTAAATAAGAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCCTAGTAGTAGTAGGC
GGCCGCTCTAGAGGATCCAAGCTTACGTACGCGTGCATGCGACGTCATAGACTCTTCTATAGTGTCACCTAAATT3′

There are 7 bp of homology between the *neo* primer and the "trapped" gene. It appears another artefact product has been amplified and analysed.

Sequence homology can be seen in 3′ end of the fourth exon going through to the end of the final exon, number 5.

EXON 4 (173 bp)

5′GGTGGTGACTTTACACGCCATAATGGCACTGGCGGCAGGTCCATCTACGGAGAGAAATTTGAGGATGAGAACTTC
ATCCTAAAGCATACAGGTCCTGGCATCTTGTCCATGGCAAATGCTGGACCAAACACAAACGGTTCCCAGTTTTTTA
TCTGCACTGCCAAGACTGAATG 3′

EXON 5 (340 bp)

5′GCTGGATGGCAAGCATGTGGTCTTTGGGAAGGTGAAAGAAGGCATGAACATTGTGGAAGCCATGGAGCGTTTTGG
GTCCAGGAATGGCAAGACCAGCAAGAAGATCACCATTTCCGACTGTGGACAGCTCTAATTTCTTTTGACTTGCGGG
CATTTTACCCATCAAACCATTCCTTCTGTAGCTCAGGAGAGCGTCCCTACCCCATCTGCTCGCAATGTCCTGTAAT
CTCTGCTCTCACTGAAGTTCTTTGGGTTCCATATTTTCCTCATTCCCCTTCAAGTCTAGCTGGATTGCAAAGT
TAAGTTTATGATTATGAATAAAAACTAAATAAGAACTGTT 3′

3′ DOWNSTREAM SEQUENCE

tgtccttgtttgtattggagtggtataggtctctaagcaactaactggct.....

3.13.9. – Clone $CaPO_4$ 11 – $D4_{48}$ showed homology to Mus Musculus Zinc finger protein 369, mRNA (cDNA clone MGC:41715 IMAGE:1365300) (Q8K1Y8)

5′TAACTGGACANCCAGANGNAATCNGTNTTCTANGGTCGNGNNTGGNCAANACTAGCAGGTACCGGTCCGGAATTC
CCGGGTCGACATCATCATCATCATATATTGCTGAAGAGCTTGGCGGCCTCCGACTGGTTCAGGNCNATGANCAGA

Homology to Q8K1Y8

GAGTTCATGATTGCTTCATTTATCCTCAAAAGGAAATAAAAGATCTGAAAATGTGTCTAATGGATAACAAGCTCAA
GTTTCAAAAATTGCACTGATATTTATATACAGGGTACTGTGTGTAAGTGAACAAAACTAAAAGTATATAGATGAAT
GAATAAACTATTGTGTGGAACTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAANANA3′

There are 9 bp of homology between the *neo* primer and the "trapped" gene. It appears another artefact product has been amplified and analysed.

Sequence homology can be seen in 3′ end of the final exon, number 7.

EXON 7 (404 bp)

5′CCCTTGGCCAGGTTCACGTGGAAGATGGCGGGGCTGGCCCCAACCGGAATGGATCCCAGC
CTCTAGTCCGGCGGAGTTCCTTTATCCCTCTTCCTGCTGGCACAAGACCTTCTGCGATTC
TTTGGAGCTGATGTTGCGGTCCACTCACCCTTGATCCCGAGGTTCTGTGGCGTGGAGCGT
ACTCTGGAGCCCTTGGCGGCCTCCGCCTGGTTCAGGCAGAAGATCAGAGAGTTCTTAATT
GCTTCATTTATCCTCAAAAGGAAATAAAAGATCTGAAAATGTGTCTAAGGTTAACAAGCT
CAAGTTTCAAAAATTGCACTGATATTTTTATACAGGGTAAAGTGTGTAAGTGAACAAAAC
TAAAAGTATATAGATGAATGAATAAACTATTGTGTGGAACTTTT 3′

3′ DOWNSTREAM SEQUENCE

ttttctgctgtgatagctgtttttctttatcaatttgacatctggaatt.....

3.13.10. – Clone $CaPO_4$ 11 – $B7_{72}$ showed homology to Mus Musculus Clone IMAGE:3968454 mRNA

5′CGGTCCGGAATTCCTTGGGTCGACATCATCATCATATATTGCTGAAGAGCTTGGCGGCCCATCGCATCCTGTCAG

Homology to clone IMAGE:3968454

ACACATCCGACGAGGAGGACGTGAGTGTTTCCATAGGGGCTGGAGAGAAACTGAGGCTGTCAGCACACACAATGCT
CCCTGGTAGTAAGGCACGAGAGTCTTCTAGTTCTAGGCAGCAAAAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGC
CTAGTAGTAGTAGGCGGCCGCTCTAGAGGATCCAAGCTTACGTACGCGTGCATGCGACGTCATAGCTCTTCTATAG
TGTCACCTAAATTCAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA
TCGCCTTGCAGCACATCCCC3′

There are 6 bp of homology between the *neo* primer and the "trapped" gene. Another artefact product has been amplified and analysed.

Sequence homology can be seen below at the 3′ end of the reference sequence leading into the poly A tail.

Reference sequence of Mm clone IMAGE:3968454

5′gttgatgatgaagcaatagcagcataaatgaagatcaagtaaggggaggagaacgatgcccaagggtgggtgttc
taaaacaccacagcaagaagactttgctctcagcaacgacatggtggagaaacaaactgggaaaaaggataaagat
aaagtttctctaaccaagactcccaagctggaccgcagtgatggcgggaaggaggtgagagaacgagccaccaagc
ggaagctgcccttcaccgtgggggccaatggagagcagaaggactcggacacagagaagcagggtcccgagcggaa
gaggatcaagaaggagcctgttgcccggaagtccgggctgttgttcggcatggggctgtctgggatccgagctgga
taccctctctctgagcgtcagcaggtggctctgctcatgcagatgactgctgaggagtctgcaaatagtccggtag
acacaacaccaaagcacccttcccagtctacagtgtgtcagaaggggacccccaactctgcctcaaaaaccaaaga
taaggtgaacaagagaaatgaacgcggagagacccgcctgcaccgagcagctatccgaggggatgcccggcgcatc
aaggaactcatcagtgagggcgcggacgtgaacgtcaaggactttgcaggctggactgcactgcacgaggcgtgta
accggggctattacgacatcgccaagcaactgttagccgctggtgcagaggtgaacactaagggcctggatgatga
cacacctttgcatgatgccgccaacaacgggcactacaaggtggtgaagttgttgttacggtatggcggaaatcct
caacaaagcaacagaaaaggagagacaccgttgaaagtagccaattcgccaacgatggtgaacctcctgttaggca
aaggcacatacacttccagtgaggagagctctactgagagctcggaagaagaggacgccccgtcattcgcaccttc
cagctccgtcgatggcaataacacagactctgagtttgagaaaggcctcaagcttaaggctaagaacccagagccc
cagaagactgtgacccccgtcaaggatgagtacgagtttgatgaggacgacgagcaggacagggtccccccggtgg
acgacaagcacctgctgaagaaagactacaggaaggaggctaaggccaacagcttcatttccatacccaagatgga
agtgaagagctactcaaaaaacaacacgcttgcaccaaagaaggcggcccatcgcatcctgtcagacacatccgac
gaggaggacgtgagtgtttccatagggggctggagagaaactgaggctgtcagcacacacaatgctccctggtagta
aggcacgagagtcttctagttctaggcagcaaaaagaaaaaaaaaaaaaaaaaa3′

3.13.11. – Clone $CaPO_4$ 8-11 – E7 showed homology to Mus Musculus RIKEN cDNA 1500001M02 gene, mRNA (cDNA clone IMAGE:4218748

5'AAGGNAANCAGCTATGGCCCATGATTTCCCCAAGCTCTAATACGCCTCCCCTATAGGGCAAGCTGGTACGCCTGC
NAGGTNCCCGGTCCCGGAATTCCCGGGTCGACATCATCATCATATATTGCTGAAGAGCTTGGCGGCCTGACCCGCC
TCTTGCGTTTGCTTGCTCACTTCCTGTCTCAGACACCCACACCCCATAGGACCTTCTGGGTGCCACTTAGTTTCTC
AGTCCCTTCCCCTCCCCCATAGGGCCTACTGTGTGCCCCTTAGTTGAGAGCTTTTTTTTTAATGTATTTATTCCAG
GCCCTTTCTGTCCCTTAGCATGAGCATAATCAGACTGGAACGGGGGACAATGTTGTAAATTGTACTGAAACCGAGA
TGCAAATAAAAAATGATCCCCTGGGAAAAAAAAAAAANGNCCTTTTGGGNNGGGGCCCNNANATTNTANGGGGCCN
CCTTTANGNAANNCCAANNTTTANAACCNGGGNNGNNANNNNAANNCCTTTTTTTNGGGCCNCAAAAAAAAAAA3'

It can be seen that there is no exact homology with the *neo* primer on this occasion, but it is a very GC rich locus and again the homology with the "trapped" gene begins immediately after the primer sequence.

The below diagram shows homology with the Mus musculus RIKEN cDNA 1500001M02 gene at the 3′ end running into the poly A tail.

Reference sequence of Mm RIKEN cDNA 1500001M02 gene, mRNA (cDNA clone IMAGE:4218748)

5'ctctcgggtgtgcagcatcaggtcagatttaaaccgccaccatgtcgagcaaaaaagcgaagaccaagaccacca
agaagcgccctcagcgcgcaacctccaatgtgttcgccatgtttgaccagtcccagatccaggagttcaaagaggc
ctttaacatgattgaccagaaccgggatggcttcattgacaaggaggacctgcacgacatgctggcgtctctgggg
aagaatcccactgatgcctacctggacgccatgatgaacgaggccccgggccccatcaatttcaccatgttcctca
ccatgtttggggagaagctgaacggcactgaccccgaggacgtcatcagaaacgccttcgcttgctttgatgagga
agccacaggcaccatccaggaggattacctgagggagctgctgaccaccatgggcgaccgcttcacagacgaggaa
gtggatgagctgtacagagaggcccccattgacaaaaaggggaacttcaactacattgagttcacacgcatcctga
agcacggcgcgaaagacaaagatgactgaagagctgcggcggctgacacgcctcttgcgtttgcttgctcacttcc
tgtctcagacacacacaccccataggacctactgcgtgccacttagtttctcagtcacttcccctcccccataggg
cctactgtgtgccacttagttgagagctttttttttaaatgtatttattccaggccctttctgtcacttagcatgt
gcataatcagactggaacggggacaatgttgtaaattgtactgaaaacgagatgcaaataaaaaatgaaccactg
tgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3′

3.13.12. – Clone $CaPO_4$ 8-11 – B6 showed homology to Mus Musculus ribosomal protein L12 (Rpl12)

5′GACATCATCATCATATATTGCTGAAGAGCTTGGCGGCCCGTGCGCCTTAGGTGTACCTGAGGTGCACCGGAGGCG
AGGTCGGCGCCACATCCGCCTTGGCCCCGAAGATCGGTCCTCTGGGTCTGTCTCCGAAGAAAGTTGGCGATGACAT
TGCCAAGGCTACCGGTGACTGGAAGGGTCTCAGAATTACAGTGAAACTGACCATCCAGAACAGACAGGCCCAGATC
GAGGTGGTGCCTTCTGCCTCAGCCCTGATCATCAAAGCCCTCAAGGAGCCACCGAGAGACAGGAGGAAGCAGAAGA
ACATTAAACACAGTGGGAACATTACTTTTGATGAGATTGTCAACATTGCCCGGCAGATGAGACACCGGTCTTTGGC
TAGAGAACTTTCTGGAACTATCAAGGAGATCCTGGGTACTGCACAGTCTGTGGGCTGCAATGTGGATGGCCGCCAC
CCTCATGACATCATAGATGACATCAACAGTGGTGCAGTGGAGTGCCCAGCTNGTTAAGAAGCAACAAGAAAATATT
CC-ATTAAAGACTATCTGATAACCAGTAAAAAAAAAAAAAAAAAAAAAAA3′

There is no homology between the *neo* primer and the "trapped" sequence. Also, it does not appear that the sequence preceding primer binding is particularly rich in GC nucleotides. It can also be seen that there are two non-homologous nucleotides between the primer sequence and the beginning of homology with the Rpl12 sequence.

Homology with the ribosomal protein L12 can be seen at the 3′ end of the single exon, as shown in the diagram below.

5′ UPSTREAM SEQUENCE

.....ctgaatgttgaaaaaaccctaaataacccaatttaaaaaaaatggggct

SINGLE EXON (629 bp)

5′TTTTCGGCTTTCGGCTCGGAGGAGGCAACGGTGCAACTTTCTTCGGTCGTCCCGAATCCG
GGTTCATCCGACACCAGCCGCCTCCACCATGCCGCCCAAGTTCGACCCCAACGAGGTCAA
AGTCGTGTACCTGAGGTGCACCGGAGGCGAGGTCGGCGCCACATCCGCCTTGGCCCCGAA
GATCGGTCCTCTGGGTCTGTCTCCGAAGAAAGTTGGCGATGACATTGCCAAGGCTACCGG
TGACTGGAAAGGTCTCAGAATTACAGTGAAACTGACCATCCAAAACAGACAGGCCCAGAT
CGAGGTGGTGCCCTCTGCCTCAGCCCTGATCATCAAAGCCCTCAAGGAGCCACCGAGAGA
CAGGAAGAAGCAGAAGAACATTAAACACAGTGGGAACATTACTTTTGATGAGATTGTCAA
CATTGCCCGGCAGATGAGACACCGGTCTCTGGCTAGAGAACTTTCTGGAACTATCAAGGA
GATCCTGGGTACTGCACAGTCTGTGGGCTGCAATGTGGATGGCCGCCACCCTCATGACAT
CATAGATGACATCAACAGTGGTGCAGTGGAGTGCCCAGCTAGTTAAGCAACAAGAAAATA
TTTCAATAAAAGACTATCTGATAACCAGT 3′

3′ DOWNSTREAM SEQUENCE

taaaaaaaaaaaaaaaaaaaaaaaaatgggttacagatctaaacagaat.....

3.13.13. – Clone $CaPO_4$11 – $E11_{48}$ showed homology to Mus Musculus ribosomal protein L28 (Rpl28)

CNGCNATACNCCANGCTCCTCAATACGACTCACTATAGGGCAAAGCTGGTACGCCTGCAGGTACCGGTCCGGAATT
CCCGGGTCGACATCATCATACATATATCTGCTGTAAGTAGCTTGGCGGCCATCCGCAGGGTCCAGTGCCATCCTTC
Homology to Rpl28

GAAGCCAGAAGCCTGTGGCGGTGAAGAGGAAACGGACCCGCCCCACCAAGAGCTCCTGAGCCCCACACGCCCGAAG
CAATAAAGAGTTACAACTGACTTCCAAAAAAAAAAAAAAAAAAANNAGCNNTNCCCNGGNNAANGNGCATTATGCCC
NCCNGGNTANANNANCNAAANTTNCNCCCGNNNAAAAGNNACCCCTTNNCTGNGGNNTANNNNCANTTAAATNAAT
TTTNGTGGCNNGANGCNTTAACNGNAANNNGGGGGGGGAAATTTTNCCGGGNNNNNAANGGGAAAAAAAA

There are 7 bp of homology between the *neo* primer and the "trapped" gene, another artefact product.

Homology with the ribosomal protein L28 begins at the 3′ terminus of exon 4 and continues to the 3′ end of exon 5, as shown in the diagram below.

EXON 4 (119 bp)

GTCAGCGAAAACCTGCCACTTCTTATGTGAGGACCACCATCAACAAGAATGCTCGGGCTA
CCCTCAGCAGCATCAGACACATGATCCGAAAGAACAAGTACCGCCCTGATCTGCGTATG

FINAL EXON, NUMBER 5 (138 bp)

GCGGCCATCCGCAGGG-CCAGTGCCATCCTTCGAAGCCAGAAGCCTGTGGTGGTGAAGAGG
AAACGGACCCGCCCCACCAAGAGCTCCTGAGCCCCACACGCCCGAAGCAATAAAGAGTCC
ACTGACTTCCATTAGGCC

3′ DOWNSTREAM SEQUENCE

tccaaaatgttttgtgctcattgagctgtggctctgcgctgtctttagtt

3.13.14. – Clone $CaPO_4$11 – $E5_{72}$ showed homology to Mus Musculus ribosomal protein L30 (Rpl30)

CGCTTTTNGCCANGCNTCCAATACNACTNNCTNTANGGCAAAGANTCGGGACACCNGCAGGTACCGGTCCGGAATT
CCCGGGTCGACATCATCATCATATATTGCTGTAAGTAGCTTGGCGGCCTGGTGTTGGTCTGGGGGTTGTCCCGCAC

Homology to Rpl30

CTAAGGCAGGAAGATGGTGGCCGCAAAGAAGACGAAAAAGTCTCTGGAGTCGATCAACTCTAGGCTCCAACTTGTT
ATGAAAAGTGGGAAGTACGTGCTGGGCTACAAGCAGACTCTGAAGATGATCAGACAAGGCAAAGCGAAGTTGGTTA
TCCTTGCCAACAACTGTCCAGCTTTGAGGAAATCTGAAATAGAGTACTATGCCATGTTGGCTAAAACTGGGGTCCA
TCACTACAGTGGCAATAATATTGAACTGGGCACAGCGTGTGGAAAATACTACAGAGTATGCACACTGGCTATCATT
GACCCAGGTGATTCTGATATTATTAGAAGCATGCCAGAACAGACTGGTGAGAAGTAAACAAGAAAGTTTTCCTTTA
ATAAAACTTTGCCAGAGCTCCTTTTGAAAAAAAAAAAAAAAAATTAANACCTTAAAAAAAAAAAAAAAAAAGTNCTA
TTCNACNCGNGGCCTANTANTATTAGGCGGCCNCTCTANAGGATCCAAGCTNACGTACCCNNGCGTGCGAGGTNTT
AGCTNTTTTATNGNGNCCCCTAANTNAATTCACTG

There are 9 bp of homology between the *neo* primer and the "trapped" gene. This is the second time the Rpl30 gene has been identified (section 3.13.3.).

Sequence homology with the Rpl30 gene is located at the 3′ end of the single exon.

SINGLE EXON (462 bp)

5′CTTTCTCGCTCCCCGGCCATCTTGGCGGC-TGGTGTTGGT-TGGGGGTTGTCCCGCACCTAAGGCAGGAAGATGG
TGGCCGCAAAGAAGACGAAAAAGTCTCTGGAGTCGATCAACTCTAGGCTCCAACTTTTTATGAAAAGTGGGAAGTA
CGTGCTGGGCTACAAACAGACTCTGAAGATGATCAGACAAGGCAAAGCGAAGTTGGTTATCCTTGCCAACAACTGT
CCAGCTTTGAGGAAATCTGAAATAGAATACTATGCCATGTTGGCTAAAACTGGGGTCCATCACTACAGTGGCAATA
ATATTGAACTGGGCACAGCGTGTGGAAAATACTACAGAGTATGCACTCTGGCTATCATTGACCCAGGTGATTCTGA
TATTATTAGAAGCATGCCAGAACAGACTGGTGAGAAGTAAACAAGAAAGTTTTCCTTTAATAAAACTTTGCCAGAG
CTCCTTTTG3′

3′ DOWNSTREAM SEQUENCE

aaaaaaaaataaataaataaaataaaataaaataaaataacaaagttcag.....

3.13.15. – Clone $CaPO_4$ 9 – $1F8_{72}$ showed homology with Similar to Mus Musculus Phosphoglycerate kinase I (XM_283048)

GTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGG
ATAACAATTTCCCACAGGAAACAGCTATGACCATGATTACGCCAAGCTCTAATACGACTCACTATAGGGAAAGCTG
GTACGCCTGCAGGTACCGGTCCGGAATTCCCGGGTCGACATCATCATCATATATTGCTGAAGAGCTTGGCGGCGAA
TGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTG
ACGAGTTCTTCTGAGGGGATCCGCTGTAAGTCTGCAGAAATTGATGATGTATTAAACAATAAAGATGTCCCCTAAA
β-geo
stop codon
AAAAAANGNAAAAAAAAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCCTAGTATGTAGTAGGCGGCCA

It can be seen from the above diagram that there is viral sequence present in this 3′RACE product and it runs as far as the β-geo stop codon. This is still not as anticipated, it is expected that the viral sequence homology would continue as far as the splice donor site in the retroviral construct which is another 46 bp downstream. Sequence homology is shown below at the 3′ end of the gene, probably untranslated region.

The reference sequence was run in a NCBI open reading frame search and the result that gave a 414 aa protein product, which was as predicted from the reference sequence information so homology was aligned with this sequence.

Reference Sequence of Mm (similar to) phosphoglycerate Kinase I (XM_283048) after locating open reading frame (ORF) using NCBI ORF Finder.

Start Codon 73 bp

5′aaaagcgcacgtctgccgcgctgttctcctcttcctcatctccgggcctttcgacctcacggtgttgccaaaatg
tcgctttccaacaagctgactttggacaagctggacgtgaaggggaagcgggtcgtgatgagggtggacttcaacg
ttcctatgaagaacaaccagataacaaacaaccaaaggatcaaggctgctgttccaagcatcaaattctgcttgga
caatggagccaagtccgttgtccttatgagccacctgggccggcctgatggtgttcccatgcctgacaagtactcc
ttagagccagttgctgctgaactcaaatctctgctgggcaaggatgttctgttcttgaaggattgtgtgggcccag
aagtcgagaatgcctgtgccaacccagcggctgggactgtcatcctgctggaaaacctccgctttcatgtagagga
agaagggaagggaaaagatgcttctggaaacaaggttaaagctgagccggccaaaattgatgctttccgagcctca
ctgtccaaactaggagatgtctatgtcaatgatgcttttgggactgcacaccgagcccatagctccatggtgggtg
tgaatctgccacagaaggctggtggatttttgatgaagaaggagctgaactactttgccaaggctttggagagtcc
agagcgacccttcctggctatcttgggaggcgctaaagttgcagacaagatccagctgatcaataatatgctagac
aaagtcaatgagatgatcattggtggtggaatggcctttaccttccttaaggtgctcaacaacatggagattggca
catctctgtatgatgaagaaggagccaagattgtcaaagatctcatgtccaaagctgagaaaaatggtgtgaagat
taccttgcctgttgactttgtcactgctgacaaatttgatgagaatgccaagactggccaagctactgtggcctct
ggtatacctgctggctggatgggcttggactgtggtactgagagcagcaagaaatatgccgaggctgtgggtcgag
ctaagcagattgtttggaatggtcctgttggggtatttgaatgggaagcctttgccaggggaaccaagtcactcat
ggatgaggtggtgaaagccacttctaggggttgcatcactatcataggtggtggagacactgccacttgctgtgcc
aaatggaacacagaggataaagtcagccatgtgagcactgggggcggtgccagtctagagctcctggaaggtaaag
tccttcctggggtggatgctctcagcaatgtttagtattttctttcctgcctttggttcctgtgcacagctcctaa

Stop Codon 1326 bp

gtcaacctagtgttttccacatctccatttggtgttagcgcaagattcagctagtggctgagatgtggcaccagga
acccttaaacagttgcacagcatctcagctcgtctttactgcatcaggatttgtctgtgttcttcaagatcccatt
taaattccttagtgactaaaaccattgtgcattgtagagggcgtctatttatattctgcctgagaaaggaagtgag
ctgtaaaggctgagctctctctctgacgtatgtagcctctggttagcttcgtcactcactgttcttgactcagcat
ggcaatctgatgaaatcccagctgtaagtctgcagaaattgatgatctattaaacaataaagatgtccact3′

Table 3.6. Summary of sequence data analysis.

CLONE	Homology with *neo* primer (bp)	GC rich locus	Viral sequence
$CaPO_4$ 9 – $1A6_{72}$ - Mm Vaccinia Related Kinase 3 (Vrk3)	6	YES	None present
$CaPO_4$ 8-11 – H1 – Mm Inhibitor of DNA binding (Id1)	0	NO	None present
$CaPO_4$ 8-11 – $F6_{48}$ - Mm Ribosomal Protein L30 (Rpl30)	0	YES	None present
$CaPO_4$ 11 – $B2_{48}$ – Mm cytochrome oxidase subunit Vb (Cox5b)	8	YES	None present
$CaPO_4$ 11 – $D1_{48}$ – Mm Histone H2A.1 (AY_158913)	7	YES	None present
$CaPO_4$11 – $H8_{72}$ – Mm proteasome (prosome, macropain) subunit, beta type 6 (Psmb6)	9	YES	None present
$CaPO_4$11 – $D4_{72}$ – Mm HRAS like suppressor 3 (Hrasls3)	9	YES	None present
$CaPO_4$ 11 – $G6_{72}$ – Mm Peptidyl-prolyl Isomerase A (PPIase)	7	YES	None present
$CaPO_4$ 11 – $D4_{48}$ – Mm Zinc finger protein 369, (Q8K1Y8)	9	YES	None present
$CaPO_4$ 11 – $B7_{72}$ – Mm clone IMAGE:3968454 mRNA	6	NO	None present
$CaPO_4$ 8-11 – E7 – Mm RIKEN cDNA 1500001M02 gene	0	YES	None present
$CaPO_4$ 8-11 – B6 – Mm Ribosomal protein L12 (Rpl12)	0	NO	None present
$CaPO_4$ 11 – $E11_{48}$ – Mm Ribosomal protein L28 (Rpl28)	7	NO	None present
$CaPO_4$ 11 – $E5_{72}$ – Mm Ribosomal protein L30 (Rpl30)	9	YES	None present
$CaPO_4$ 9 – $1F8_{72}$ – Mm (similar to) Phosphoglycerate kinase I (XM_283048)	0	NO	**PRESENT**

3.14. INJECTING CLONES INTO HOST BLASTOCYSTS

After the above detailed investigation of the sequence data, obtained from 3′ RACE PCR products of resistant clones, it is clear that more analysis needs to be done to elucidate the true trapping events for these clones.

However, it can be seen that there is one result which appears to be a true product for phosphoglycerate kinase 1 as it definitely contains viral construct sequence (see 3.13.14).

At the time of printing this clone is being injected into host blastocysts.

CHAPTER 4

DISCUSSION

4. DISCUSSION

Due to the results being presented in a single chapter, this discussion chapter encompasses this whole gene trapping project and looks at possible explanations of anomalies, improvements of results and future work.

4.1. TRANSFECTION OF RETROVIRAL VECTORS INTO ECOTROPIC BOSC PACKAGING CELLS AND SUBSEQUENT INFECTION OF ES CELLS WITH VIRION PRODUCED

It was a major achievement to obtain G418 resistant ES cell colonies and to be able to expand them in culture. Overcoming this hurdle took quite a considerable time and was a culmination of many amended factors in both the transfection and infection protocols.

4.1.1. Using pCX-GFP plasmid individually as positive control

Amending the transfection positive control to a separate reaction, with the pCX-GFP plasmid being transfected individually into BOSC cells allowed more of the proviral DNA to be introduced. There was never an occasion when the transfection reaction did not work, as green BOSC cells were always visible about 12 hours after transfection, regardless of the method used. It can not be ascertained whether this amendment had a positive effect on the transfection reactions, as no quantitative measurements were taken. Also, no G418 resistant colonies were coming through at the time of this change in the protocol so no comparisons can be made with possible virion production. However, it is felt that allowing more proviral DNA to be transfected could only have a positive bearing on the situation.

4.1.2. Using harvested virion immediately

Using harvested virion immediately, instead of storing at -80°C for future infection reactions, again, cannot be shown conclusively to have improved results. However, it is definitely felt that this was an improvement, combined with other amendments made after the infection reaction, G418 resistant colonies began to appear.

4.1.3. Use of feeder cells

The fact that feeder cells were originally being used during infection was a definite pitfall. Mitotically inactivated feeder layers will mop up the retroviral particles as they can still be infected, but without cell division the virus cannot integrate. The undifferentiated ES cells were maintained using medium supplemented with LIF as opposed to using the feeder layer.

It also seemed unfavourable to have feeders present during selection of the ES cells. Selection continued over a period of approximately 12 days, during which time STO feeders were dying, curling up, coming away from the plate and bringing any resistant colonies with them. STO-NEO feeders were used to try and extend the period of attachment, but even the resistant feeders were curling up after 12 days of selection.

This being said, it was realised that feeders were actually needed when resistant ES cell clones were picked and transferred to 96 well plates. There are so few ES cells being transferred, the feeders are needed to promote cell attachment and viability. This was a major turning point as resistant ES cell clones began to survive being picked and grew in culture. Prior to this, a small number of ES cell clones had survived selection but never survived picking.

It is essential that prior to molecular manipulation all feeders are removed from the expanded, resistant ES cell clones, for example, if STO-NEO feeders are used the *neo* portion can interfere with primer binding during subsequent RACE reactions.

It can be concluded that certain features of feeder cells are necessary at particular times during a gene trapping experiment but for some procedures it is better to substitute medium supplemented with LIF to maintain pluripotentiality. Feeder cells are thought to have generalised effects, such as medium detoxification, in addition to more specific effects in promoting cell attachment and viability and of course, preventing ES cell differentiation (Nichols et al., 1990).

4.2. COMPARING SHUTTLE AND SPLICE ACCEPTOR CONSTRUCTS: RESULTS OF INFECTION PROCEDURE

All retroviral vectors were transfected into BOSC packaging cells using both the lipofectamine and calcium phosphate methods. Neither emerged as an optimal method but as it was the calcium phosphate method being used when results began to appear, this method was maintained. As the same number of ES cells were seeded and the same amount of construct DNA was transfected for subsequent infections, this enabled comparisons to be made between a number of transfection attempts (table 3.1.).

The shuttle and splice acceptor series constructs were compared as an infection assay, as opposed to an actual particle assay because no viral titres were known. The infection assay was calculated by dividing the total number of $G418^R$ colonies recovered for each retroviral construct by the total number of successful infections for each construct. This resulting figure gives an average value of the number of $G418^R$ colonies retrieved for each construct during an infection reaction (table 3.2.).

4.2.1. Splice acceptor versus shuttle series

It was clear that the splice acceptor series gave rise to a far greater number of resistant ES cell clones than the shuttle series. The shuttle series are activated from insertion into exons but as the splice acceptor constructs survive splicing events they can be activated from intron as well as exon insertions. This means that the splice acceptor constructs have more productive sites available when compared to the shuttle series. A major factor to be considered is that as all the retroviral vectors require a cellular promoter for activation, the maximum number of genomic targets equals the number of expressed genes.

4.2.2. Splice acceptor with IRES versus splice acceptor without IRES

It can also be seen that the splice acceptor construct with an IRES included upstream of the β-geo reporter gave the best result out of all the constructs (table 3.2.). The fact that the IRES performed as expected, by increasing the number of $G418^R$ colonies retrieved when compared to the splice acceptor construct without IRES, was

confirmed by applying a non-parametric, Mann Whitney test. The resulting p value of 0.0196 confirmed that the two constructs gave significantly different results.

The splice acceptor construct without the IRES has several limitations due to the requirement of translational fusions with endogenous tagged sequences and only some insertions will be fully productive. Only those trapping events generating a transcriptional fusion between the reporter and the coding region of the tagged gene in the correct frame and orientation can be detected. Insertions of the β-geo in the wrong frame or orientation, or fusion to the 5′ or 3′ untranslated regions of the host transcript will not be detected. Also, a partial or complete loss of reporter and/or selection activity might occur when β-geo protein is fused to certain endogenous protein sequences and some classes of genes might be absent or under represented. For instance, it has been shown that β-gal activity is lost when β-geo is fused to a signal peptide, therefore β-geo fusion with secreted molecules cannot be detected (Skarnes et al., 1995).

The splice acceptor construct containing the IRES almost doubled the number of $G418^R$ colonies recovered during infection (table 3.2.). The IRES allows for cap-independent translation of the β-geo from fusion transcripts with host sequences without the above constraints. Production of reporter/selection activity is independent of the reading frame of the tagged gene when IRES is present. Therefore, translation of β-geo depends only on the orientation and half of the insertions within chromosomal transcriptional units should be productive. As a consequence, the total number of $G418^R$ colonies is higher and fewer experiments are needed to obtain more colonies.

4.3. PLASMID RESCUE SERIES RESULTS

Unfortunately, no reproducible results were achieved in this area of the project. After amending many factors (section 3.11.) involved with plasmid rescue more resistant bacterial colonies were obtained but no resulting sequence data of any worth was attained. A brief discussion of successful amendments follows.

4.3.1. Glycogen

Using glycogen when precipitating DNA from the 500µl ligation reaction was a definite improvement as resistant bacterial colonies began to appear more readily. Prior to this, as the reactions were so dilute, the DNA was probably being lost. It can be seen that (table 3.3.), for $CaPO_46$ where glycogen was not used, a total of 3 resistant bacterial colonies appeared. However, in $CaPO_49$ where glycogen was used 14 resistant bacterial colonies appeared, which is a considerable improvement. It is also notable that some of these resistant bacterial colonies actually grew up when transferred to liquid growth medium.

Using glycogen as a carrier for nucleic acid precipitation definitely improved the return of resistant bacterial colonies, which was a step in the right direction.

4.3.2. Rolling circle amplification (RCA)

To try and improve results, rolling circle amplification (RCA) was attempted (section 1.8.). Resistant bacterial colonies appeared after RCA, but there was not really any great increase in the numbers obtained. The plasmid DNA retrieved from two separate shuttle clones was treated with RCA and electroporated. When the resulting resistant bacterial colonies were grown up and the plasmid DNA isolated and examined by agarose gel electrophoresis, the resulting plasmids all appeared to be the same size (~4 kb). This similarity was very unexpected as two different clones were used. It was considered whether these were random products of the RCA and that 4 kb was the maximum amplification size that could be reached during the incubation period. However, this does not account for the ability of the bacteria to survive selection. Sequencing a number of these products resulted in unreadable sequence with a considerable amount of background. This would normally suggest that single bacterial colonies had not been picked, but with so few colonies present on the plate this would have been very unlikely.

Occasionally the picked bacterial colonies did not grow when transferred to liquid medium. These colonies were particularly small and had taken an extended period of time to appear, sometimes ≥ 24 hours. It may be due to the length of incubation that some of colonies managed to survive due to the reducing activity of the ampicillin. When the colonies were then transferred to ampicillin rich liquid medium, growth could not be maintained.

It can be concluded that RCA offered no great advantage to the plasmid rescue method of gene trapping. Due to the continued lack of readable sequence data obtained after sequencing any plasmid rescue products, some were sent to a commercial sequencing company (Oswel Research Products Ltd., University of Southampton, UK) where they were also unable to sequence the plasmids effectively. As no sequence data was going to be obtainable without considerably more investigative work, attention was focused on the splice acceptor series.

With hindsight, it must be stated that if the unacceptable condition of the genomic DNA being used (Fig 3.8) had been appreciated earlier, steps could have been taken that may have improved the outcome of the plasmid rescue portion of this project.
Due to RNA contamination and lack of good quality gDNA, it was unlikely that there would be a sufficient amount of gDNA available after digestion for ligation and subsequent plasmid rescue.

4.4. SPLICE ACCEPTOR SERIES RESULTS

Once G418^R ES cell colonies were obtainable, which indicates successful gene trap events, efforts were centred on obtaining sequence data to identify the insertion location of the construct.

4.4.1. 3′ RACE PCR

Primarily, attempts were made with 5′ RACE PCR (Frohman et al., 1988; Skarnes et al., 1992). However, it proved to be extremely cumbersome when compared to the more accessible 3′ RACE PCR procedure. Results began to appear with 3′ RACE PCR and it was decided to proceed with this route of investigation. This was an original idea as 3′ RACE PCR has never been used to analyse splice acceptor traps in previous studies.

It had to be considered that by concentrating on 3′ RACE PCR, it was insertion events occurring in the 3′ regions of trapped genes that were being optimally amplified. Insertion events at the 3′ end can lead to more subtle phenotypes, with the majority of the gene remaining intact to produce a truncated protein. It is possible that the truncated protein retains sufficient normal activity to rescue the phenotype.

Another concern was whether the distance between the two priming positions for 3′ RACE PCR would be too great if the insertion event had occurred towards the 5′ end of the gene. The fact that 3′ RACE PCR is the amplification of cDNA, there is no concern over the size of intronic sequence. The average size of mRNA is approximately 2-3 kb and that of untranslated regions is about 300 bp. With advances in technology it is possible to efficiently produce PCR products up to 10 kb in length (Nielson et al., 1994). These facts considered, it was felt that the benefits of the more refined 3′ RACE PCR procedure outweighed the possible reduction in PCR products due to 5′ insertion events.

4.4.2. Beta Galactosidase staining

The β-geo gene depends on the expression of the endogenous gene for the production of *neo* and *lacZ* activities. Therefore, trapping with a β-geo vector should theoretically select for genes that are expressed in ES cells and the proportion of β-gal positive

colonies among the G418^R colonies should be close to 100% (Friedrich and Soriano, 1991). Using the splice acceptor traps only about 50% gave a positive result when stained for *lacZ* activity. There was a broad range in the distribution and intensity of β-gal staining among the different gene trapped cell lines, suggesting that the splice acceptor vectors are able to capture genes with different levels of expression in ES cells. The fact that some clones gave lighter patterns of staining than others and not all cells of a "clonal" population stained blue is in agreement with previous work showing heterogeneity of staining in β-gal colonies (MacGregor et al., 1987).

Theoretically, every G418^R clone should be positive for β-gal activity when the IRES construct is used. The low frequency of β-gal positive cells observed might be due to the capture of genes with weak promoter activities so that *lacZ* expression cannot be detected. The fact that β-geo is a fusion protein, suggests expect that all G418^R ES cell lines represent active gene trap events and should be β-gal positive, but the low frequency of β-gal positive cells, perhaps reflects the higher sensitivity of *neo* than β-gal activity.

The trapping of a weak promoter is also suggested by very faint and sometimes barely detectable dot-like expression found in several gene-trapped cell lines only after extensive β-gal staining. Another theory of this faint staining after extended periods of incubation could be that the pH optimum for the bacterial enzyme is between pH 7.0 and 7.5 (Reithel and Kim, 1960) whereas the optima of the two mammalian β-galactosidases lie between pH 3 and 6 (Lojda, 1970). Therefore, by extended staining incubation it may be possible that the staining solution becomes more acidic leading to endogenous β-gal activity being seen on these occasions.

The major advantage for the inclusion of the *beta galactosidase* gene will mainly be appreciated once germline chimeras have been created. The progeny can then be studied in detail to identify particular areas of expression with this staining technique.

4.4.3. High throughput

In principle, high-throughput analysis sounds very straightforward and it definitely does seem the way forward. There are many ES cell lines that have been created

during gene trapping experiments and are still awaiting analysis. For example, in this study alone a total of 671 $G418^R$ ES cell lines were isolated using splice acceptor constructs and around 230 were subjected to 3′ RACE PCR amplification, this was less than half, leaving about another 400 to be analysed.

An attempt was made during this project to create a high throughput protocol, beginning when the $G418^R$ colonies had been harvested and transferred to 96 well plates, continuing through RNA isolation, 3′ RACE PCR amplification up to the sequencing reaction. This proved difficult due to time constraints and a lack of experience with the new robot (Multiprobe II HT EX). This being said, it has to be acknowledged that a high throughput protocol would prevent the need for ES cell lines to be stored in liquid nitrogen awaiting analysis.

An example of high throughput analysis has been demonstrated by Hansen and co-workers. Using four different gene trap vectors and robotics after the stage of manually picking resistant clones, they generated 5,142 sequences adjacent to the gene trap integration sites (gene-trap sequence tags; http://genetrap.de) from >11,000 ES cell clones (Hansen et al., 2003). The preparation of mRNA, subsequent 5′ RACE PCR and direct sequencing of PCR products were all automated procedures, thereby allowing the analysis of >11,000 ES cell clones. This is definitely the way gene trap screens should be progressing, thereby reducing the number of unanalysed ES cell lines.

One area of the gene trap phenomenon that would greatly benefit from automation is the picking of $G418^R$ ES cell cones. This is an extremely uncomfortable and time-consuming process when attempted manually. It would necessary for the identification of any resistant ES cell clones to be apparent to an automated system. Perhaps one way would be a colour marker to be identified by a scanning beam. The ES cells need to remain viable after picking so it could not be the blue of *lacZ* staining, a more feasible option would be green fluorescent protein (GFP). GFP could be introduced into the cells as a fusion protein with *neo*, replacing the *lacZ* of present constructs. Perhaps if a Cre-mutant *lox* system (Araki et al., 1999) was utilised it

would be possible to exchange GFP for *lacZ* activity in the ES cell line at a later stage of experimental procedures.

If a high throughput method had been established in this study, perhaps more than one true gene trap event would have been recovered, but possibly, also a lot of artefact sequence data. This demonstrates the necessity for the thorough analysis of resulting data at all stages of experimentation. It must be stated that even though no high throughput protocol was developed in this study, due to optimising procedures for use with the robot, a definite improvement in laboratory technique was achieved with manual procedures being streamlined.

4.5. SEQUENCING RESULTS OF THE SPLICE ACCEPTOR SERIES

When analysing 3′ RACE PCR sequence products by comparing with NCBI GenBank data, a number of interesting insertion locations were discovered (table 3.5.). There was one in particular that caused excitement.

4.5.1. Mm Vaccinia-related kinase 3 (VRK3)

VRK3 is a member of the VRK proteins, which have recently been identified in the human kinome as a new family of p53 regulators (Manning et al., 2002). The VRK proteins, of which there are three, are members of a new signalling pathway but their targets have not yet been characterised despite their presence in almost all cell types tested. In adult organs the three genes are widely expressed and the ratio of each is similar, suggesting that VRK genes are likely to play a basic role in cell biology (Vega et al., 2003). A major concern is that as they are expressed during early phases of liver development suggesting that there is a significant risk that knock-out animals will be non-viable, as in *C. elegans* (Kamath et al., 2003), unless they are generated as conditional knock-out mice and the genes are induced after liver development. This was most exciting as it was thought we had this gene trapped in an ES cell line ready to inject. Also, it appeared that the gene was trapped at the 3′ end, leading to the possible production of a hypomorph with a mild phenotype, reducing the possibility of embryo lethality.

4.6. SEQUENCING ANOMALIES

After obtaining the preliminary sequence analysis results detailed in table 3.5, it was felt that the exact insertion location of the constructs should be investigated further, at the nucleotide level, before proceeding any further (section 3.13.).

From the pAMP1 cloned 3′ RACE products it was expected that we would see viral sequence running from the 3′ RACE primer sequence to the end of the β-geo sequence and further downstream to the splice donor site. At this site, novel sequence relating to the endogenous trapped gene would be identified. Unfortunately, this is not what occurred on the majority of occasions. It can be seen from table 3.6, that 14 out of 15 (93%) of the sequencing results showed no viral sequence. This means that when the resulting sequence from the pAMP1 cloned 3′ RACE product was analysed closely, the 3′ RACE primer sequence could be identified along with some of the cloning region from the pAMP1 cloning vector (fig. 3.13.), but there was no traceable sequence relating to the construct. Immediately after the identified 3′ RACE primer sequence, endogenous gene sequence was found. Without the presence of any construct sequence, questions were raised as to the efficacy of these "gene trap" events.

When studied closely, it appeared that 12 out of 14 of these sequenced pAMP1 cloned 3′ RACE products, showed a considerable amount of homology between the 3′ RACE primer sequence and the endogenous "trapped" gene. This homology of the 3′ RACE primer sequence with the endogenous "trapped" genes varied between 6 and 9 bp. It can be safely assumed that these 3′ RACE products are artefact and cannot be included in any future analysis of these ES cell lines.

It appears that the 3′ RACE primer sequence is GC rich and unspecific. However, it has to be considered that about 230 $G418^R$ ES cell clones were subjected to 3′ RACE PCR and out of all those analysed by agarose gel electrophoresis only 15 revealed a product band. This seems most strange, surely if the primer was so unspecific these products would have been picked up time and time again, as the population of mRNA is the same on each occasion? It had been expected that a reduced number of 3′ RACE PCR products would be seen, as amplification from the polyA would be

unlikely to reach as far as the *neo* priming locus, if insertions were in the 5′ region of endogenous genes. If a lot of 3′ RACE PCR products were being seen, then this may have prompted investigative action sooner. When a discrete, single band was not seen from a 3′ RACE PCR product being analysed on an agarose gel, then nested PCR was attempted. The fact that most of the 3′ RACE PCR products were artefacts explains why the nested PCR reaction never picked up any products, because the specific product had not been present in the first instance. Perhaps nested PCR should be used to specifically amplify a certain product from a large, varied population of initial amplifications.

There are two results that do not show any particular homology with the primer and priming does not appear to have begun in an obviously GC rich locus, but there is still no viral sequence present. These are probably the result of mispriming early on in the 3′ RACE PCR cycles, as amplification progressed misprimed products primed from a locus within a certain distance from the polyA to produce an amplified product.

All of these results point to the fact that the 3′ RACE primer sequence is not suitable for this amplification procedure. The primer sequence needs to be redesigned and possibly used in conjunction with a nested PCR reaction.

4.6.1. Similar to Mm phosphoglycerate kinase I (pgkI)

One pAMP1 cloned 3′ RACE sequence result does look more promising (section 3.13.15.), similar to Mm phosphoglycerate kinase I. It can be seen that there is definitely viral sequence present in this cloned 3′ RACE product and it runs as far as the β-geo stop codon. This is still an anomaly as it is expected that the viral sequence homology would continue as far as the splice donor site corresponding to the retroviral construct, which is another 46 bp downstream. It may be that there was a cryptic splice donor site, which caused this early splicing event, or it could simply be a putative sequence mistake introduced by the molecular manipulation during RACE PCR procedures.

It appears from the analysed sequence information, that this ES cell line is a true gene trap event as it contains viral construct sequence running from immediately after the

3′ RACE primer sequence for 93 bp downstream to the β-geo stop codon. The homology between the trapped sequence and the identified reference sequence (XM_283048) appears to be in the 3′ untranslated region. The fact that the sequence homology of the trapped gene sequence has been matched with "similar to" pgkI after searching the NCBI GenBank database using the Blastn homology program, required further investigation. The "similar to" pgkI sequence (XM_283048) was the product of NCBI Annotation Project and has been directly submitted (17.02.03) to the National Centre for Biotechnology Information.

When the "similar to" pgkI was aligned with the NCBI GenBank database pgkI gene there was a large area of 100% homology at the 5′ end of both sequences. This homology did not extend into the 3′ end of the two sequences. Both the "similar to" pgkI and the pgkI gene sequences were compared against the mouse genome sequence using the NCBI GenBank database and both mapped to the same chromosomal locations. This suggests that they are in fact the same gene but maybe the "similar to" sequence is a splice variant of the pgkI sequence.

This 3′ location, where the "similar to" pgkI and the pgkI sequences do not show homology is where the gene trap insertion is expected to have occurred. More genomic DNA analysis needs to be done to confirm this insertion location.
However, with no record of a knock-out for pgkI having been published, at the time of printing this ES cell line is being injected into host blastocysts.

4.7. CONCLUSION OF AIMS

The retroviral gene trap vectors examined in this study performed with a varying degree of success, depending on the sequence information contained. It was confirmed that the plasmid rescue construct pSHUT+124K SD126-1 does not perform as intended due to the incorrect location of the *neo* cassette (fig. 3.7.). The other plasmid rescue construct, pSHUT+124K-A, did result in a reasonable number of $G418^R$ ES cell clones. However, there was no success past this point of the plasmid rescue, gene trap procedure. This is possibly due to technical difficulties and not

attributable to the construct itself. Further work needs to be carried out to try and evaluate the true performance of this construct and ascertain where the problems lie.

The splice acceptor series vectors both resulted in a good number of $G418^R$ ES cell clones and also produced sequence data. Due to the problems arising with 3′ RACE PCR it is not possible to make any speculations as to the quality of the trapping events and subsequent sequence information obtained. Additional sequence information contained in these constructs was assessed. The splice acceptor construct (pCSI+SD126-2) containing an IRES performed as expected, by almost doubling the number of $G418^R$ ES cell clones retrieved. This construct also contains a 330 bp *lox* P stuffer region which when trapped lines are crossed through a cre expressing strain, should be lost, aiding the identification of heterozygous and homozygous progeny. This was not tested as no transgenics were produced.

The splice acceptor construct (pCRR+SD126 T7 I-Sce-X) contains a T7 promoter just upstream of the splice donor site. This promoter site can be utilised once RNA has been harvested from $G418^R$ colonies and cDNA prepared. In theory, T7 RNA polymerase will be able to copy more RNA from the T7 promoter site, through the splice donor and trapped transcript to the polyA and appended sequence. By performing 3′ RACE PCR on this enriched template it would be more likely that the correct region would be amplified. This procedure was attempted on 16 separate occasions, but no RNA amplification was identified and subsequently no 3′ RACE PCR products. It can be said that this modification does not function as expected, but again, it may be technical procedures that need attention and not a malfunction of the construct.

Three out of four constructs produced $G418^R$ ES cell clones, confirming their ability to function as productive gene trap constructs. However, it was only the IRES modification that has been shown conclusively to function as expected.

4.8. FUTURE WORK

Firstly, it would be wise to attempt a genomic PCR on the pgkI ES cell line to confirm the insertion location is as expected. This would mean designing primers to amplify the boundary regions, thereby confirming the locus of the gene trap event.

The fact that genomic PCR (section 3.12.5.) amplified a 470 bp region of the β-geo gene in a number of splice acceptor trapped ES cell lines allows us to confirm that they are gene trap clones. All of the ES cell lines isolated survived G418 selection, so it can therefore be assumed that they are all gene trap events. This means that there are a total of 671 splice acceptor $G418^R$ ES cell clones that need to be analysed. It would be interesting to repeat 3′ RACE PCR using an improved primer on the 15 ES cell lines that have already been analysed, to see if any similar results are obtained or if they were just artefact products. Maybe some adaptations to the 3′ RACE PCR protocol would be advantageous, reducing the number of cycles to 15 during the first round of PCR and then perform nested PCR on this specific product. This should ensure that only specific amplification is possible.

The plasmid rescue construct needs more investigation. The fact that $G418^R$ ES cell lines were retrieved, indicates that the construct performed as an insertional tag, but further work is needed to be able to obtain sequence data from rescued plasmids. Plasmid rescue is a very neat method of retrieving trapped sequence as demonstrated by a number of groups (Araki et al., 1999; Hicks et al., 1997).

If possible, it would be interesting to attempt to analyse the splice acceptor trapped ES cell lines by 5′ RACE PCR to see how many trap events may have been missed by using the 3′ RACE PCR procedure.

4.9. SUMMARY

Large-scale insertional mutations in mammalian cells can be induced effectively with gene traps. By selecting for gene expression, recombinants are obtained in which the reporter gene is fused to the regulatory elements of an endogenous gene. Transcripts generated by these fusions reflect the activity of individual cellular genes and serve as molecular tags to identify and/or clone any genes linked to specific functions (Floss and Wurst, 2002; Hicks et al., 1995; Stanford et al., 2001). Application of this technique in a genome-wide manner should allow the identification of most, if not all, active transcripts in the genome and thus it is an important tool for genome annotation. More importantly, gene trapping in mouse ES cells enables the establishment of ES cell libraries with mutations in most genes, which then can be used to make transgenic mice. This opens the possibility to assign a function to each gene in the context of an entire organism.

It has been shown that gene trap vectors can disrupt all functional classes of genes and are highly mutagenic in transgenic mice (Hansen et al., 2003). However, it must be considered that individual gene trap vectors contain their own biases for insertion into the genome. They complement each other in gene targeting, suggesting that the most effective way of saturating the mouse genome with mutations is by using a combination of different gene trap vectors.

In the future, new entrapment vectors and automation will have an important impact on tagged sequence mutagenesis.

BIBLIOGRAPHY

Adam, M.A. and Dusty Miller, A. (1988) Identification of a Signal in a Murine Retrovirus That is Sufficient for Packaging of Nonretroviral RNA into Virions. *Journal of Virology*, **62**, 3802-3806.

Adams, M.D., Rudner, D.Z. and Rio, D.C. (1996) Biochemistry and regulation of pre-mRNA splicing. *Curr Opin Cell Biol*, **8**, 331-339.

Akiyama, N., Matsuo, Y., Sai, H., Noda, M. and Kizaka-Kondoh, S. (2000) Identification of a Series of Transforming Growth Factor *B*-Responsive Genes by Retrovirus-Mediated Gene Trap Screening. *Molecular and Cellular Biology*, **20**, 3266-3273.

Allen, N.D., Cran, D.G., Barton, S.C., Hettle, S., Reik, W. and Surani, M.A. (1988) Transgenes as probes for active chromosomal domains in mouse development. *Nature*, **333**, 852-855.

Anderson, K.V. and Ingham, P.W. (2003) The transformation of the model organism: a decade of developmental genetics. *Nature Genetics*, **33**, 285-293.

Araki, K., Imaizumi, T., Sekimoto, T., Yoshinobu, K., Yoshimuta, J., Akizuki, M., Miura, K., Araki, M. and Yamamura, K. (1999) Exchangeable Gene Trap Using The Cre/Mutated *Lox* System. *Cellular and Molecular Biology*, **45**, 737-750.

Armentano, D., Yu, S.-F., Kantoff, P.W., von Ruden, T., Anderson, W.F. and Gilboa, E. (1987) Effect of Internal Viral Sequences on the Utility of Retroviral Vectors. *Journal of Virology*, **61**, 1647-1650.

Baker, R.K., Haendel, M.A., Swanson, B.J., Shambaugh, J.C., Micales, B.K. and Lyons, G.E. (1997) *In Vitro* Preselection of Gene-Trapped Embryonic Stem Cell Clones for Characterising Novel Developmentally Regulated Genes in the Mouse. *Developmental Biology*, **185**, 201-214.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M. and Gautheret, D. (2000) Patterns of variant polyadenylation signal usage in human genes. *Genome Res*, **10**, 1001-1010.

Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K. and Gehring, W.J. (1989) P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes and Development*, **3**, 1288-1300.

Bender, M.A., Palmer, T.D., Gelinas, R.E. and Dusty Miller, A. (1987) Evidence that the Packaging Signal of Moloney Murine Leukemia Virus Extends into the *gag* Region. *Journal of Virology*, **61**, 1639-1646.

Blanco, L., Bernad, A., Lazaro, J.M., Martin, G., Garmendia, C. and Salas, M. (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem*, **264**, 8935-8940.

Blanco, L. and Salas, M. (1984) Characterization and purification of a phage phi 29-encoded DNA polymerase required for the initiation of replication. *Proc Natl Acad Sci U S A*, **81**, 5325-5329.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. (1997) The Complete Genome Sequence of Escherichai coli K-12. *Science*, **277**, 1453-1474.

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

Breitbart, R.E. and Nadal-Ginard, B. (1987) Developmentally induced, muscle-specific trans factors control the differential splicing of alternative and constitutive troponin T exons. *Cell*, **49**, 793-803.

Brinster, R.L., Chen, H.Y., Trumbauer, M., Senear, A.W., Warren, R. and Palmiter, R.D. (1981) Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell*, **27**, 223-231.

Bronson, S.K. and Smithies, O. (1994) Altering mice by homologous recombination using embryonic stem cells. *Journal of Biological Chemistry*, **269**, 27155-27158.

Brown, S.D. and Balling, R. (2001) Systematic approaches to mouse mutagenesis. *Curr Opin Genet Dev*, **11**, 268-273.

Bukrinsky, M.I., Sharova, N., Dempsey, M.P., Stanwick, T.L., Bukrinskaya, A.G., Haggerty, S. and Stevenson, M. (1992) Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl. Acad. Sci. USA*, **89**, 6580-6584.

Burdon, T., Smith, A.G. and Savatier, P. (2002) Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends in Cell Biology*, **12**, 432-438.

Capecchi, M. (1989) The New Mouse Genetics: Altering the Genome by Gene Targeting. *Trends in Genetics*, **5**, 70-76.

Chowdhury, K., Bonaldo, P., Torres, M., Stoykova, A. and Gruss, P. (1997) Evidence for the stochastic integration of gene trap vectors into the mouse germline. *Nucleic Acids Research*, **25**, 1531-1536.

Chuck, A.S., Clarke, M.F. and Palsson, B.O. (1996) Retroviral Infection Is Limited by Brownian Motion. *Human Gene Therapy*, **7**, 1527-1534.

Chuck, A.S. and Palsson, B.O. (1996) Consistent and high rates of gene transfer can be obtained using flow-through transduction over a wide range of retroviral titers. *Hum Gene Ther*, **7**, 743-750.

Coffin, J.M. (1979) Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J Gen Virol*, **42**, 1-26.

Cone, R.D. and Mulligan, R.C. (1984) High-efficiency gene transfer into mammalian cells: Generation of helper-free recombinant retrovirus with broad mammalian host range. *Proc. Natl. Acad. Sci. USA*, **81**, 6349-6353.

Couldrey, C., Carlton, M.B.L., Ferrier, J., Colledge, W.H. and Evans, M.J. (1998) Disruption of Murine alpha-Enolase by a Retroviral Gene Trap Results in Early Embryonic Lethality. *Developmental Dynamics*, **212**, 284-292.

Craigie, R. (1992) Hotspots and warm spots: integration specificity of retroelements. *Trends in Genetics*, **8**, 187-190.

Davis, H.E., Morgan, J.R. and Yarmush, M.L. (2002) Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes. *Biophysical Chemistry*, 159-172.

Davis, S., Aldrich, T.H., Stahl, N., Pan, L., Taga, T., Kishimoto, T., Ip, N.Y. and Yancopoulos, G.D. (1993) LIFR beta and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor. *Science*, **260**, 1805-1808.

Dymecki, S.M. (1996) Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proc Natl Acad Sci U S A*, **93**, 6191-6196.

Eizinger, A., Jungblut, B. and Sommer, R. J. (1999) Evolutionary Change in the Functional Specificity of Genes. *Trends in Genetics,* **15**, 197-202.

Engels, W.R. (1993) Contributing software to the internet: the Amplify program. *Trends in Biochemical Sciences*, **18**, 448-450.

Evans, M. and Kaufman, M.H. (1981) Establishment in Culture of pluripotential cells from Mouse Embryos. *Nature*, **292**, 154-156.

Evans, M.J. and Hunter, S. (2002) Source and nature of embryonic stem cells. *C.R. Biologies*, **325**, 1-5.

Ewing, B. and Green, P. (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nature Genetics*, **25**, 232-234.

Floss, T. and Wurst, W. (2002) Functional Genomics by Gene-Trapping in Embryonic Stem Cells. *Methods in Molecular Biology*, **185**, 347-379.

Friedrich, G. and Soriano, P. (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes and Development*, **5**, 1513-1523.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA*, **85**, 8998-9002.

Gearing, D.P., Thut, C.J., VandeBos, T., Gimpel, S.D., Delaney, P.B., King, J., Price, V., Cosman, D. and Beckmann, M.P. (1991) Leukemia inhibitory factor receptor is structurally related to the IL-6 signal transducer, gp130. *Embo J*, **10**, 2839-2848.

Gilbert, W. (1978) Why Genes in Pieces? *Nature*, **271**, 501.

Gilboa, E., Goff, S., Shields, A., Yoshimura, F., Mitra, S. and Baltimore, D. (1979) In vitro synthesis of a 9 kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus. *Cell*, **16**, 863-874.

Goff, S.P. (2001) Intracellular trafficking of retroviral genomes during the early phase of infection: viral exploitation of cellular pathways. *The Journal of Gene Medicine*, **3**, 517-528.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) Life with 6000 genes. *Science*, **274**, 546, 563-547.

Gossler, A., Joyner, A.L., Rossant, J. and Skarnes, W.C. (1989) Mouse Embryonic Stem Cells and Reporter Constructs to Detect Developmentally Regulated Genes. *Science*, **244**, 463-465.

Hansen, J., Floss, T., Van Sloun, P., Fuchtbauer, E.M., Vauti, F., Arnold, H.H., Schnutgen, F., Wurst, W., von Melchner, H. and Ruiz, P. (2003) A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc. Natl. Acad. Sci. USA*, **100**, 9918-9922.

Hardison, R.C., Oeltjen, J. and Miller, W. (1997) Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res*, **7**, 959-966.

Hicks, G.C., Shi, E.-g., Lei, X.-M., Li, C.-H., Pawlak, M. and Ruley, H.E. (1997) Functional genomics in mice by tagged sequence mutagenesis. *Nature Genetics*, **16**, 338-344.

Hicks, G.G., Shi, E.G., Chen, J., Roshon, M., Williamson, D., Scherer, C. and Ruley, H.E. (1995) Retrovirus gene traps. *Methods Enzymol*, **254**, 263-275.

Hill, D.P. and Wurst, W. (1993) Gene and enhancer trapping: mutagenic strategies for developmental studies. *Curr Top Dev Biol*, **28**, 181-206.

Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987) HPRT-Deficient (Lesch-Nyhan) Mouse Embryos Derived from Germline Colonization by Cultured Cells. *Nature*, **326**, 292-295.

Ishida, Y. and Leder, P. (1999) RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells. *Nucleic Acids Research*, **27**, e35 i-viii.

Jaenisch, R. (1976) Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc Natl Acad Sci U S A*, **73**, 1260-1264.

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

Jareborg, N., Birney, E. and Durbin, R. (1999) Comparative Analysis of Noncoding Regions of 77 Orthologous Mouse and Human Gene Pairs. *Genome Research*, **9**, 815-824.

Jenkins, N.A. and Copeland, N.G. (1985) High frequency germline acquisition of ecotropic MuLV proviruses in SWR/J-RF/J hybrid mice. *Cell*, **43**, 811-819.

Joyner, A.L. (1991) Gene Targeting and Gene Trap Screens Using Embryonic Stem Cells: New Approaches to Mammalian Development. *BioEssays*, **13**, 649-656.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P. and Ahringer, J. (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, **421**, 231-237.

Kan, Z., Rouchka, E.C., Gish, W.R. and States, D.J. (2001) Gene Structure Prediction and Alternative Splicing Analysis Using Genomically Aligned ESTs. *Genome Research*, **11**, 889-900.

Kim, D.G., Kang, H.M., Jang, S.K. and Shin, H.S. (1992) Construction of a bifunctional mRNA in the mouse by using the internal ribosome entry site of the encephalomyocarditis virus. *Molecular and Cellular Biology*, **12**, 3636-3643.

Kim, J.W., Closs, E.I., Albritton, L.M. and Cunningham, J.M. (1991) Transport of cationic amino acids by the mouse ecotropic retrovirus receptor. *Nature*, **352**, 725-728.

Kimura, S., Niwa, H., Moriyama, M., Araki, K., Abe, K., Miike, T. and Yamamura, K.-I. (1994) Improvement of germ line transmission by targeting beta-galactosidase to nuclei in transgenic mice. *Development Growth and Differentiation*, **36**, 521-527.

Komada, M., McLean, D.J., Griswold, M.D., Russell, L.D. and Soriano, P. (2000) E-MAP-115, encoding a microtubule-associated protein, is a retinoic acid-inducible gene required for spermatogenesis. *Genes and Development*, **14**, 1332-1342.

Kotani, H., Newton, P.B., 3rd, Zhang, S., Chiang, Y.L., Otto, E., Weaver, L., Blaese, R.M., Anderson, W.F. and McGarrity, G.J. (1994) Improved methods of retroviral vector transduction and production for gene therapy. *Hum Gene Ther*, **5**, 19-28.

Kothary, R., Clapoff, S., Brown, A., Campbell, R., Peterson, A. and Rossant, J. (1988) A transgene containing *lacZ* inserted into the *dystonia* locus is expressed in neural tube. *Nature*, **335**, 435-437.

Lakso, M., Sauer, B., Mosinger, B., Jr., Lee, E.J., Manning, R.W., Yu, S.H., Mulder, K.L. and Westphal, H. (1992) Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc Natl Acad Sci U S A*, **89**, 6232-6236.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey,

J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., Szustakowki, J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S. and Chen, Y.J. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

Lewis, P.F. and Emerman, M. (1994) Passage through Mitosis Is Required for Oncoretroviruses but Not for Human Immunodeficiency Virus. *Journal of Virology*, **68**, 510-516.

Lojda, Z. (1970) Indigogenic methods for glycosidases. II. An improved method for beta-D-galactosidase and its application to localization studies of the enzymes in the intestine and in other tissues. *Histochemie*, **23**, 266-288.

MacGregor, G.R., Mogg, A.E., Burke, J.F. and Caskey, C.T. (1987) Histochemical Staining of Clonal Mammalian Cell Lines Expressing *E. coli* β Galactosidase Indicates Heterogeneous Expression of the Bacterial Gene. *Somatic Cell and Molecular Genetics*, **13**, 253-265.

Mann, R., Mulligan, R.C. and Baltimore, D. (1983) Construction of a Retrovirus Packaging Mutant and Its Use to Produce Helper-Free Defective Retrovirus. *Cell*, **33**, 153-159.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912-1934.

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

Miller, A.D. (1990) Retrovirus Packaging Cells. *Human Gene Therapy*, **1**, 5-14.

Miller, A.D. and Buttimore, C. (1986) Redesign of Retrovirus Packaging Cell Lines To Avoid Recombination Leading to Helper Virus Production. *Molecular and Cellular Biology*, **6**, 2895-2902.

Miller, A.D., Law, M.-F. and Verma, I.M. (1985) Generation of Helper-Free Amphotropic Retroviruses That Transduce a Dominant-Acting, Methotrexate-Resistant Dihydrofolate Reductase Gene. *Molecular and Cellular Biology*, **5**, 431-437.

Miller, A.D. and Rosman, G.J. (1989) Improved Retroviral Vectors for Gene Transfer and Expression. *BioTechniques*, **7**, 980-990.

Miller, A.D., Trauber, D.R. and Buttimore, C. (1986) Factors Involved in Production of Helper Virus-Free Retrovirus Vectors. *Somatic Cell and Molecular Genetics*, **12**, 175-183.

Mills, A.A. and Bradley, A. (2001) From mouse to man: generating megabase chromosome rearrangements. *Trends Genet*, **17**, 331-339.

Mironov, A.A., Fickett, J.W. and Gelfand, M.S. (1999) Frequent alternative splicing of human genes. *Genome Res*, **9**, 1288-1293.

Mitchell, K.J., Pinson, K.I., Kelly, O.G., Brennan, J., Zupicich, J., Scherz, P., Leighton, P.A., Goodrich, L.V., Lu, X., Avery, B.J., Tate, P., Dill, K., Pangilinan, E., Wakenight, P., Tessier-Lavigne, M. and Skarnes, W.C. (2001) Functional Analysis of Secreted and Transmembrane Proteins Critical to Mouse Development. *Nature Genetics*, **28**, 241-249.

Modrek, B. and Lee, C. (2002) A genomic view of alternative splicing. *Nat Genet*, **30**, 13-19.

Mountford, P.S. and Smith, A.G. (1995) Internal Ribosome Entry Sites and Dicistronic RNAs in Mammalian Transgenesis. *Trends in Genetics*, **11**, 179-184.

Nagy, A., Perrimon, N., Sandmeyer, S. and Plasterk, R. (2003) Tailoring the genome: the power of genetic approaches. *Nat Genet*, **33 Suppl**, 276-284.

Neilan, E.G. and Barsh, G.S. (1999) Gene Trap Insertional Mutagenesis in Mice: New Vectors and Germ Line Mutations in Two Novel Genes. *Transgenic Research*, **8**, 451-458.

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

Nielson, K.B., Schoettlin, W., Bauer, J.C. and Mathur, E. (1994) *Taq* extender PCR Additive for improved length, yield and reliability of PCR product. *Strategies*, **7**, 27-29.

Niwa, H., Araki, K., Kimura, S., Taniguchi, S., Wakasugi, S. and Yamamura, K. (1993) An Efficient Gene-Trap Method Using Poly A Trap Vectors and Characterization of Gene-Trap Events. *Journal of Biochemistry*, **113**, 343-349.

Niwa, H., Burdon, T., Chambers, I. and Smith, A. (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev*, **12**, 2048-2060.

Nolan, P.M., Peters, J., Strivens, M., Rogers, D., Hagan, J., Spurr, N., Gray, I.C., Vizor, L., Brooker, D., Whitehill, E., Washbourne, R., Hough, T., Greenaway, S., Hewitt, M., Liu, X., McCormack, S., Pickford, K., Selley, R., Wells, C., Tymowska-Lalanne, Z., Roby, P., Glenister, P., Thornton, C., Thaung, C., Stevenson, J.A., Arkell, R., Mburu, P., Hardisty, R., Kiernan, A., Erven, A., Steel, K.P., Voegeling, S., Guenet, J.L., Nickols, C., Sadri, R., Nasse, M., Isaacs, A., Davies, K., Browne, M., Fisher, E.M., Martin, J., Rastan, S., Brown, S.D. and Hunter, J. (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet*, **25**, 440-443.

O'Kane, C.J. and Gehring, W.J. (1987) Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, **84**, 9123-9127.

Oliver, S.G. (1996) From DNA sequence to biological function. *Nature*, **379**, 597-600.

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

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

Palu, G., Parolin, C., Takeuchi, Y. and Pizzato, M. (2000) Progress with retroviral gene vectors. *Reviews in Medical Virology*, **10**, 185-202.

Pear, W.S., Nolan, G.P., Scott, M.L. and Baltimore, D. (1993) Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl. Acad. Sci. USA*, **90**, 8392-8396.

Reithel, F.J. and Kim, J.C. (1960) Studies on the beta-galactosidase isolated from Escherichia coli ML 308. 1. The effect of some ions on enzymic activity. *Arch Biochem Biophys*, **90**, 271-277.

Robertson, E. (1987) *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*. IRL Press Ltd, Oxford.

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

Roe, T., Reynolds, T.C., Yu, G. and Brown, P.O. (1993) Integration of murine leukemia virus DNA depends on mitosis. *Embo J*, **12**, 2099-2108.

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. and Breindl, M. (1987) Retrovirus Integration and Chromatin Structure: Moloney Murine Leukemia Proviral Integration Sites Map Near DNase I-Hypersensitive Sites. *Journal of Virology*, **61**, 336-343.

Salminen, M., Meyer, B.I. and Gruss, P. (1998) Efficient Poly A Trap Approach Allows the Capture of Genes Specifically Active in Differentiated Embryonic Stem Cells and in Mouse Embryos. *Developmental Dynamics*, **212**, 326-333.

Sambrook, J. and Gething, M.J. (1989) Chaperones, paperones. *Nature*, **342**, 224-225.

Scherdin, U., Rhodes, K. and Breindl, M. (1990) Transcriptionally Active Genome Regions Are Preferred Targets for Retrovirus Integration. *Journal of Virology*, **64**, 907-912.

Schweizer, M., Roberts, L.M., Holtke, J., Takabayashi, K., Muller, G., Hoja, U., Hollerer, E., Schuh, B. and Scheizer, E. (1986) Molecular structure and expression of fatty acid synthetase genes in yeast. *Biochem Soc Trans*, **14**, 572-574.

Shirai, M., Miyashita, A., Ishii, N., Itoh, Y., Satokata, I., Watanabe, Y.G. and Kuwano, R. (1996) A Gene Trap Strategy for Identifying the Gene Expressed in the Embryonic Nervous System. *Zoological Science*, **13**, 277-283.

Skarnes, W.C., Auerbach, B.A. and Joyner, A.L. (1992) A gene trap approach in mouse embryonic stem cells: the *lac Z* reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes and Development*, **6**, 903-918.

Skarnes, W.C., Moss, J.E., Hurtley, S.M. and Beddington, R.S.P. (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl. Acad. Sci. USA*, **92**, 6592-6596.

Soriano, P. (1995) Gene Targeting in ES Cells. *Annual Reviews in Neuroscience*, **18**, 1-18.

Soriano, P., Friedrich, G. and Lawinger, P. (1991) Promoter Interactions in Retrovirus Vectors Introduced into Fibroblasts and Embryonic Stem Cells. *Journal of Virology*, **65**, 2314-2319.

Speck, N.A. and Baltimore, D. (1987) Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol Cell Biol*, **7**, 1101-1110.

Spence, S.E., Gilbert, D.J., Swing, D.A., Copeland, N.G. and Jenkins, N.A. (1989) Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic Srev lines of mice. *Mol Cell Biol*, **9**, 177-184.

Stanford, W., Cohn, J. and Cordes, S. (2001) Gene Trap Mutagenesis: Past, Present and Beyond. *Nature Reviews Genetics*, **2**, 756-768.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L. and et al. (1992) The C. elegans genome sequencing project: a beginning. *Nature*, **356**, 37-41.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J. and Higashinakagawa, T. (1995) Gene trap capture of a novel mouse gene, *jumonji*, required for neural tube formation. *Genes and Development*, **9**, 1211-1222.

Temin, H.M. (1981) Structure, Variation and Synthesis of Retrovirus Long Terminal Repeat. *Cell*, **27**, 1-3.

Temin, H.M. (1982) Function of the Retrovirus Long Terminal Repeat. *Cell*, **28**, 3-5.

Thomas, T., Voss, A.K., Chowdhury, K. and Gruss, P. (2000) A New Gene Trap Construct Enriching for Insertion Events Near the 5' End of Genes. *Transgenic Research*, **9**, 395-404.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S. and Jones, J.M. (1998) Embryonic stem cell lines derived from human blastocysts. *Science*, **282**, 1145-1147.

Toyoshima, K. and Vogt, P.K. (1969) Enhancement and inhibition of avian sarcoma viruses by polycations and polyanions. *Virology*, **38**, 414-426.

Tsukiyama, T., Niwa, O. and Yokoro, K. (1989) Mechanism of suppression of the long terminal repeat of Moloney leukemia virus in mouse embryonal carcinoma cells. *Mol Cell Biol*, **9**, 4670-4676.

Varmus, H.E. (1982) Form and Function of Retroviral Proviruses. *Science*, **216**, 812-820.

Varmus, H.E., Heasley, S., Kung, H.J., Oppermann, H., Smith, V.C., Bishop, J.M. and Shank, P.R. (1978) Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. *J Mol Biol*, **120**, 55-82.

Vega, F.M., Gonzalo, P., Gaspar, M.L. and Lazo, P.A. (2003) Expression of the VRK (vaccinia-related kinase) gene family of p53 regulators in murine hematopoietic development. *FEBS Lett*, **544**, 176-180.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei,

Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001) The sequence of the human genome. *Science*, **291**, 1304-1351.

Vijaya, S., Steffen, D.L. and Robinson, H.L. (1986) Acceptor Sites for Retroviral Integrations Map Near DNase I-Hypersensitive Sites in Chromatin. *Journal of Virology*, **60**, 683-692.

von Melchner, H., DeGregori, J.V., Rayburn, H., Reddy, S., Friedel, C. and Ruley, H.E. (1992) Selective Disruption of Genes Expressed in Totipotent Embryonal Stem Cells. *Genes and Development*, **6**, 919-927.

Wang, H., Kavanaugh, M.P., North, R.A. and Kabat, D. (1991) Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter. *Nature*, **352**, 729-731.

Ware, L.M. and Axelrad, A.A. (1972) Inherited Resistance to N- and B-tropic Murine Leukemia Viruses *In Vitro*: Evidence that Congenic Mouse Strains SIM and SIM.R Differ at the Fv-1 Locus. *Virology*, **50**, 339-348.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S.E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M.R., Brown, D.G., Brown, S.D., Bult, C., Burton, J., Butler, J., Campbell, R.D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A.T., Church, D.M., Clamp, M., Clee, C., Collins, F.S., Cook, L.L., Copley, R.R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K.D., Deri, J., Dermitzakis, E.T., Dewey, C., Dickens, N.J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D.M., Eddy, S.R., Elnitski, L., Emes, R.D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G.A., Flicek, P., Foley, K., Frankel, W.N., Fulton, L.A., Fulton, R.S., Furey, T.S., Gage, D., Gibbs, R.A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T.A., Green, E.D., Gregory, S., Guigo, R., Guyer, M., Hardison, R.C., Haussler, D., Hayashizaki, Y., Hillier, L.W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D.B., Johnson, L.S., Jones, M., Jones, T.A., Joy, A., Kamal, M., Karlsson, E.K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W.J., Kirby, A., Kolbe, D.L., Korf, I., Kucherlapati, R.S., Kulbokas, E.J., Kulp, D., Landers, T., Leger, J.P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D.R., Mardis, E.R., Matthews, L., Mauceli, E., Mayer, J.H., McCarthy, M., McCombie, W.R., McLaren, S., McLay, K., McPherson, J.D., Meldrim, J., Meredith, B., Mesirov, J.P., Miller, W., Miner, T.L., Mongin, E., Montgomery, K.T., Morgan, M., Mott, R., Mullikin, J.C., Muzny, D.M., Nash, W.E., Nelson, J.O., Nhan, M.N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M.J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K.H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C.S., Poliakov, A., Ponce, T.C., Ponting, C.P., Potter, S., Quail, M., Reymond, A., Roe, B.A., Roskin, K.M., Rubin, E.M., Rust, A.G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M.S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J.B., Slater, G., Smit, A., Smith, D.R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J.P., Von Niederhausern, A.C., Wade, C.M., Wall, M., Weber, R.J., Weiss, R.B., Wendl, M.C., West, A.P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R.K., Winter, E., Worley, K.C., Wyman, D., Yang, S., Yang, S.P., Zdobnov, E.M., Zody, M.C. and Lander, E.S. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520-562.

Weber, F., de Villiers, J. and Schaffner, W. (1984) An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. *Cell*, **36**, 983-992.

Weiss, R. (1998) Viral RNA-dependent DNA polymerase RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Rev Med Virol*, **8**, 3-11.

Weiss, R.A. and Tailor, C.S. (1995) Retrovirus Receptors. *Cell*, **82**, 531-533.

Westerman, K.A. and Leboulch, P. (1996) Reversible immortalization of mammalian cells mediated by retroviral transfer and site specific recombination. *Proc. Natl. Acad. Sci. USA*, **93**, 8971-8976.

Wiles, M.V., Vauti, F., Otte, J., Fuchtbauer, E.M., Ruiz, P., Fuchtbauer, A., Arnold, H.H., Lehrach, H., Metz, T., von Melchner, H. and Wurst, W. (2000) Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. *Nat Genet*, **24**, 13-14.

Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A.L., Hill, D.P., Guillemot, F., Gasca, S., Cado, D., Auerbach, A. and Ang, S.-L. (1995) A Large-Scale Gene-Trap Screen for Insertional Mutations in Developmentally Regulated Genes in Mice. *Genetics*, **139**, 889-899.

Xu, C., Inokuma, M.S., Denham, J., Golds, K., Kundu, P., Gold, J.D. and Carpenter, M.K. (2001) Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnology*, **19**, 971-974.

Yoshida, M., Yagi, T., Furuta, Y., Takayanagi, K., Kominami, R., Takeda, N., Tokunaga, T., Chiba, J., Ikawa, Y. and Aizawa, S. (1995) A new strategy of gene trapping in ES cells using 3'RACE. *Transgenic Research*, **4**, 277-287.

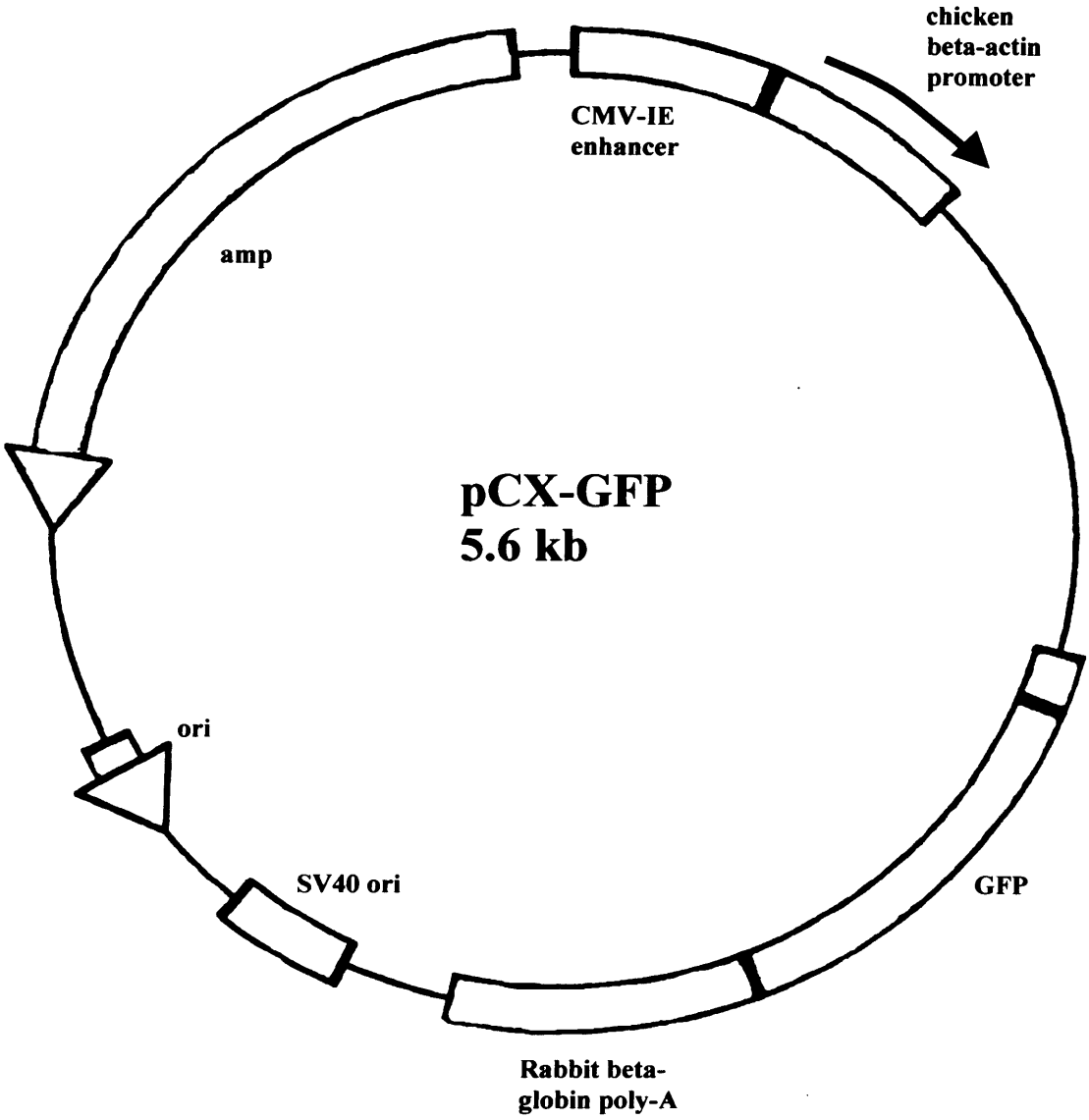
Yu, Y. and Bradley, A. (2001) Engineering Chromosomal Rearrangements in Mice. *Nature Reviews Genetics*, **2**, 780-790.

Zambrowicz, B.P., Friedrich, G.A., Buxton, E.C., Lilleberg, S.L., Person, C. and Sands, A.T. (1998) Disruption and sequence identification of 2000 genes in mouse embryonic stem cells. *Nature*, **392**, 608-611.

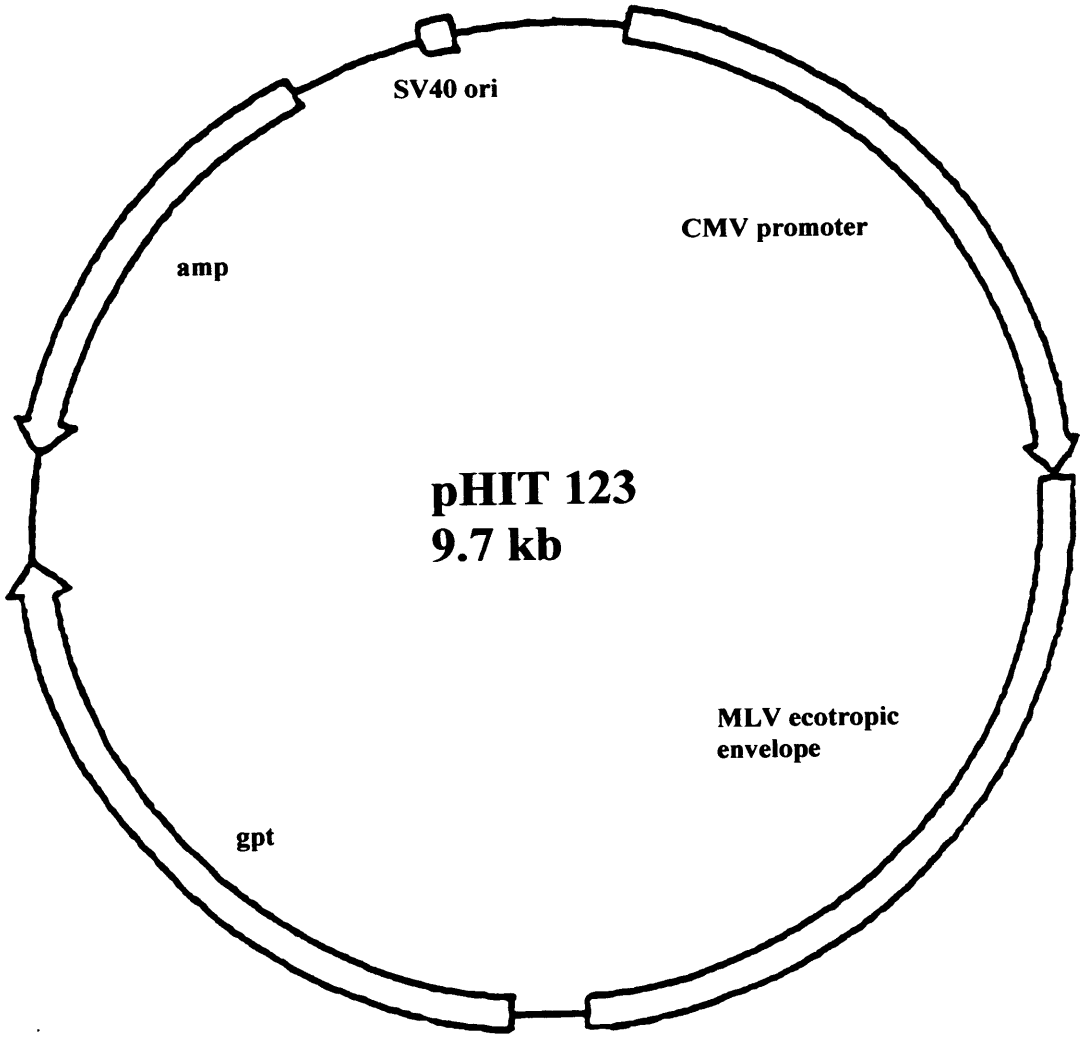
APPENDIX I

PLASMIDS USED

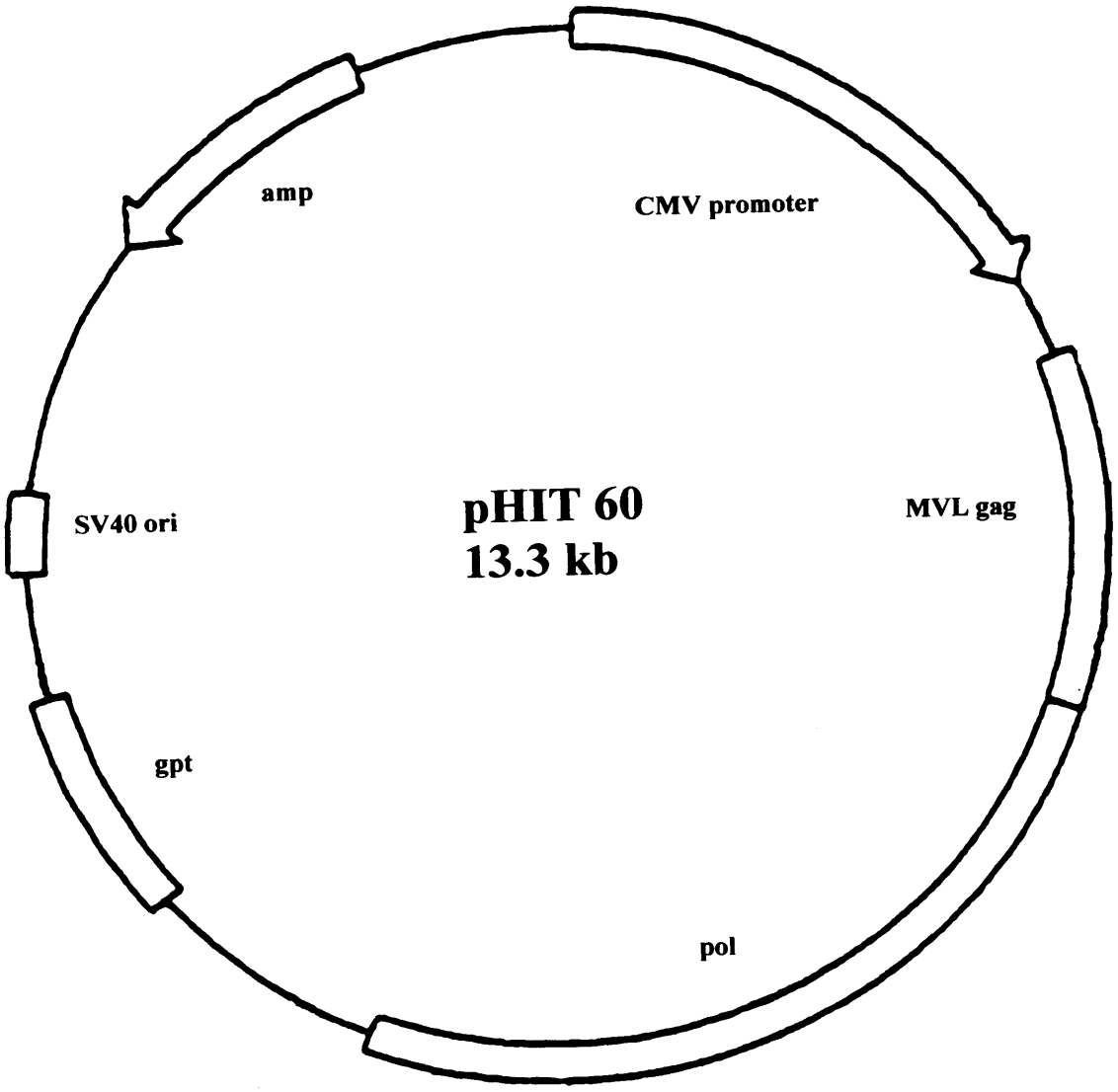
pCX-GFP



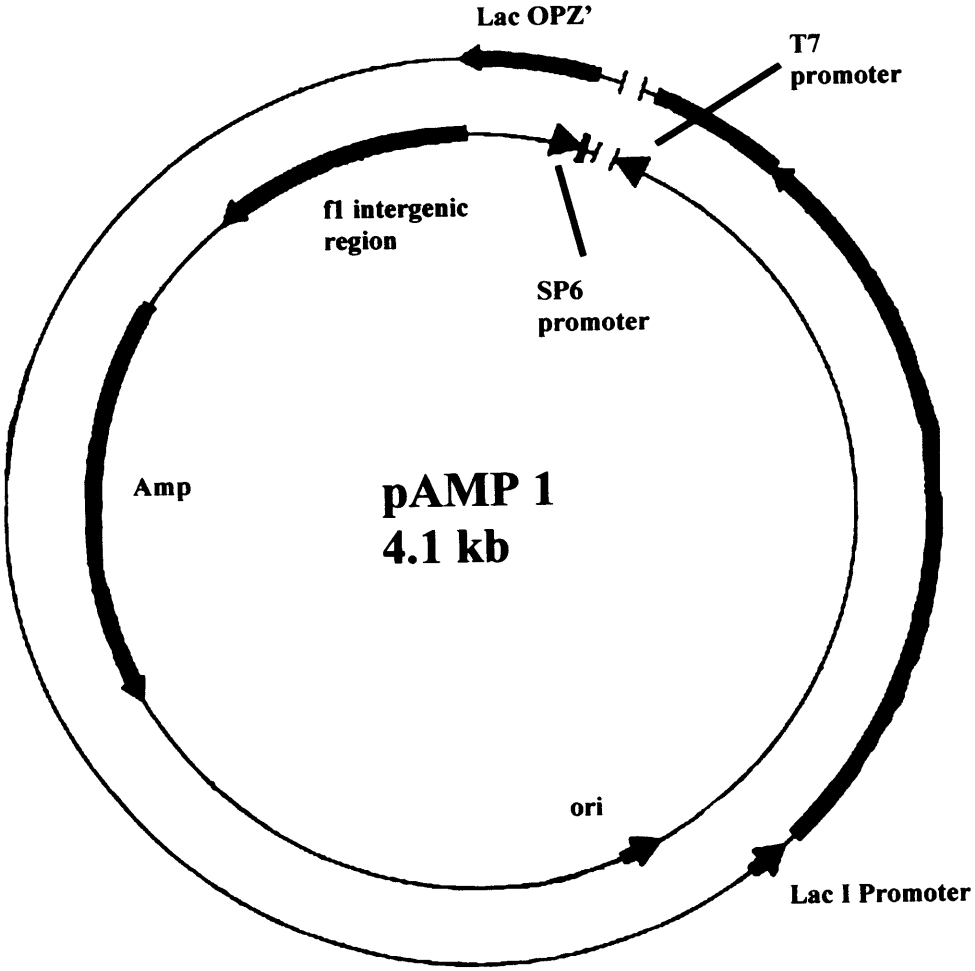
pHIT 123


SV40 ori
CMV promoter
amp
pHIT 123
9.7 kb
MLV ecotropic
envelope
gpt

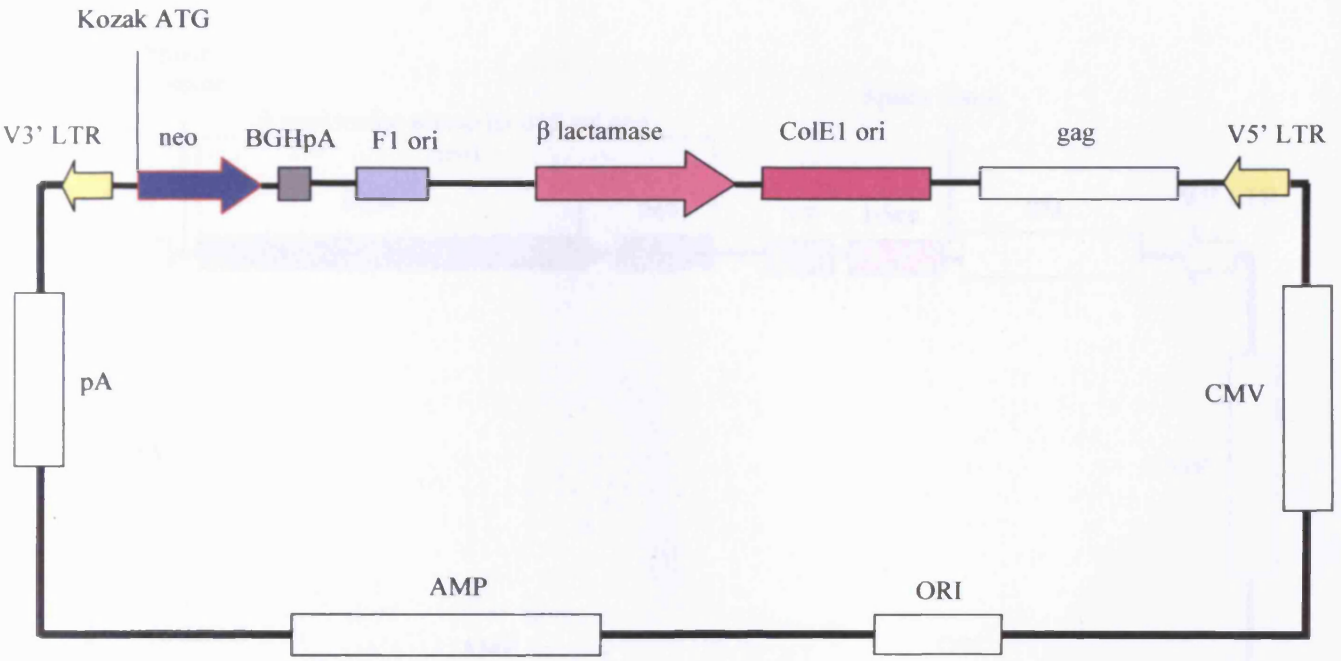
pHIT 60



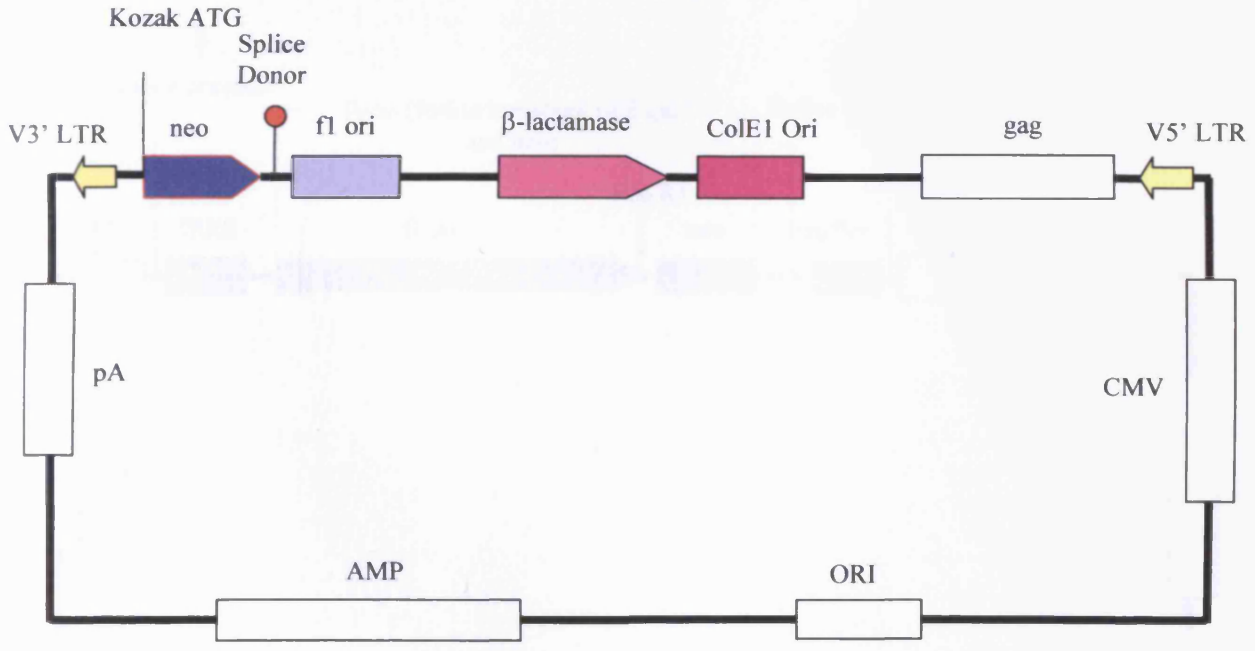
pAMP 1


Lac OPZ'
T7
promoter
f1 intergenic
region
SP6
promoter
Amp
pAMP 1
4.1 kb
ori
Lac I Promoter

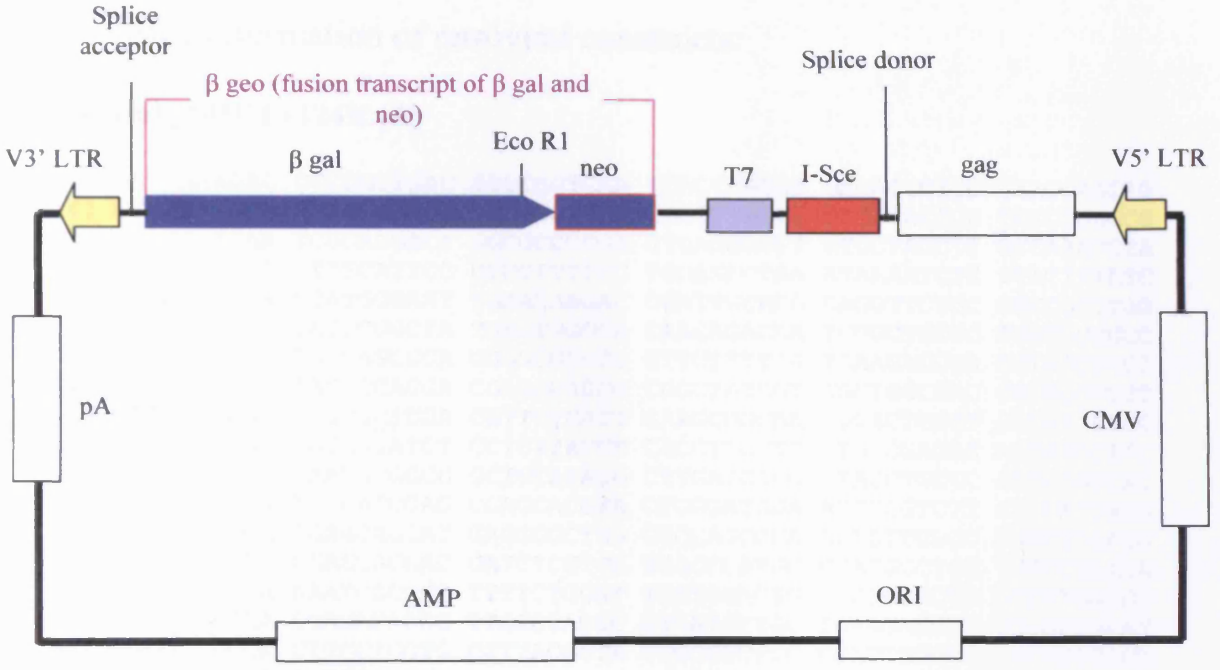
pSHUT + 124K (A)



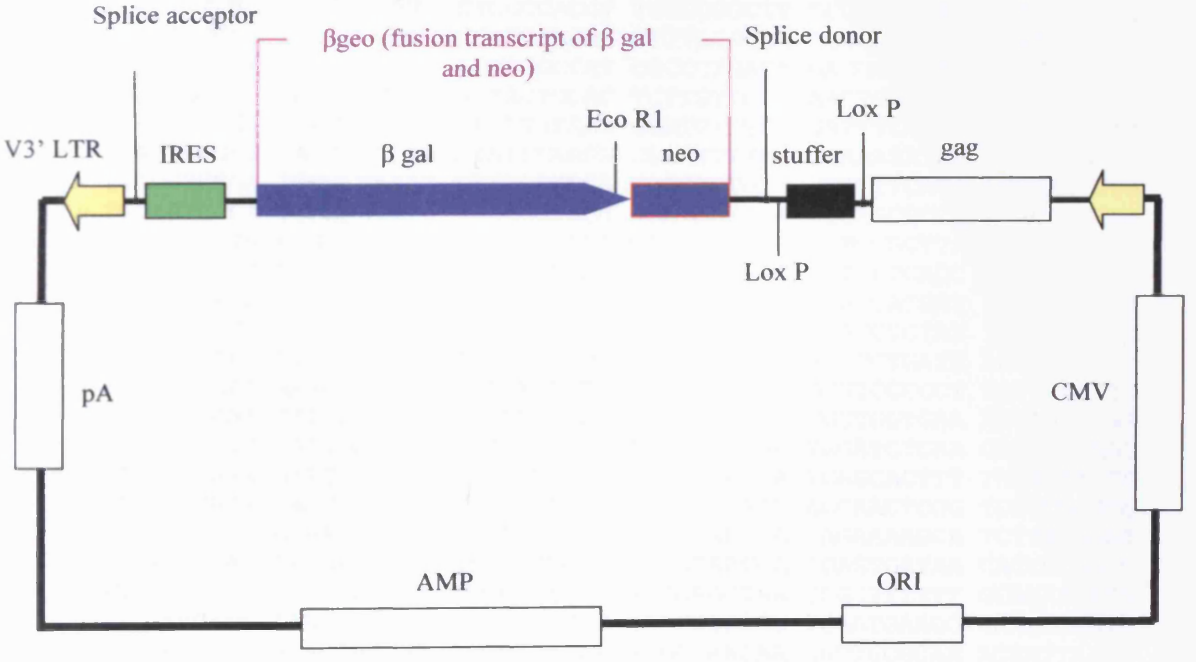
pSHUT + 124K SD126 (1)



pCRR + SD126 T7 I-Sce (X)



pCSI + SD126 L4 (2)



APPENDIX II

Sequence information of retroviral constructs:

proviral pSHUT+124K (A)

```
   1 AATGAAAGAC CCCCGCTGAC GGGTAGTCAA TCACTCAGAG GAGACCCTCC CAAGGAACAG
  61 CGAGACCACA AGTCGGATGC AACTGCAAGA GGGTTTATTG GATACACGGG TACCCGGGCG
 121 ACTCAGTCAA TCGGAGGACT GGCGCCCCGA GTGAGGGGTT GTGCTAGCTT GCCAAACCTA
 181 CAGGTGGGGT CTTTCATTCC CCCCTTTTTC TGGAGACTAA ATAAAATCTT TTATTTTATC
 241 GATAAGCTTA CCATGGGAAT TGAACAAGAC GGATTGCACG CAGGTTCTCC GGCCGCTTGG
 301 GTGGAGAGGC TATTCGGCTA TGACTGGGCA CAACAGACAA TCGGCTGCTC TGATGCCGCC
 361 GTGTTCCGGC TGTCAGCGCA GGGGCGCCCG GTTCTTTTTG TCAAGACCGA CCTGTCCGGT
 421 GCCCTGAATG AACTGCAGGA CGAGGCAGCG CGGCTATCGT GGCTGGCCAC GACGGGCGTT
 481 CCTTGCGCAG CTGTGCTCGA CGTTGTCACT GAAGCGGGAA GGGACTGGCT GCTATTGGGC
 541 GAAGTGCCGG GGCAGGATCT CCTGTCATCT CACCTTGCTC CTGCCGAGAA AGTATCCATC
 601 ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG CTACCTGCCC ATTCGACCAC
 661 CAGCGAAACA TCGCATCGAG CGAGCACGTA CTCGGATGGA AGCCGGTCTT GTCGATCAGG
 721 ATGATCTGGA CGAAGAGCAT CAGGGGCTCG CGCCAGCCGA ACTGTTCGCC AGGCTCAAGG
 781 CGCGCATGCC CGACGGCGAG GATCTCGTCG TGACCCATGG CGATGCCTGC TTGCCGAATA
 841 TCATGGTGGA AAATGGCCGC TTTTCTGGAT TCATCGACTG TGGCCGGCTG GGTGTGGCGG
 901 ACCGCTATCA GGACATAGCG TTGGCTACCC GTGATATTGC TGAAGAGCTT GGCGGCGAAT
 961 GGGCTGACCG CTTCCTCGTG CTTTACGGTA TCGCCGCTCC CGATTCGCAG CGCATCGCCT
1021 TCTATCGCCT TCTTGACGAG TTCTTCTGAT CTAGAGAGCT CAGCTGTGCC TTCTAGTTGC
1081 CAGCCATCTG TTGTTTGCCC CTCCCCCGTG CCTTCCTTGA CCCTGGAAGG TGCCACTCCC
1141 ACTGTCCTTT CCTAATAAAA TGAGGAAATT GCATCGCATT GTCTGAGTAG GTGTCATTCT
1201 ATTCTGGGGG GTGGGGTGGG GCAGGACAGC AAGGGGGAGG ATTGGGAAGA CAATAGCAGG
1261 CATGCTGGGG ATGCGGTGGG CTCTATGGCT TCTGAGGCGG AAAGAACGTC GACGAATTCG
1321 AGCTCGGCCG ACTTGGCCAA TTCGCCCTAT AGTGAGTCGT ATTACAATTC ACTGGCCGTC
1381 GTTTTACAAC GTCGTGACTG GGAAAACCCT GGCGTTACCC AACTTAATCG CCTTGCAGCA
1441 CATCCCCCTT TCGCCAGCTG GCGTAATAGC GAAGAGGCCC GCACCGATCG CCCTTCCCAA
1501 CAGTTGCGCA GCCTGAATGG CGAATGGACG CGCCCTGTAG CGGCGCATTA AGCGCGGCGG
1561 GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT
1621 TCGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC
1681 GGGGGCTCCC TTTAGGGTTC CGATTTAGAG CTTTACGGCA CCTCGACCGC AAAAAACTTG
1741 ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA
1801 CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTTCCA AACTGGAACA ACACTCAACC
1861 CTATCTCGGT CTATTCTTTT GATTTATAAG GGATTTTGCC GATTTCGGCC TATTGGTTAA
1921 AAAATGAGCT GATTTAACAA ATATTTAACG CGAATTTTAA CAAAATATTA ACGTTTACAA
1981 TTTCGCCTGA TGCGGTATTT TCTCCTTACG CATCTGTGCG GTATTTCACA CCGCATATGG
2041 TGCACTCTCA GTACAATCTG CTCTGATGCC GCATAGTTAA GCCAGCCCCG ACACCCGCCA
2101 ACACCCGCTG ACGCGCCCTG ACGGGCTTGT CTGCTCCCGG CATCCGCTTA CAGACAAGCT
2161 GTGACCGTCT CCGGGAGCTG CATGTGTCAG AGGTTTTCAC CGTCATCACC GAAACGCGCG
2221 AGACGAAAGG GCCTCGTGAT ACGCCTATTT TTATAGGTTA ATGTCATGAT AATAATGGTT
2281 TCTTAGACGT CAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT
2341 TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCCTGATA AATGCTTCAA
2401 TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTCGCCCT TATTCCCTTT
2461 TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT
2521 GCTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGGTAAG
2581 ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TTTCCAATGA TGAGCACTTT TAAAGTTCTG
2641 CTATGTCATA CACTATTATC CCGTATTGAC GCCGGGCAAG AGCAACTCGG TCGCCGGGCG
2701 CGGTATTCTC AGAATGACTT GGTTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT
2761 GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC
2821 AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT GCACAACATG
2881 GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC
2941 GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGCCAACAA CGTTGCGCAA ACTATTAACT
3001 GGCGAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA GGCGGATAAA
3061 GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATAAATCT
```

3121 GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC
3181 TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA
3241 CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT AACTGTCAGA CCAAGTTTAC
3301 TCATATATAC TTTAGATTGA TTTAAAACTT CATTTTTAAT TTAAAAGGAT CTAGGTGAAG
3361 ATCCTTTTTG ATAATCTCAT GACCAAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG
3421 TCAGACCCCG TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTTCT GCGCGTAATC
3481 TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG
3541 CTACCAACTC TTTTTCCGAA GGTAACTGGC TTCAGCAGAG CGCAGATACC AAATACTGTC
3601 CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC GCCTACATAC
3661 CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC
3721 GGGTTGGACT CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTCGGGCTG AACGGGGGGT
3781 TCGTGCACAC AGCCCAGCTT GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT
3841 GAGCTATGAG AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC
3901 GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT
3961 TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA
4021 GGGGGGCGGA GCCTATCGAA AAACGCCAGC AACGCGGCCT TTTTACGGTT CCTGGCCTTT
4081 TGCTGGCCTT TTGCTCACAT GTTCTTTCCT GCGTTATCCC CTGATTCTGT GGATAACCGT
4141 ATTACCGCCT TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG
4201 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC CGCCTCTCCC CGCGCGTTGG
4261 CCGATTCATT AATGCAGCTG GCACGACAGG TTTCCCGACT GGAAAGCGGG CAGTGAGCGC
4321 AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTTATGCTT
4381 CCGGCTCGTA TGTTGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT
4441 GACCATGATT ACGCCAAGCT ATTTAGGTGA CACTATAGAA TACTCAAGCT TATGCATGCG
4501 GCCGCATCTA GCGAATTCGG CGCCTAGAGA AGGAGTGAGG GCTGGATAAA GGGAGGATCG
4561 AGGCGGGGTC GAACGAGGAG GTTCAAGGGG GAGAGACGGG GCGGATGGAG GAAGAGGAGG
4621 CGGAGGCTTA GGGTGTACAA AGGGCTTGAC CCAGGGAGGG GGGTCAAAAG CCAAGGCTTC
4681 CCAGGTCACG ATGTAGGGGA CCTGGTCTGG GTGTCCATGC GGGCCAGGTG AAAAGACCTT
4741 GATCTTAACC TGGGTGATGA GGTCTCGGTT AAAGGTGCCG TCTCGCGGCC ATCCGACGTT
4801 AAAGGTTGGC CATTCTGCAG AGCAGAAGGT AACCCAACGT CTCTTCTTGA CATCTACCGA
4861 CTGGTTGTGA GCGATCCGCT CGACATCTTT CCAGTGACCT AAGGTCAAAC TTAAGGGAGT
4921 GGTAACAGTC TGGCCCATAT TCTCAGACAA ATACAGAAAC ACAGTCAGAC AGAGACAACA
4981 CAGAACGATG CTGCAGCAGA CAAGACGCGC GGCGCGGCTT CGGTCCCAAA CCGAAAGCAA
5041 AAATTCAGAC GGAGGCGGGA ACTGTTTTAG GTTCTCGTCT CCTACCAGAA CCACATATCC
5101 CTCCTCTAAG GGGGGTGCAC CAAAGAGTCC AAAACGATCG GGATTTTTGG ACTCAGGTCG
5161 GGCCACAAAA ACGGCCCCCG AAGTCCCTGG GACGTCTCCC AGGGTTGCGG CCGGGTGTTC
5221 CGAACTCGTC AGTTCCACCA CGGGTCCGCC AGATACAGAG CTAGTTAGCT AACTAGTACC
5281 GACGCAGGCG CATAAAATCA GTCATAGACA CTAGACAATC GGACAGACAC AGATAAGTTG
5341 CTGGCCAGCT TACCTCCCGG TGGTGGGTCG GTGGTCCCTG GGCAGGGGTC TCCCGATCCC
5401 GGACGAGCCC CCAAATGAAA GACCCCCGCT GACGGGTAGT CAATCACTCA GAGGAGACCC
5461 TCCCAAGGAA CAGCGAGACC ACAAGTCGGA TGCAACTGCA AGAGGGTTTA TTGGATACAC
5521 GGGTACCCGG GCGACTCAGT CAATCGGAGG ACTGGCGCCC CGAGTGAGGG GTTGTGCTAG
5581 ATAACTTCGT ATAGTATACA TTATACGAAG TTATCTAGCT TGCCAAACCT ACAGGTGGGG
5641 TCTTTCATT

proviral pSHUT+124k SD126 (1)

```
   1 AATGAAAGAC CCCCGCTGAC GGGTAGTCAA TCACTCAGAG GAGACCCTCC CAAGGAACAG
  61 CGAGACCACA AGTCGGATGC AACTGCAAGA GGGTTTATTG GATACACGGG TACCCGGGCG
 121 ACTCAGTCAA TCGGAGGACT GGCGCCCCGA GTGAGGGGTT GTGCTAGCTT GCCAAACCTA
 181 CAGGTGGGGT CTTTCATTCC CCCCTTTTTC TGGAGACTAA ATAAAATCTT TTATTTTATC
 241 GATAAGCTTA CCATGGGAAT TGAACAAGAC GGATTGCACG CAGGTTCTCC GGCCGCTTGG
 301 GTGGAGAGGC TATTCGGCTA TGACTGGGCA CAACAGACAA TCGGCTGCTC TGATGCCGCC
 361 GTGTTCCGGC TGTCAGCGCA GGGGCGCCCG GTTCTTTTTG TCAAGACCGA CCTGTCCGGT
 421 GCCCTGAATG AACTGCAGGA CGAGGCAGCG CGGCTATCGT GGCTGGCCAC GACGGGCGTT
 481 CCTTGCGCAG CTGTGCTCGA CGTTGTCACT GAAGCGGGAA GGGACTGGCT GCTATTGGGC
 541 GAAGTGCCGG GGCAGGATCT CCTGTCATCT CACCTTGCTC CTGCCGAGAA AGTATCCATC
 601 ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG CTACCTGCCC ATTCGACCAC
 661 CAAGCGAAAC ATCGCATCGA GCGAGCACGT ACTCGGATGG AAGCCGGTCT TGTCGATCAG
 721 GATGATCTGG ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG AACTGTTCGC CAGGCTCAAG
 781 GCGCGCATGC CCGACGGCGA GGATCTCGTC GTGACCCATG GCGATGCCTG CTTGCCGAAT
 841 ATCATGGTGG AAAATGGCCG CTTTTCTGGA TTCATCGACT GTGGCCGGCT GGGTGTGGCG
 901 GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG CTGAAGAGCT TGGCGGCGAA
 961 TGGGCTGACC GCTTCCTCGT GCTTTACGGT ATCGCCGCTC CCGATTCGCA GCGCATCGCC
1021 TTCTATCGCC TTCTTGACGA GTTCTTCTGA TCTAGAACTA GTGGATCCAG CTTATCGAGC
1081 TTGCCAACTA CAGCAGGTAA GCATATTTAT GAAAAATGCT GATTGTGTTA GCTACTTGTG
1141 TCAGTGTTGT GATAGTCGAC GAATTCGAGC TCGGCCGACT TGGCCAATTC GCCCTATAGT
1201 GAGTCGTATT ACAATTCACT GGCCGTCGTT TTACAACGTC GTGACTGGGA AAACCCTGGC
1261 GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGAA
1321 GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGACGCGC
1381 CCTGTAGCGG CGCATTAAGC GCGGCGGGTG TGGTGGTTAC GCGCAGCGTG ACCGCTACAC
1441 TTGCCAGCGC CCTAGCGCCC GCTCCTTTCG CTTTCTTCCC TTCCTTTCTC GCCACGTTCG
1501 CCGGCTTTCC CCGTCAAGCT CTAAATCGGG GGCTCCCTTT AGGGTTCCGA TTTAGAGCTT
1561 TACGGCACCT CGACCGCAAA AAACTTGATT TGGGTGATGG TTCACGTAGT GGGCCATCGC
1621 CCTGATAGAC GGTTTTTCGC CCTTTGACGT TGGAGTCCAC GTTCTTTAAT AGTGGACTCT
1681 TGTTCCAAAC TGGAACAACA CTCAACCCTA TCTCGGTCTA TTCTTTTGAT TTATAAGGGA
1741 TTTTGCCGAT TTCGGCCTAT TGGTTAAAAA ATGAGCTGAT TTAACAAATA TTTAACGCGA
1801 ATTTTAACAA AATATTAACG TTTACAATTT CGCCTGATGC GGTATTTTCT CCTTACGCAT
1861 CTGTGCGGTA TTTCACACCG CATATGGTGC ACTCTCAGTA CAATCTGCTC TGATGCCGCA
1921 TAGTTAAGCC AGCCCCGACA CCCGCCAACA CCCGCTGACG CGCCCTGACG GGCTTGTCTG
1981 CTCCCGGCAT CCGCTTACAG ACAAGCTGTG ACCGTCTCCG GGAGCTGCAT GTGTCAGAGG
2041 TTTTCACCGT CATCACCGAA ACGCGCGAGA CGAAAGGGCC TCGTGATACG CCTATTTTTA
2101 TAGGTTAATG TCATGATAAT AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT
2161 GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG
2221 AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA
2281 CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC
2341 CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC
2401 ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT
2461 CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTCATACAC TATTATCCCG TATTGACGCC
2521 GGGCAAGAGC AACTCGGTCG CCGGGCGCGG TATTCTCAGA ATGACTTGGT TGAGTACTCA
2581 CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC
2641 ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG
2701 GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA
2761 CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG
2821 CCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA
2881 TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG
2941 GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT
3001 GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT
3061 CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG
3121 CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT
3181 TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT
3241 TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT
3301 TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA
3361 GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC
3421 AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC
3481 AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

3541 GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG
3601 GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC
3661 TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA GCGCCACGCT TCCCGAAGGG
3721 AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG
3781 CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT
3841 GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATCGAAAAA CGCCAGCAAC
3901 GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG
3961 TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC
4021 CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCCAATA
4081 CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT
4141 CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT CACTCATTAG
4201 GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA
4261 TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG CCAAGCTATT TAGGTGACAC
4321 TATAGAATAC TCAAGCTTAT GCATGCGGCC GCATCTAGCG AATTCGGCGC CTAGAGAAGG
4381 AGTGAGGGCT GGATAAAGGG AGGATCGAGG CGGGGTCGAA CGAGGAGGTT CAAGGGGGAG
4441 AGACGGGGCG GATGGAGGAA GAGGAGGCGG AGGCTTAGGG TGTACAAAGG GCTTGACCCA
4501 GGGAGGGGGG TCAAAAGCCA AGGCTTCCCA GGTCACGATG TAGGGGACCT GGTCTGGGTG
4561 TCCATGCGGG CCAGGTGAAA AGACCTTGAT CTTAACCTGG GTGATGAGGT CTCGGTTAAA
4621 GGTGCCGTCT CGCGGCCATC CGACGTTAAA GGTTGGCCAT TCTGCAGAGC AGAAGGTAAC
4681 CCAACGTCTC TTCTTGACAT CTACCGACTG GTTGTGAGCG ATCCGCTCGA CATCTTTCCA
4741 GTGACCTAAG GTCAAACTTA AGGGAGTGGT AACAGTCTGG CCCATATTCT CAGACAAATA
4801 CAGAAACACA GTCAGACAGA GACAACACAG AACGATGCTG CAGCAGACAA GACGCGCGGC
4861 GCGGCTTCGG TCCCAAACCG AAAGCAAAAA TTCAGACGGA GGCGGGAACT GTTTTAGGTT
4921 CTCGTCTCCT ACCAGAACCA CATATCCCTC CTCTAAGGGG GGTGCACCAA AGAGTCCAAA
4981 ACGATCGGGA TTTTTGGACT CAGGTCGGGC CACAAAAACG GCCCCCGAAG TCCCTGGGAC
5041 GTCTCCCAGG GTTGCGGCCG GGTGTTCCGA ACTCGTCAGT TCCACCACGG GTCCGCCAGA
5101 TACAGAGCTA GTTAGCTAAC TAGTACCGAC GCAGGCGCAT AAAATCAGTC ATAGACACTA
5161 GACAATCGGA CAGACACAGA TAAGTTGCTG GCCAGCTTAC CTCCCGGTGG TGGGTCGGTG
5221 GTCCCTGGGC AGGGGTCTCC CGATCCCGGA CGAGCCCCCA AATGAAAGAC CCCCGCTGAC
5281 GGGTAGTCAA TCACTCAGAG GAGACCCTCC CAAGGAACAG CGAGACCACA AGTCGGATGC
5341 AACTGCAAGA GGGTTTATTG GATACACGGG TACCCGGGCG ACTCAGTCAA TCGGAGGACT
5401 GGCGCCCCGA GTGAGGGGTT GTGCTAGCTT GCCAAACCTA CAGGTGGGGT CTTTCATT

proviral pCSI+SD126 L4 (2)

```
   1 AATGAAAGAC CCCCGCTGAC GGGTAGTCAA TCACTCAGAG GAGACCCTCC CAAGGAACAG
  61 CGAGACCACA AGTCGGATGC AACTGCAAGA GGGTTTATTG GATACACGGG TACCCGGGCG
 121 ACTCAGTCAA TCGGAGGACT GGCGCCCCGA GTGAGGGGTT GTGCTAGCTT GCCAAACCTA
 181 CAGGTGGGGT CTTTCATTCC CCCCTTTTTC TGGAGACTAA ATAAAATCTT TTATTTTATC
 241 TATGGCTCGT ACTCTATAGG CTTCAGCTGG TGATATTGTT GAGTCAAAAC TAGAGCCTGG
 301 ACCACTGATA TCCTGTCTTT AACAAATTGG ACTAATCGAT CTCGACCTCG AGACTAGTGG
 361 ATCCCCCGGG CTGCAGATCT GTAGGGCGCA GTAGTCCAGG GTTTTCCTTG ATGATGTCAT
 421 ACTTATCCCT GTCCCTTTTT TTTCCACAGC TCGCGGTTGA GGACAAACTC TTCGCGGTCT
 481 TTCCAGTGGG GATCGACGGT ATCGATAAAG CTAATTCCGC CCCCCCCCCC CCCCCCCCCT
 541 CTCCCTCCCC CCCCCCTAAC GTTACTGGCC GAAGCCGCTT GGAATAAGGC CGGTGTGCGT
 601 TTGTCTATAT GTTATTTTCC ACCATATTGC CGTCTTTTGG CAATGTGAGG GCCCGGAAAC
 661 CTGGCCCTGT CTTCTTGACG AGCATTCCTA GGGGTCTTTC CCCTCTCGCC AAAGGAATGC
 721 AAGGTCTGTT GAATGTCGTG AAGGAAGCAG TTCCTCTGGA AGCTTCTTGA AGACAAACAA
 781 CGTCTGTAGC GACCCTTTGC AGGCAGCGGA ACCCCCCACC TGGCGACAGG TGCCTCTGCG
 841 GCCAAAAGCC ACGTGTATAA GATACACCTG CAAAGGCGGC ACAACCCCAG TGCCACGTTG
 901 TGAGTTGGAT AGTTGTGGAA AGAGTCAAAT GGCTCTCCTC AAGCGTATTC AACAAGGGGC
 961 TGAAGGATGC CCAGAAGGTA CCCCATTGTA TGGGATCTGA TCTGGGGCCT CGGTGCACAT
1021 GCTTTACGTG TGTTTAGTCG AGGTTAAAAA ACGTCTAGGC CCCCCGAACC ACGGGGACGT
1081 GGTTTTCCTT TGAAAAACAC GATGATAATA TGGCCATGAG CTTGATGATC TGTGACATGG
1141 CGGATCCCGT CGTTTTACAA CGTCGTGACT GGGAAAACCC TGGCGTTACC CAACTTAATC
1201 GCCTTGCAGC ACATCCCCCT TTCGCCAGCT GGCGTAATAG CGAAGAGGCC CGCACCGATC
1261 GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGCG CTTTGCCTGG TTTCCGGCAC
1321 CAGAAGCGGT GCCGGAAAGC TGGCTGGAGT GCGATCTTCC TGAGGCCGAT ACTGTCGTCG
1381 TCCCCTCAAA CTGGCAGATG CACGGTTACG ATGCGCCCAT CTACACCAAC GTAACCTATC
1441 CCATTACGGT CAATCCGCCG TTTGTTCCCA CGGAGAATCC GACGGGTTGT TACTCGCTCA
1501 CATTTAATGT TGATGAAAGC TGGCTACAGG AAGGCCAGAC GCGAATTATT TTTGATGGCG
1561 TTAACTCGGC GTTTCATCTG TGGTGCAACG GGCGCTGGGT CGGTTACGGC CAGGACAGTC
1621 GTTTGCCGTC TGAATTTGAC CTGAGCGCAT TTTTACGCGC CGGAGAAAAC CGCCTCGCGG
1681 TGATGGTGCT GCGTTGGAGT GACGGCAGTT ATCTGGAAGA TCAGGATATG TGGCGGATGA
1741 GCGGCATTTT CCGTGACGTC TCGTTGCTGC ATAAACCGAC TACACAAATC AGCGATTTCC
1801 ATGTTGCCAC TCGCTTTAAT GATGATTTCA GCCGCGCTGT ACTGGAGGCT GAAGTTCAGA
1861 TGTGCGGCGA GTTGCGTGAC TACCTACGGG TAACAGTTTC TTTATGGCAG GGTGAAACGC
1921 AGGTCGCCAG CGGCACCGCG CCTTTCGGCG GTGAAATTAT CGATGAGCGT GGTGGTTATG
1981 CCGATCGCGT CACACTACGT CTGAACGTCG AAAACCCGAA ACTGTGGAGC GCCGAAATCC
2041 CGAATCTCTA TCGTGCGGTG GTTGAACTGC ACACCGCCGA CGGCACGCTG ATTGAAGCAG
2101 AAGCCTGCGA TGTCGGTTTC CGCGAGGTGC GGATTGAAAA TGGTCTGCTG CTGCTGAACG
2161 GCAAGCCGTT GCTGATTCGA GGCGTTAACC GTCACGAGCA TCATCCTCTG CATGGTCAGG
2221 TCATGGATGA GCAGACGATG GTGCAGGATA TCCTGCTGAT GAAGCAGAAC AACTTTAACG
2281 CCGTGCGCTG TTCGCATTAT CCGAACCATC CGCTGTGGTA CACGCTGTGC GACCGCTACG
2341 GCCTGTATGT GGTGGATGAA GCCAATATTG AAACCCACGG CATGGTGCCA ATGAATCGTC
2401 TGACCGATGA TCCGCGCTGG CTACCGGCGA TGAGCGAACG CGTAACGCGA ATGGTGCAGC
2461 GCGATCGTAA TCACCCGAGT GTGATCATCT GGTCGCTGGG GAATGAATCA GGCCACGGCG
2521 CTAATCACGA CGCGCTGTAT CGCTGGATCA AATCTGTCGA TCCTTCCCGC CCGGTGCAGT
2581 ATGAAGGCGG CGGAGCCGAC ACCACGGCCA CCGATATTAT TTGCCCGATG TACGCGCGCG
2641 TGGATGAAGA CCAGCCCTTC CCGGCTGTGC CGAAATGGTC CATCAAAAAA TGGCTTTCGC
2701 TACCTGGAGA GACGCGCCCG CTGATCCTTT GCGAATACGC CCACGCGATG GGTAACAGTC
2761 TTGGCGGTTT CGCTAAATAC TGGCAGGCGT TTCGTCAGTA TCCCCGTTTA CAGGGCGGCT
2821 TCGTCTGGGA CTGGGTGGAT CAGTCGCTGA TTAAATATGA TGAAAACGGC AACCCGTGGT
2881 CGGCTTACGG CGGTGATTTT GGCGATACGC CGAACGATCG CCAGTTCTGT ATGAACGGTC
2941 TGGTCTTTGC CGACCGCACG CCGCATCCAG CGCTGACGGA AGCAAAACAC CAGCAGCAGT
3001 TTTTCCAGTT CCGTTTATCC GGGCAAACCA TCGAAGTGAC CAGCGAATAC CTGTTCCGTC
3061 ATAGCGATAA CGAGCTCCTG CACTGGATGG TGGCGCTGGA TGGTAAGCCG CTGGCAAGCG
3121 GTGAAGTGCC TCTGGATGTC GCTCCACAAG GTAAACAGTT GATTGAACTG CCTGAACTAC
3181 CGCAGCCGGA GAGCGCCGGG CAACTCTGGC TCACAGTACG CGTAGTGCAA CCGAACGCGA
3241 CCGCATGGTC AGAAGCCGGG CACATCAGCG CCTGGCAGCA GTGGCGTCTG GCGGAAAACC
3301 TCAGTGTGAC GCTCCCCGCC GCGTCCCACG CCATCCCGCA TCTGACCACC AGCGAAATGG
3361 ATTTTTGCAT CGAGCTGGGT AATAAGCGTT GGCAATTTAA CCGCCAGTCA GGCTTTCTTT
3421 CACAGATGTG GATTGGCGAT AAAAAACAAC TGCTGACGCC GCTGCGCGAT CAGTTCACCC
3481 GTGCACCGCT GGATAACGAC ATTGGCGTAA GTGAAGCGAC CCGCATTGAC CCTAACGCCT
```

3541 GGGTCGAACG CTGGAAGGCG GCGGGCCATT ACCAGGCCGA AGCAGCGTTG TTGCAGTGCA
3601 CGGCAGATAC ACTTGCTGAT GCGGTGCTGA TTACGACCGC TCACGCGTGG CAGCATCAGG
3661 GGAAAACCTT ATTTATCAGC CGGAAAACCT ACCGGATTGA TGGTAGTGGT CAAATGGCGA
3721 TTACCGTTGA TGTTGAAGTG GCGAGCGATA CACCGCATCC GGCGCGGATT GGCCTGAACT
3781 GCCAGCTGGC GCAGGTAGCA GAGCGGGTAA ACTGGCTCGG ATTAGGGCCG CAAGAAAACT
3841 ATCCCGACCG CCTTACTGCC GCCTGTTTTG ACCGCTGGGA TCTGCCATTG TCAGACATGT
3901 ATACCCCGTA CGTCTTCCCG AGCGAAAACG GTCTGCGCTG CGGGACGCGC GAATTGAATT
3961 ATGGCCCACA CCAGTGGCGC GGCGACTTCC AGTTCAACAT CAGCCGCTAC AGTCAACAGC
4021 AACTGATGGA AACCAGCCAT CGCCATCTGC TGCACGCGGA AGAAGGCACA TGGCTGAATA
4081 TCGACGGTTT CCATATGGGG ATTGGTGGCG ACGACTCCTG GAGCCCGTCA GTATCGGCGG
4141 AATTCCAGCT GAGCGCCGGT CGCTACCATT ACCAGTTGGT CTGGTGTCAG GGGATCCCCC
4201 GGGCTGCAGC CAATATGGGA TCGGCCATTG AACAAGATGG ATTGCACGCA GGTTCTCCGG
4261 CCGCTTGGGT GGAGAGGCTA TTCGGCTATG ACTGGGCACA ACAGACAATC GGCTGCTCTG
4321 ATGCCGCCGT GTTCCGGCTG TCAGCGCAGG GGCGCCCGGT TCTTTTTGTC AAGACCGACC
4381 TGTCCGGTGC CCTGAATGAA CTGCAGGACG AGGCAGCGCG GCTATCGTGG CTGGCCACGA
4441 CGGGCGTTCC TTGCGCAGCT GTGCTCGACG TTGTCACTGA AGCGGGAAGG GACTGGCTGC
4501 TATTGGGCGA AGTGCCGGGG CAGGATCTCC TGTCATCTCA CCTTGCTCCT GCCGAGAAAG
4561 TATCCATCAT GGCTGATGCA ATGCGGCGGC TGCATACGCT TGATCCGGCT ACCTGCCCAT
4621 TCGACCACCA AGCGAAACAT CGCATCGAGC GAGCACGTAC TCGGATGGAA GCCGGTCTTG
4681 TCGATCAGGA TGATCTGGAC GAAGAGCATC AGGGGCTCGC GCCAGCCGAA CTGTTCGCCA
4741 GGCTCAAGGC GCGCATGCCC GACGGCGAGG ATCTCGTCGT GACCCATGGC GATGCCTGCT
4801 TGCCGAATAT CATGGTGGAA AATGGCCGCT TTTCTGGATT CATCGACTGT GGCCGGCTGG
4861 GTGTGGCGGA CCGCTATCAG GACATAGCGT TGGCTACCCG TGATATTGCT GAAGAGCTTG
4921 GCGGCGAATG GGCTGACCGC TTCCTCGTGC TTTACGGTAT CGCCGCTCCC GATTCGCAGC
4981 GCATCGCCTT CTATCGCCTT CTTGACGAGT TCTTCTGATC TAGAACTAGT GGATCCAGCT
5041 TATCGAGCTT GCCAACTACA GCAGGTAAGC ATATTTATGA AAAATGCTGA TTGTGTTAGC
5101 TACTTGTGTC AGTGTTGTGA TAGTCGACAT AACTTCGTAT AGCATACATT ATACGAAGTT
5161 ATACTAGATG TGTGGAATTG TGGCGGATAA CATTTCACAC ATCTAGTGAC GGCTACACCA
5221 ACAGTATCTA TACACTGTCC ATCAGCAGCG CCACACAGTT CGGCAATGTG CCCTGGTACA
5281 GTGAGGCCTG CTCCTCCACA CTGGCCACCA CCTACAGCAG CGGCAACCAG AATGAGAAGC
5341 AGATCGTGAG TCTTGCCTGG CTGGGGCTGG GAAGGAGGCC TCCTGCTCTC CTCAGGGCCA
5401 CATTAGGCTC TCTCTGACCT CATCCCTGTG ATAGTGGCTC AGGCATCATG AGATCCTGTG
5461 TGGAATTGTG AGCGGATAAC AATTTCACAC ATGATCTATA ACTTCGTATA GCATACATTA
5521 TACGAAGTTA TGATCTCGAG TTGTGATAGT CGAGGCGCCT AGAGAAGGAG TGAGGGCTGG
5581 ATAAAGGGAG GATCGAGGCG GGGTCGAACG AGGAGGTTCA AGGGGGAGAG ACGGGGCGGA
5641 TGGAGGAAGA GGAGGCGGAG GCTTAGGGTG TACAAAGGGC TTGACCCAGG GAGGGGGGTC
5701 AAAAGCCAAG GCTTCCCAGG TCACGATGTA GGGGACCTGG TCTGGGTGTC CATGCGGGCC
5761 AGGTGAAAAG ACCTTGATCT TAACCTGGGT GATGAGGTCT CGGTTAAAGG TGCCGTCTCG
5821 CGGCCATCCG ACGTTAAAGG TTGGCCATTC TGCAGAGCAG AAGGTAACCC AACGTCTCTT
5881 CTTGACATCT ACCGACTGGT TGTGAGCGAT CCGCTCGACA TCTTTCCAGT GACCTAAGGT
5941 CAAACTTAAG GGAGTGGTAA CAGTCTGGCC CATATTCTCA GACAAATACA GAAACACAGT
6001 CAGACAGAGA CAACACAGAA CGATGCTGCA GCAGACAAGA CGCGCGGCGC GGCTTCGGTC
6061 CCAAACCGAA AGCAAAAATT CAGACGGAGG CGGGAACTGT TTTAGGTTCT CGTCTCCTAC
6121 CAGAACCACA TATCCCTCCT CTAAGGGGGG TGCACCAAAG AGTCCAAAAC GATCGGGATT
6181 TTTGGACTCA GGTCGGGCCA CAAAAACGGC CCCCGAAGTC CCTGGGACGT CTCCCAGGGT
6241 TGCGGCCGGG TGTTCCGAAC TCGTCAGTTC CACCACGGGT CCGCCAGATA CAGAGCTAGT
6301 TAGCTAACTA GTACCGACGC AGGCGCATAA AATCAGTCAT AGACACTAGA CAATCGGACA
6361 GACACAGATA AGTTGCTGGC CAGCTTACCT CCCGGTGGTG GGTCGGTGGT CCCTGGGCAG
6421 GGGTCTCCCG ATCCCGGACG AGCCCCCAAA TGAAAGACCC CCGCTGACGG GTAGTCAATC
6481 ACTCAGAGGA GACCCTCCCA AGGAACAGCG AGACCACAAG TCGGATGCAA CTGCAAGAGG
6541 GTTTATTGGA TACACGGGTA CCCGGGCGAC TCAGTCAATC GGAGGACTGG CGCCCCGAGT
6601 GAGGGGTTGT GCTAGCTAGC TTGCCAAACC TACAGGTGGG GTCTTTCATT

proviral pCRR+SD126T7 I-Sce (X)

```
   1 AATGAAAGAC CCCCGCTGAC GGGTAGTCAA TCACTCAGAG GAGACCCTCC CAAGGAACAG
  61 CGAGACCACA AGTCGGATGC AACTGCAAGA GGGTTTATTG GATACACGGG TACCCGGGCG
 121 ACTCAGTCAA TCGGAGGACT GGCGCCCCGA GTGAGGGGTT GTGCTAGCTT GCCAAACCTA
 181 CAGGTGGGGT CTTTCATTCC CCCCTTTTTC TGGAGACTAA ATAAAATCTT TTATTTTATC
 241 TATGGCTCGT ACTCTATAGG CTTCAGCTGG TGATATTGTT GAGTCAAAAC TAGAGCCTGG
 301 ACCACTGATA TCCTGTCTTT AACAAATTGG ACTAATCGAT CTCGACCTCG AGACTAGTGG
 361 ATCCCCCGGG CTGCAGATCT GTAGGGCGCA GTAGTCCAGG GTTTTCCTTG ATGATGTCAT
 421 ACTTATCCCT GTCCCTTTTT TTTCCACAGC TCGCGGTTGA GGACAAACTC TTCGCGGTCT
 481 TTCCAGTGGG GATCGACGGT ATCGATAAGC TTGATGATCT GTGACATGGC GGATCCCGTC
 541 GTTTTACAAC GTCGTGACTG GGAAAACCCT GGCGTTACCC AACTTAATCG CCTTGCAGCA
 601 CATCCCCCTT TCGCCAGCTG GCGTAATAGC GAAGAGGCCC GCACCGATCG CCCTTCCCAA
 661 CAGTTGCGCA GCCTGAATGG CGAATGGCGC TTTGCCTGGT TTCCGGCACC AGAAGCGGTG
 721 CCGGAAAGCT GGCTGGAGTG CGATCTTCCT GAGGCCGATA CTGTCGTCGT CCCCTCAAAC
 781 TGGCAGATGC ACGGTTACGA TGCGCCCATC TACACCAACG TAACCTATCC CATTACGGTC
 841 AATCCGCCGT TTGTTCCCAC GGAGAATCCG ACGGGTTGTT ACTCGCTCAC ATTTAATGTT
 901 GATGAAAGCT GGCTACAGGA AGGCCAGACG CGAATTATTT TTGATGGCGT TAACTCGGCG
 961 TTTCATCTGT GGTGCAACGG GCGCTGGGTC GGTTACGGCC AGGACAGTCG TTTGCCGTCT
1021 GAATTTGACC TGAGCGCATT TTTACGCGCC GGAGAAAACC GCCTCGCGGT GATGGTGCTG
1081 CGTTGGAGTG ACGGCAGTTA TCTGGAAGAT CAGGATATGT GGCGGATGAG CGGCATTTTC
1141 CGTGACGTCT CGTTGCTGCA TAAACCGACT ACACAAATCA GCGATTTCCA TGTTGCCACT
1201 CGCTTTAATG ATGATTTCAG CCGCGCTGTA CTGGAGGCTG AAGTTCAGAT GTGCGGCGAG
1261 TTGCGTGACT ACCTACGGGT AACAGTTTCT TTATGGCAGG GTGAAACGCA GGTCGCCAGC
1321 GGCACCGCGC CTTTCGGCGG TGAAATTATC GATGAGCGTG GTGGTTATGC CGATCGCGTC
1381 ACACTACGTC TGAACGTCGA AAACCCGAAA CTGTGGAGCG CCGAAATCCC GAATCTCTAT
1441 CGTGCGGTGG TTGAACTGCA CACCGCCGAC GGCACGCTGA TTGAAGCAGA AGCCTGCGAT
1501 GTCGGTTTCC GCGAGGTGCG GATTGAAAAT GGTCTGCTGC TGCTGAACGG CAAGCCGTTG
1561 CTGATTCGAG GCGTTAACCG TCACGAGCAT CATCCTCTGC ATGGTCAGGT CATGGATGAG
1621 CAGACGATGG TGCAGGATAT CCTGCTGATG AAGCAGAACA ACTTTAACGC CGTGCGCTGT
1681 TCGCATTATC CGAACCATCC GCTGTGGTAC ACGCTGTGCG ACCGCTACGG CCTGTATGTG
1741 GTGGATGAAG CCAATATTGA AACCCACGGC ATGGTGCCAA TGAATCGTCT GACCGATGAT
1801 CCGCGCTGGC TACCGGCGAT GAGCGAACGC GTAACGCGAA TGGTGCAGCG CGATCGTAAT
1861 CACCCGAGTG TGATCATCTG GTCGCTGGGG AATGAATCAG GCCACGGCGC TAATCACGAC
1921 GCGCTGTATC GCTGGATCAA ATCTGTCGAT CCTTCCCGCC CGGTGCAGTA TGAAGGCGGC
1981 GGAGCCGACA CCACGGCCAC CGATATTATT TGCCCGATGT ACGCGCGCGT GGATGAAGAC
2041 CAGCCCTTCC CGGCTGTGCC GAAATGGTCC ATCAAAAAAT GGCTTTCGCT ACCTGGAGAG
2101 ACGCGCCCGC TGATCTTTGC GAATACGCCC ACGCGATGGG TAACAGTCTT GGCGGTTTCG
2161 CTAAATACTG GCAGGCGTTT CGTCAGTATC CCCGTTTACA GGGCGGCTTC GTCTGGGACT
2221 GGGTGGATCA GTCGCTGATT AAATATGATG AAAACGGCAA CCCGTGGTCG GCTTACGGCG
2281 GTGATTTTGG CGATACGCCG AACGATCGCC AGTTCTGTAT GAACGGTCTG GTCTTTGCCG
2341 ACCGCACGCC GCATCCAGCG CTGACGGAAG CAAAACACCA GCAGCAGTTT TTCCAGTTCC
2401 GTTTATCCGG GCAAACCATC GAAGTGACCA GCGAATACCT GTTCCGTCAT AGCGATAACG
2461 AGCTCCTGCA CTGGATGGTG GCGCTGGATG GTAAGCCGCT GGCAAGCGGT GAAGTGCCTC
2521 TGGATGTCGC TCCACAAGGT AAACAGTTGA TTGAACTGCC TGAACTACCG CAGCCGGAGA
2581 GCGCCGGGCA ACTCTGGCTC ACAGTACGCG TAGTGCAACC GAACGCGACC GCATGGTCAG
2641 AAGCCGGGCA CATCAGCGCC TGGCAGCAGT GGCGTCTGGC GGAAAACCTC AGTGTGACGC
2701 TCCCCGCCGC GTCCCACGCC ATCCCGCATC TGACCACCAG CGAAATGGAT TTTTGCATCG
2761 AGCTGGGTAA TAAGCGTTGG CAATTTAACC GCCAGTCAGG CTTTCTTTCA CAGATGTGGA
2821 TTGGCGATAA AAAACAACTG CTGACGCCGC TGCGCGATCA GTTCACCCGT GCACCGCTGG
2881 ATAACGACAT TGGCGTAAGT GAAGCGACCC GCATTGACCC TAACGCCTGG GTCGAACGCT
2941 GGAAGGCGGC GGGCCATTAC CAGGCCGAAG CAGCGTTGTT GCAGTGCACG GCAGATACAC
3001 TTGCTGATGC GGTGCTGATT ACGACCGCTC ACGCGTGGCA GCATCAGGGG AAAACCTTAT
3061 TTATCAGCCG GAAAACCTAC CGGATTGATG GTAGTGGTCA AATGGCGATT ACCGTTGATG
3121 TTGAAGTGGC GAGCGATACA CCGCATCCGG CGCGGATTGG CCTGAACTGC CAGCTGGCGC
3181 AGGTAGCAGA GCGGGTAAAC TGGCTCGGAT TAGGGCCGCA AGAAAACTAT CCCGACCGCC
3241 TTACTGCCGC CTGTTTTGAC CGCTGGGATC TGCCATTGTC AGACATGTAT ACCCCGTACG
3301 TCTTCCCGAG CGAAAACGGT CTGCGCTGCG GGACGCGCGA ATTGAATTAT GGCCCACACC
3361 AGTGGCGCGG CGACTTCCAG TTCAACATCA GCCGCTACAG TCAACAGCAA CTGATGGAAA
3421 CCAGCCATCG CCATCTGCTG CACGCGGAAG AAGGCACATG GCTGAATATC GACGGTTTCC
3481 ATATGGGGAT TGGTGGCGAC GACTCCTGGA GCCCGTCAGT ATCGGCGGAA TTCCAGCTGA
```

3541 GCGCCGGTCG CTACCATTAC CAGTTGGTCT GGTGTCAGGG GATCCCCCGG GCTGCAGCCA
3601 ATATGGGATC GGCCATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC GCTTGGGTGG
3661 AGAGGCTATT CGGCTATGAC TGGGCACAAC AGACAATCGG CTGCTCTGAT GCCGCCGTGT
3721 TCCGGCTGTC AGCGCAGGGG CGCCCGGTTC TTTTTGTCAA GACCGACCTG TCCGGTGCCC
3781 TGAATGAACT GCAGGACGAG GCAGCGCGGC TATCGTGGCT GGCCACGACG GGCGTTCCTT
3841 GCGCAGCTGT GCTCGACGTT GTCACTGAAG CGGGAAGGGA CTGGCTGCTA TTGGGCGAAG
3901 TGCCGGGGCA GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAGTA TCCATCATGG
3961 CTGATGCAAT GCGGCGGCTG CATACGCTTG ATCCGGCTAC CTGCCCATTC GACCACCAAG
4021 CGAAACATCG CATCGAGCGA GCACGTACTC GGATGGAAGC CGGTCTTGTC GATCAGGATG
4081 ATCTGGACGA AGAGCATCAG GGGCTCGCGC CAGCCGAACT GTTCGCCAGG CTCAAGGCGC
4141 GCATGCCCGA CGGCGAGGAT CTCGTCGTGA CCCATGGCGA TGCCTGCTTG CCGAATATCA
4201 TGGTGGAAAA TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC
4261 GCTATCAGGA CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC GGCGAATGGG
4321 CTGACCGCTT CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC ATCGCCTTCT
4381 ATCGCCTTCT TGACGAGTTC TTCTGATCTA GTAATACGAC TCACTATAGG GCGATCGAGT
4441 TACGCTAGGG ATAACAGGGT AATATAGTCT AGAACTAGTG GATCCAGCTT ATCGAGCTTG
4501 CCAACTACAG CAGGTAAGCA TATTTATGAA AAATGCTGAT TGTGTTAGCT ACTTGTGTCA
4561 GTGTTGTGAT AGTCGACGGC GCCTAGAGAA GGAGTGAGGG CTGGATAAAG GGAGGATCGA
4621 GGCGGGGTCG AACGAGGAGG TTCAAGGGGG AGAGACGGGG CGGATGGAGG AAGAGGAGGC
4681 GGAGGCTTAG GGTGTACAAA GGGCTTGACC CAGGGAGGGG GGTCAAAAGC CAAGGCTTCC
4741 CAGGTCACGA TGTAGGGGAC CTGGTCTGGG TGTCCATGCG GGCCAGGTGA AAAGACCTTG
4801 ATCTTAACCT GGGTGATGAG GTCTCGGTTA AAGGTGCCGT CTCGCGGCCA TCCGACGTTA
4861 AAGGTTGGCC ATTCTGCAGA GCAGAAGGTA ACCCAACGTC TCTTCTTGAC ATCTACCGAC
4921 TGGTTGTGAG CGATCCGCTC GACATCTTTC CAGTGACCTA AGGTCAAACT TAAGGGAGTG
4981 GTAACAGTCT GGCCCATATT CTCAGACAAA TACAGAAACA CAGTCAGACA GAGACAACAC
5041 AGAACGATGC TGCAGCAGAC AAGACGCGCG GCGCGGCTTC GGTCCCAAAC CGAAAGCAAA
5101 AATTCAGACG GAGGCGGGAA CTGTTTTAGG TTCTCGTCTC CTACCAGAAC CACATATCCC
5161 TCCTCTAAGG GGGGTGCACC AAAGAGTCCA AAACGATCGG GATTTTTGGA CTCAGGTCGG
5221 GCCACAAAAA CGGCCCCCGA AGTCCCTGGG ACGTCTCCCA GGGTTGCGGC CGGGTGTTCC
5281 GAACTCGTCA GTTCCACCAC GGGTCCGCCA GATACAGAGC TAGTTAGCTA ACTAGTACCG
5341 ACGCAGGCGC ATAAAATCAG TCATAGACAC TAGACAATCG GACAGACACA GATAAGTTGC
5401 TGGCCAGCTT ACCTCCCGGT GGTGGGTCGG TGGTCCCTGG GCAGGGGTCT CCCGATCCCG
5461 GACGAGCCCC CAAATGAAAG ACCCCCGCTG ACGGGTAGTC AATCACTCAG AGGAGACCCT
5521 CCCAAGGAAC AGCGAGACCA CAAGTCGGAT GCAACTGCAA GAGGGTTTAT TGGATACACG
5581 GGTACCCGGG CGACTCAGTC AATCGGAGGA CTGGCGCCCC GAGTGAGGGG TTGTGCTAGC
5641 TAGCTTGCCA AACCTACAGG TGGGGTCTTT CATT

APPENDIX III

Sequence data obtained from 3′ RACE products:

$CaPO_4$ 8-11 - E7 – Mus musculus RIKEN cDNA 1500001M02 gene, mRNA (cDNA clone IMAGE: 4218748)

NNNACGNNNNTNTNCNCCCCCGNNCNNNNNNAAACCCCCCNCCCCTNTTTTTTTTTTTTGNGG
CCCNAAAAAAAGGNNTTNNNNTNNCNNCCCNGGTTNTAAANNTTGGNNTTNCNTAAAGGNGG
CCCCNTANAATNTNNGGGCCCCNNCCCAAAAGGNCNTTTTTTTTTTTTCCCAGGGGATCATTTT
TTATTTGCATCTCGGTTTCAGTACAATTTACAACATTGTCCCCCGTTCCAGTCTGATTATGCTC
ATGCTAAGGGACAGAAAGGGCCTGGAATAAATACATTAAAAAAAAAAGCTCTCAACTAAGGGG
CACACAGTAGGCCCTATGGGGGAGGGGAAGGGACTGAGAAACTAAGTGGCACCCAGAAGGT
CCTATGGGGTGTGGGTGTCTGAGACAGGAAGTGAGCAAGCAAACGCAAGAGGCGGGTCAGGC
CGCAAAGCT

$CaPO_4$ 9 - $1A6_{72}$ - Mus musculus vaccinia related kinase 3 (AI428238)

TTTTTTTTTNNTTTTGATTAAAAACAGTTTATTATGCCATCACAGAAATGCCATAGGGAGTGGA
GTGGCTGCTGCCGCC

$CaPO_4$ 9 - $1F8_{72}$ - Similar to phosphoglycerate kinase 1 (XM283048)

TTTTNCNTTTTTTTTTAGGGGACATCTTTATTGTTTAATACATCATCAATTTCTGCAGACTTACA
GC

$CaPO_4$ 8-11 - H1 – Mus musculus Idb1 Inhibitor of DNA binding 1 (NM010495)

TTTTTTTTTTTTTTCTTTTTATCTAAAGGGTTNGTTTAATAACAACAAAAAACTCACCCCC
ATAGAACTATTGTAAAACAATATTTTCAGCCAGGGATCATTGTAATATACAATACAAAAT
ATTTCCTCAGAAATCCGAGAAGCACGAAATGTGACCATTTAAAAAAAAAATGTCTGCTTT
TTCAATAAAACAGAAACACGCGGGGTTGATTAACCCCCTCCCCAAAGTCTCTGGAGGCTG
AAAGGTGGAGAGGGTGAGGCTCTGTTGATCAAACCCTTTACCCACTGGACCGATCCGCCA
AGAGTCCGGGGGCTTGCGGTAGTGTCTTTCCCAGAGATCCCCTGGGGAACCGAGAGCACT
TTTTTCCTCTTGCCTCCTGAAGGGCTGGAGGCCATCTGGGCCCTCAGTGCGCCGCCTAAG
CGACACAAGATGCGATCGTCGGCTGGAACAAATGCCGCCTCGG

$CaPO_4$ 8-11 - B6 – Mus musculus ribosomal protein L12 (Rpl12) (XM_192893.1)

GACATCATCATCATATATTGCTGAAGAGCTTGGCGGCCCGTGCGCCTTAGGTGTACCTGAGGT
GCACCGGAGGCGAGGTCGGCGCCACATCCGCCTTGGCCCCGAAGATCGGTCCTCTGGGTCTGT
CTCCGAAGAAAGTTGGCGATGACATTGCCAAGGCTACCGGTGACTGGAAGGGTCTCAGAATT
ACAGTGAAACTGACCATCCAGAACAGACAGGCCCAGATCGAGGTGGTGCCTTCTGCCTCAGC
CCTGATCATCAAAGCCCTCAAGGAGCCACCGAGAGACAGGAGGAAGCAGAAGAACATTAAA
CACAGTGGNAACATTACTTTTGATGAGATTGTCAACATTGCCCGGCAGATGAGACACCGGTCT
TTGGCTAGAGAACTTTCTGGAACTATCAAGGAGATCCTGGGTACTGCACAGTCTGTGGGCTGC
AATGTGGATGGCCGCCACCCTCATGACATCATAGATGACATCAACAGTGGTGCAGTGGAGTG
CCCANCTNGTTAAGAAGCAACAAGAAAATNTTCCATTAAAGACTATCTGATAACCAGTAAAA
AAAAAAAAAAAAAAAAAAAA

$CaPO_4$ 11 - $F6_{48}$ – Mus musculus ribosomal protein L30 (Rpl30)

TGGTGTTGGTCTGGGGGTTGTCCCGCACCTAAGGCAGGAAGATGGTGGCCGCAAAGAAGACG
AAAAAGTCTCTGGAGTCGATCAACTCTAGGCTCCAACTTGTTATGAAAAGTGGGAAGTACGTG
CTGGGCTACAAGCAGACTCTGAAGATGATCAGACAAGGCAAAGCGAAGTTGGTTATCCTTGC
CAACAACTGTCCAGCTTTGAGGAAATCTGAAATAGAGTACTATGCCATGTTGGCTAAAACTGG
GGTCCATCACTACAGTGGCAATAATATTGAACTGGGCACAGCGTGTGGAAAATACTACAGAG
TATGCACACTGGCTATCATTGACCCAGGTGATTCTGATATTATTAGAAGCATGCCAGAACAGA
CTGGTGAGAAGTAAACAAGAAAGTTTTCCTTTAATAAAACTTTGCCAGAGCTCCTTTTGAAAA
AAAAAAAAAAAAAAAAAAAGTNCTNTTCAACCCGGGGCCTANTANTATTAGGCGGCCGCTTTA
AAGGATCCAAGNTTACNTACGCGNGCNTGCNANNNNATANNTTTTTTATNGGGNCNCCTAAN
TNAATTCACTGGCCGNGGTTTNNAACGNNNTGGGGNNAAAAAAANNNNNNNNNNN

$CaPO_4$ 11 - $E11_{48}$ – Mus musculus ribosomal protein L28

ACATATATCTGCTGTAAGTAGCTTGGCGGCCATCCGCAGGGTCCAGTGCCATCCTTCGAAGCC
AGNAAGCCTGTGGCGGTGAAGAGGAAACGGACCCGCCCCACCAAGNAGCTCCTGAGCCCCAC
ACGCCCGAAGCAATAAAGAGTCCACTGACTTCCAAAAAAAAAAAAAAAANAANTAGTCGAAN
CNNNNANNAAANNTTTTTGGCGGCGNGGNNTAAAGGATCCAAGNTNNNGCCCTTTTAAANGA
AANANNTTAACTTTNCTATNGNGGCANCTAAATTCTTTTTTTTGGNNGNCCNTTAANAANATT
CNGNNGGGGAAANTTTNANANNNNGGGNANGGGAAAAAANNNNNNN

$CaPO_4$ 11 - $F11_{48}$ – Arabidopsis thaliana Wee 1

GAATTCACTAGTGATTNAGAGAAGATATCNNCTTTGTCAAGATGATCATAGTTCTCATTAAGT
ATTTCTTGAGGCATATAACGTGCATCCCCCTCTTCGATCGGCTGGCTCTTATCAAGAAGAGTTG
CACATCCAAAATCACCAAGCTTATACACACCACTTTTCACATAAATATTATCTGGCTTTACATC
TAAATGAGCTACCCCTCTCTGATGGATATATTGCAATGCCTTGGCTACCTGATACATTGCTTCC
AAAACTGCTACCTCCGAAGATAGTTTAGAATATTTTTTGTTGGATAAGCTGTGGTCACAGAGC
TCCATTTGGATGTAAAGATGCTCATTCTCAAACCACGAAATCGAATTCCCGCGGCCGGCAGTT
TTGGAAGCAATGTATCAGGTAGCCAAGGCATTGCAATATATCCATCAGAGAGGGGTAGCTCA
TTTAGATGTAAAGCCAGATAATATTTATGTGAAAAGTGGTGTGTATAAGCTTGGTGATTTTGG
ATGTGCAACTCTTCTTGATAAGAGCCAGCCGATCGAAGAGGGGGATGCACGTTATATGCCTCA
AGAAATACTTAATGAGAACTATGATCATCTTGACAAAGTGGATATCTTCTCTAATCACTAGTG
AATTCG

$CaPO_4$ 11 - $B2_{48}$ – Mus musculus cytochrome oxidase subunit Vb

ATATTGCTGAAGAGCTTGGCGGCGCAGGCCCTGAGGGCCCACGGCCCCCGTGGCGCGGCCGT
GACCCGCTCCATGGCTTCTGGAGGTGGTGTCCCCACTGATGAGGAGCAGGCTACTGGGCTGGA
GAGGGAGATCATGATAGCAGCACAGAAGGGACTGGACCCATACAATATGCTACCTCCAAAGG
CAGCTTCAGGCACCAAGGAAGACCCTAATCTAGTCCCGTCCATCAGCAACAAGAGAATAGTG
GGCTGCATCTGTGAAGAGGACAACTGTACTGTCATCTGGTTTTGGCTGCACAAAGGCGAGAGT
CAGCGATGCCCCAACTGTGGAACCCATTACAAGCTGGTGCCCCACCAAATGGCCCACTGAGCC
CCTGTGTTATCTTTTCAGAATGTAAAGAAAAACTTCTCTCTAATAAAGATTAGCCATTGCACTG
GCTCCTCCCATAAAAAAAAAAAAAAAAGTACTANTCGACTNCTAGGCCTACGCGTGGCCTAGC
TNGTNGCNCGCGGGCGCTNTACANGATCCNAGCTTACGTACGACGTGCATGCNTACTTCATAG
CNTTTCNATAGTGNCACCTANNGTCAATTCACTGNACNGCNGTTTNACNNGGNANNGACTGG
GAAAACCCTGNCGTTACCCATNTTAATCGACTTGCATTNATCCCCCTNTNGNCAATTGAAGNA
GTACCGAAAAAGNCCGCTNCGATNGACNTTNCCAAANTGATGNANNNTGAATGGCNACNGGN
ACCGCCCTNTAACNNCCNTNAACNCGGGGGNNNNGGNGGNTNNCCCANNNNTNCCN

$CaPO_4$ 11 - $D1_{48}$ – Mus musculus Histone H2A.1 (AY158913)

GTATNGCCNTGATTNCNCCNAGCTTTAATNCGNCTCCCTATAGGGAAANCTGGTACGCCTGCA
GGTACCGGTCCGGAATTCCCGGGTCGACATCATCATCATATATTGCTGAAGAGCTTGGCGGCC
GTGCTGGAGTACCTGACGGCCGAGATCCTGGAGCTGGCGGGCAACGCGGCCCGCGACAACAA
GAAAACGCGCATCATCCCGCGCCACCTGCAGCTGGCCATCCGCAACGACGAGGAGCTCAACA
AGCTGCTGGGCCGCGTGACCATCGCGCAGGGCGGCGTCCTGCCCAACATCCAGGCCGTGCTGC
TGCCCAAGAAGACCGAGAGCCACCACAAGGCCAAGGGAAAATAAGACCAGCCGTTCACCCAC
CCGAAAAAAAAAAAAAAAAA

$CaPO_4$ 11 - $E5_{72}$ – Mus musculus ribosomal protein L30

CGCTTTTNGCCANGCNTCCAATACNACTNNCTNTANGGCAAAGANTCGGGACACCNGCAG
GTACCGGTCCGGAATTCCCGGGTCGACATCATCATCATATATTGCTGTAAGTAGCTTGGC
GGCCTGGTGTTGGTCTGGGGGTTGTCCCGCACCTAAGGCAGGAAGATGGTGGCCGCAAAG
AAGACGAAAAAGTCTCTGGAGTCGATCAACTCTAGGCTCCAACTTGTTATGAAAAGTGGG
AAGTACGTGCTGGGCTACAAGCAGACTCTGAAGATGATCAGACAAGGCAAAGCGAAGTTGGT
TATCCTTGCCAACAACTGTCCAGCTTTGAGGAAATCTGAAATAGAGTACTATGCCATG
TTGGCTAAAACTGGGGTCCATCACTACAGTGGCAATAATATTGAACTGGGCACAGCGTGT
GGAAAATACTACAGAGTATGCACACTGGCTATCATTGACCCAGGTGATTCTGATATTATT
AGAAGCATGCCAGAACAGACTGGTGAGAAGTAAACAAGAAAGTTTTCCTTTAATAAAACT
TTGCCAGAGCTCCTTTTGAAAAAAAAAAAAAAAAATTAANACCTTAAAAAAAAAAAAAAAAAA
GTNCTATTCNACNCGNGGCCTANTANTATTAGGCGGCCNCTCTANAGGATCCAAGCTNACGTA
CCCNNGCGTGCGAGGTNTTAGCTNTTTTATNGNGNCCCCTAANTNAATTCACTG

$CaPO_4$ 11 - $D4_{72}$ – Mus musculus HRAS like suppressor 3 (NM139269)

TTTTTTTTTTTTTTTTTTCCTTTCTTGTGGTTTATTGTATGTTGTGAAAATAAACCTAATTAAAAG
CAATAGAATCTATAAATCAAACTCCGAACCCTCTAGCAAAAGAAGAGCCCAAAAACTGGGCA
GTCATTCAGCTCATTGCTTCTGTTTCTTGTTTCTGGAGAGCATGACTCCAACGAGGCCCAGGG

$CaPO_4$ 11 - $G6_{72}$ – Mus musculus Peptidylprolyl isomerase A (NM008907)

TTTTTTTTTTTTTTCTTATTTAGTTTTTATTCATAATCATAAACTTAACTTTGCAATCCAGCTAGA
CTTGAAGGGGAATGAGGAAAATATGGAACCCAAAGAACTTCAGTGAGAGCAGAGATTACAG
GACATTGCGAGCAGATGGGGTAGGGACGCTCTCCTGAGCTACAGAAGGAATGGTTTGATGGG
TAAAATGCCCGCAAGTCAAAAGAAATTAGAGCTGTCCACAGTCGGAAATGGTGATCTTCTTGC
TGGTCTTGCCATTCCTGGACCCAAAACGCTCCATGGCTTCCACAATGTTCATGCCTTCTTTCAC
CTTCCCAAAGACCACATGCCTGCCATCCAGCCATTCAGTCTTGGCAGTGCAGATAAAAAACTG
GGAACCGTTTGTGTTTGGTCCAGCATTTGCCATGGACAAGATGCCAGGACCTGTATG
CTTTAGGATGAAGTTCTCATCCTCAAACTTCTCTCCGTAGATGGACCTG

$CaPO_4$ 11 - $B7_{72}$ – Mus musculus clone IMAGE:3968454 mRNA (BC005576.1)

TTTTTTTTTTTTTTTTGCTGCCTAGAACTAGAAGACTCTCGTGCCTTACTACCAGGGAGCATTGT
GTGTGCTGACAGCCTCAGTTTCTCTCCAGCCCCTATTNNAACACTCACGTCCTCCTCGTCGGAT
GTGTCTGACAGGATGCGATGGG

$CaPO_4$ 11 - $H8_{72}$ – Mus musculus proteasome (prosome, macropain) subunit, beta type 6 (NM008946.2)

GACCTTTTTTTTTTTTTTTTTTTTTTTTTTTCTCTTTAACAAACTGTTTATTAGCTTCTGCGTCGGT
AAGAACCATCCTTTACTACTCCCTAGAACACCAGGCCTCAGGGAGGTGGCAACGTGGCAATG
GTGAACTTGGGGATTTGGTCTCCCAAAAGCACCTGCCGCTCTACCCCTGACACCTGAATGG

$CaPO_4$ 11 - $D4_{48}$ – Mus musculus zinc finger protein 369, mRNA (cDNA clone MGC:41715 IMAGE:1365300)

CATCATATAGTATGCATGAAGAGCTTGGCGNCCTCCGACTGGTNCAGNNCNATGANCAGANN
AGNTCATGATTGCTTCATTTATCCTCAAAAGGAAATAAAAGATCTGAAAATGTGTCTAATGGA
TAACAAGCTCAAGTTTCAAAAATTGCACTGATATTTATATACAGGGTACTGTGTGTAAGTGAA
CAAAACTAAAAGTATATAGATGAATGAATAAACTATTGTGTGGAACTTTTTTAAAAAAAAAA
ANANNANAAAAAAAANANANNNNNNNNNNNNN