TARGETED TRANSGENESIS AND THE 186 SITE-SPECIFIC RECOMBINATION SYSTEM

A thesis submitted for the degree of

Doctor of Philosophy

at the

UNIVERSITY OF ADELAIDE

Department of Biochemistry

South Australia

by

SHARON JANE HARRISON
B.Sc. (Hons.)

April, 1999

Statement

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

To the best of my knowledge and belief, this thesis contains no material previously published or written by any other person except where due reference is made in the text.

I consent to this thesis being made available for photocopying and loan.

Sharon J. Harrison

Summary

Efforts are being made to find alternative and more efficient means of generating transgenic livestock, to perform precise genetic modifications in mammals, other than the mouse, such as a gene "knock-out", and to insert a new gene or replace a defective gene, as required for successful gene therapy in humans.

Current methods for producing transgenic farm animals rely on the use of microinjection into fertilised eggs which has several inherent disadvantages as the site of insertion is random, the copy number uncontrollable and gene expression is subject to position effects from the surrounding DNA context. In contrast, homologous recombination to generate transgenic mice has been very successful in contributing to our understanding of basic biological phenomena such as mammalian physiology and development, and for producing animal models of human disease. Currently this technology is dependent on the availability of embryonic stem cells which do not yet exist for any livestock species. Although homologous recombination has been tried in fertilised mouse eggs the success rate is too low for consideration as a routine method. The Cre/*lox* and FLP/FRT recombinase systems have been shown to operate in eukaryotic systems via microinjection into fertilised mouse eggs (excision), but the drawback with these systems is their non-directionality, one protein directs both excision and integration.

Nuclear transfer has been used to produce cloned sheep (Dolly), cows and mice from adult cells. Combining the techniques of homologous recombination in a foetal primary cell line, and nuclear transfer, Schnieke's team (1997) successfully produced transgenic sheep. At present the technology is inefficient. Furthermore, successful homologous recombination demands construction of a targeting vector containing vast regions of isogenic homology (species specific), a time consuming and onerous task.

The temperate coliphages like λ, P2 and 186 can enter either a lytic or lysogenic lifecycle upon infection of their *Escherichia coli* host. As different combinations of proteins mediate the excisive and integrative pathways these systems have the directionality lacking in the one protein systems like Cre and FLP. However, none have been shown to function in a eukaryotic environment. The recombinase systems may offer an alternative route to homologous recombination for the initial introduction of the target site, and that is by using closely matching, naturally occurring target sites in combination with an altered integrase, one could bypass the homologous recombination step altogether

The aim of the work described in this thesis was to characterise the 186 integration reaction and, to conduct preliminary investigations into the possible use of such a system for targeted transgenesis in livestock species. In Chapter 2 the *in vitro* requirements for 186 integrative site-specific recombination were investigated. Chapter 3 details characterisation of the 186-*att*B site in which I located its precise position, sites of strand cross-over, order of strand exchange and conducted preliminary mutational analysis to identify bases important for recombination. Chapter 4 describes *in vitro* mouse genomic DNA recombination experiments using naked DNA and chromatin containing *att*B sites. Lastly, *in vivo* investigations were conducted whereby active 186-intasomes were microinjected into fertilised mouse eggs containing genomic copies of 186-*att*B. The resultant progeny were then analysed for 186-mediated integration events.

Acknowledgments

I would like to take this opportunity to wholeheartedly thank the many who have given their support, be it intellectual, financial or emotional, and allowed me to fulfil a personal goal.

To the University of Adelaide for the support of a postgraduate scholarship (UAS), and to the Department of Biochemistry, in particular Dr. B. Egan and Prof. P. Rathjen for permission to undertake this study. To BresaGen, Dr. J. Smeaton, Dr. A. Robins and Dr. P. Tolstoshev, for providing the opportunity of further study in the first instance, for allowing me to combine work/Ph.D and family commitments, and also for their generous financial support.

A special thankyou goes to my two supervisors, Dr. Barry Egan (Dept. Biochem.) and Dr. Peter Wigley (BresaGen), for helpful and critical advice and for taking my education as a future scientist seriously.

To my friends and colleagues Lisa, Rick, Allan H., Blair, Angelo, Katina, Meera, George, Roger, Ela and the Mol. Biol. Club (G08), your company and good humour (at times warped!) was immensely appreciated. Also, thankyou to Steve MacIlfatrick for performing the fertilised egg microinjections and, together with Dr. Zhong Tao Du, for coordinating the mouse breeding.

To Alan, my husband, and Timothy, our son, a very special thankyou for your immeasurable support, encouragement and love from beginning to end, and all the bits in between.

Contents

CHAPTER 1

Introduction

Chapter 1

Introduction

Genetics is a powerful tool for studying the function of different gene products as well as the developmental and physiological consequences of a normal or aberrant protein. Until recently, the most potent genetic applications have been confined to relatively simple organisms whose genomes are more easily manipulated. Currently there are only a few tools available to precisely modify the genetic make-up of animals, and then predominantly only in mice. Consequently the widespread use of such technology for the efficient and precise modification of a livestock animal's genome awaits further developments and refinements.

1.1 Transgenesis

Technological advances in biology have enabled the development of methods that allow foreign genes to be introduced into the germline of animals. Animals containing foreign DNA inserted into their genomes are said to be transgenic (Gordon *et al.,* 1980) and the foreign DNA is referred to as the transgene (for reviews see Palmiter and Brinster, 1986; Jaenisch *et al.,* 1988). Transgenic animals provide new insights into the mechanisms of development and developmental gene regulation that are not evident from studies in cultured cells (Brinster *et al.,* 1988).

1.1.1 Biotechnological applications

Transgenesis also offers a number of biotechnological applications such as genetic engineering of livestock. In addition to improving existing traits (Vize *et al.,* 1988; Palmiter *et al.,* 1982), animals with entirely new properties can be produced. Presently there is a considerable amount of interest in transgenic animals for the production of recombinant biomedical proteins and for the generation of transgenic pigs as organ donors (xenotransplantation). One approach to the production of recombinant proteins involves targeting transgene expression to the mammary epithelial cells, permitting harvesting of the protein product from milk (Whitelaw *et*

al., 1991; Wilmut *et al.*, 1994). Such a powerful technology will enable transgenic animals to produce high-value human biomedical proteins that contain the correct post-translational modifications that are essential for their function. For example, Paleyander and coworkers (1997) recently reported the successful production of functional human factor VIII, vital for prophylactic treatment of haemophiliacs, in the milk of transgenic pigs.

There exists a world-wide organ shortage for lifesaving transplantation operations, and one approach to solving this problem could be xenotransplantation - the transplantation into humans of organs from other species. Pigs are currently considered the most likely source of organs for human xenotransplantation due to their anatomical and physiological similarity to humans, and the relative ease with which they can be bred in large numbers. Many research groups around the world, including BresaGen (Australia), are investigating the feasibility of pig to human organ donation. A severe form of rejection known as hyperacute rejection has been a major barrier to the use of xenografts (Platt *et al.*, 1991). All mammals express on the endothelium of their organs galactose α1-3 galactose moeities, known as GAL epitopes, which are generated by α1-3 galactosyltransferase. Humans (and Old World monkeys) have lost the expression of this enzyme (Galili *et al.*, 1991) and by exposure to the GAL epitopes, via food/bacteria in the gut, have naturally circulating anti-GAL antibody, demonstrated to be central to the hyperacute rejection (Tearle *et al.*, 1996). Also implicated in the hyperacute rejection response is the incompatibility between human complement and porcine complement regulatory factors (CRF). Recently, van Denderen (1997) showed that mice transgenic for the human complement regulator decay-accelerating factor (DAF), which also had the α1-3 galactosyltransferase gene knocked out, when challenged with human blood or plasma, had higher levels of protection from complement-mediated injury. Generating transgenic pigs for organ transplantation will involve precise genetic manipulation to ablate the α1-3 galactosyltransferase gene and to introduce human DAF and membrane-bound human CRF genes. At the present time there exists no technology to achieve a

precise modification of an animal's genome, apart from mice, for which embryonic stem cells are available.

Genetic engineering of livestock is expected to have a profound effect on the future of the agricultural industry. Exploitation of the technology has been used to create transgenic animals which express either extra growth hormone (GH) or the insulin-like growth factors (IGFs), both of which regulate biological processes that influence growth rate and feed efficiency. In particular, pigs transgenic for the GH gene have been shown to have alterations in carcass composition that results in a marked reduction in subcutaneous fat (Vize *et al.,* 1988; Pursel *et al.,* 1989).

Despite the encouraging success there has been in creating transgenic animals of agricultural and biomedical importance there remains to be addressed the need to improve the efficiency and ease of introducing genes into the germline of animals and to develop techniques to introduce single copies of genes at precise locations.

1.2 Microinjection

Microinjection of exogenous DNA into a pronucleus of fertilised eggs is the current method employed to create transgenic farm animals. Fertilised oocytes are collected from superovulated donors by flushing the oviducts and the exogenous gene construct is injected into a pronucleus of zygotes before transfer into the oviducts or uteri of synchronised recipients. The resultant progeny then have to be tested for integration and expression of the foreign gene. The success in using this technique was first shown in mice by Gordon (1980), and followed later by Brinster (1981) and Constantini (1981). Furthermore, evidence was provided that at least some of the transgenes could be expressed (Brinster *et al.*, 1981) and importantly that the foreign genes were incorporated into the germline (Gordon *et al.,* 1981; Palmiter *et al.,* 1982 and 1983). Additionally, Palmiter (1982) demonstrated the transmission of the transgene to 50% of the founder animal's offspring, suggesting that these genes are stably integrated into one of the chromosomes.

1.2.1 Mechanism of integration of a transgene by microinjection

The average frequency of success (transgenic animals per eggs injected and transferred) is estimated to be 2-5% for mice and only 0.59% for pigs and 0.74% for sheep (Grosveld and Kollias, 1992). Such low frequencies constrain the adoption of this technology by the livestock industry. Although improvements in micromanipulation and microsurgical techniques for embryo re-implantation into pseudopregnant recipients may achieve a greater success rate in the future, the major constraint in using the microinjection technique remains that the mechanism for foreign DNA integration into the genome is not fully understood and occurs in a random and uncontrollable manner. The integration of one or more copies usually occurs at a single chromosomal site and if multiple copies are integrated they usually occur tandemly as a head to tail array. Multiple copies are thought to promote the inactivation of the inserted gene, via DNA methylation (Mehtali *et al.*, 1990) and/or heterochromatin formation (Dorer *et al.*, 1994; Garrick *et al.*, 1998).

It was thought that random chromosomal breaks may serve as integration sites for the foreign DNA (Brinster *et al.*, 1985), but the frequencies that different chromosomes are targeted suggests that there may be preferential sites for integration, although the number characterised so far is small (Rijkers *et al.*, 1994). At the site of transgene integration the host cellular sequences have frequently undergone multiple rearrangements such as duplications, inversions, deletions and even chromosomal translocations. A number of workers have characterised the junctional sequences (Mark *et al.*, 1992; Mu Chen. *et al.*, 1995). One relationship found has been the presence of very short (1-6 nucleotide), regions of homology at some genome-transgene junctions. Also, in a number of transgene integration sites, one feature of the cellular DNA which has been found is the presence of topoisomerase 1 consensus sites (Wilkie and Palmiter, 1987; Hamada *et al.*, 1993). Eukaryotic type 1 topoisomerase induces single strand nicks in DNA and catalyses rejoining of the 3' end of the nicked DNA to the 5' hydroxyl of another DNA fragment. If it cut close to another nick it could produce double strand breaks.

Topoisomerase 1 has been proposed to have a role in strand exchange during recombination (Halligen *et al.*, 1982).

1.2.1.i Mosaicism

Pronuclear injection can generate hemizygous or mosaic embryos with respect to the transgene depending on when the injected DNA integrates into the genome (Wilkie *et al.*, 1986). If the integration event occurs before the first round of DNA replication, hemizygous transgenic animals are produced. If integration occurs after replication mosaic embryos are produced. Depending on whether the trophoectoderm and/or inner cell mass (ICM) cell lineages of the blastula carry the transgene determines if the animal will be transgenic. Only those embryos with ICM cells which carry the transgene have a chance of generating a transgenic animal.

1.2.2 Position effects

Gene expression can be influenced by the chromosomal position of the gene, and this is termed a position effect. The eukaryotic genome is arranged into functional domains that influence gene expression. Euchromatin regions are highly transcribed, whereas heterochromatin regions are not. Transcriptionally quiet regions are often close to centromeres (eg. *Drosophila melanogaster;* Hazelrigg *et al.*, 1984) or telomeres (eg. *Saccharomyces cerevisiae;* Gottschling *et al.*, 1990).

It has been demonstrated in mice that the site of integration of the transgene influences its expression (Nandi *et al.*, 1988). For example, some transgenic mice produced do not express the foreign gene at all, which may be due to its integration into heterochromatin domains or its proximity to a silencer element. Of those mice that do express the gene, the levels of expression are influenced by the surrounding chromosomal context, independent of copy number. There is also the possibility of the foreign DNA integrating into an endogenous gene and disrupting its function, or inappropriately activating a gene by providing an

enhancer/promoter element. Creating animals homozygous for the transgene by interbreeding of the F1 progeny can reveal undesirable secondary effects, such as severe abnormalities and even lethality depending on the site of integration. Such an undesirable result is costly in both time and money. Some of the transgene expression problems have been overcome by including regulatory elements such as the β-globin locus control region (Grosveld *et al.*, 1987) or 'insulator' elements which protect the transgene from the effects of nearby regulatory elements (Eissenberg and Elgin, 1991).

1.2.3 Summary

Currently microinjection has several major limitations that thwart its widespread use in creating transgenic animals for the agricultural industry, these primarily being:

1. inefficiency,
2. the random site of integration,
3. the variable copy number,
4. expression of the transgene is subject to position effects,
5. the possibility of insertional inactivation of an endogenous gene.

It becomes apparent that new approaches must be developed and optimised to overcome these limitations for transgenesis to become a viable option for the agricultural and biomedical industries. Approaches being undertaken include the use of embryonic stem cells, homologous recombination and, relevant to the work presented in this thesis, the targeted insertion of transgenes using site-specific recombination systems.

1.3 Embryonic Stem Cells and Transgenesis

1.3.1 Embryonic stem cells

Murine embryonic stem (ES) cells are pluripotent (each can give rise to a whole embryo) stem cell lines established from the inner cell mass of the 3.5 day blastocyst (Jackson *et al.*, 1989). The cells retain their pluripotency following culture (in the presence of the growth factor leukaemia-inhibiting factor) and manipulation *in vitro*, and upon re-implantation into a host mouse blastocyst can contribute to all cell lineages, including the germline of the resulting chimeric animal. Results have shown that the cultured ES cells can recolonise embryos efficiently and integrate uniformly to give a widespread distribution in the resulting animals (Robertson *et al.*, 1986). Evans and Kaufman (1981), and independently Martin (1981), were the first researchers to demonstrate the establishment in culture of ES cells from early mouse embryos. The technique of culturing ES cells has also been established for the hamster (Doetschman *et al.*, 1988). Porcine ES cells have been cultured *in vitro*, however conditions and techniques for maintaining porcine (and bovine and ovine) ES cells remain to be optimised (Shim *et al.*, 1997), and it has not yet been demonstrated that they can colonise a host blastocyst to form germline chimeras (pers. comm. Dr. P Wigley, BresaGen, Australia). Technology in this area is advancing rapidly as demonstrated by the recent report of the isolation of human pluripotent stem cells (Shamblott *et al.*, 1998).

1.3.2 Transgenic ES cells

Embryonic stem cells provide an invaluable tool for manipulating the mammalian germline. Modified murine ES cells can be used as a vehicle to generate mice of any desired genotype. The major advantage of creating transgenic ES cells is that they can be identified and characterised before injection into host blastocysts to generate chimeras. This directly contrasts with the conventional method of transgenesis discussed in section 1.2 whereby transgenic animals can only be identified at the post-implantation stage, either by characterising tissue from the animal post-natally or from the foetus *in utero*.

Genes can be introduced into ES cells by standard transfection methods but, as with microinjection into fertilised eggs, this results in random integration of the DNA into the genome, giving rise to variable gene expression due to chromosomal position effects and copy number variations between clones. The site-specific integration of a single copy of the transgene into a pre-selected site in the genome should provide a means of creating clonal transgenic lines of animals with identical gene expression patterns.

1.3.3 Homologous recombination: Gene targeting in ES cells

Gene targeting in ES cells allows the modification of any cloned gene to be introduced into the germline of an animal. Briefly, a target vector containing the modified gene, flanked by homologous chromosomal DNA, is introduced into ES cells by standard transfection methods. Since there is a bias towards random insertion, in the order of 10^1 to 10^3 fold higher than that of insertion by recombination between the introduced DNA and the complementary sequence present in the genome (Danks, 1994), only in a small percentage of the cells does the targeting vector pair with the cognate chromosomal DNA sequence and by homologous recombination the modified gene is transferred to the genome (Thomas *et al.,* 1987; Capecchi *et al.,* 1989; Joyner, 1991 and 1992). Enrichment and screening procedures can be used to select only ES cells containing the correctly integrated modified gene. For instance, a promoter trap targeting vector, designed to use the transcriptional machinery of the endogenous target gene to drive the positive selection cassette (cloned in-frame with the endogenous translated product) in the targeting vector can yield an enrichment of about 100-fold for targeted clones (Joyner, 1992). The correctly targeted ES cells are then clonally expanded, maintained as pure populations and injected into host blastocysts. The resulting animals are chimeric, consisting of cells from the host blastocyst and the altered ES cells. If there is contribution to the germline, heterozygotes can be generated, and interbreeding of the heterozygous siblings will generate animals homozygous for the gene.

1.3.3.i Antibiotic resistance: A cautionary note

Screening procedures used to select for homologous recombination events involve the inclusion of a genetic marker in the target vector that produces a selectable change in cell phenotype. One commonly used selectable marker is the bacterial neomycin resistance (*neo*^{R}) gene. Provided the target vector has significant homologous DNA (> 1-2 kb) on both sides of the altered gene, and overall contains 5-10 kb of homologous DNA, at least 10-20% of the cells, and possibly as high as 80%, that are *neo*^{R} will contain the correctly targeted mutation.

Although homologous recombination will minimise the problems of random integration and variable copy number effects, a successful targeting event often results in the inclusion of a selectable marker such as the *neo*^{R} gene in the genome. Creating transgenic animals of agricultural importance with antibiotic resistance markers may not be permitted by regulatory bodies such as the Genetic Manipulation Advisory Council (GMAC) in Australia. Neomycin is still used today for the treatment of some human disease states (Vita *et al.*, 1998; Connolly *et al.*, 1997), therefore to preserve the life-saving potential of antibiotics, the spread of resistance genes at all levels needs to be halted. Refraining from the use of selective antibiotics in animal husbandry will help to prevent the entry of transmittable antibiotic-resistance genes into the human food chain (Perreten *et al.*, 1997). Precise gene insertion without incorporating a resistance gene could be achieved using site-specific recombination systems.

1.4 Male Stem Cell-Mediated Gene Transfer

A novel approach for potentially producing transgenic animals was reported by Brinster and Zimmermann (1994) and by Brinster and Avarbock (1994) and involved the use of male stem cells. Both report that stem cells isolated from the testes of donor mice could repopulate sterile testes of a recipient when injected into seminiferous tubules.

Preceding puberty and continuing through adult life, stem cell spermatogonia undergo continuous replication in a process known as stem cell renewal. Additionally, a small fraction of this population undergo differentiation to produce spermatozoa. After fertilisation with the female counterpart, their genes contribute to embryogenesis and ultimately to the differentiation of every cell in the body. Stem cell spermatogonia are basically the only self-renewing cell type in the adult capable of genetic contribution to the next generation, and in this sense can be thought of as being totipotent. Brinster and Avarbock (1994) reported that progeny can result from transplanted donor spermatogonia. The production of offspring from transplanted stem cell spermatogonia indicates this method may provide an additional path to producing transgenic livestock. One drawback of this technique is that currently there is no method to culture the cells *in vitro,* with drug selection, following transfection with exogenous DNA. Recently Kim, *et al* (1997) addressed this issue and reported on a new method of transferring foreign DNA into male germ cells of both mice and pigs. Using liposome/bacterial LacZ gene complexes they injected directly into the seminiferous tubules. In both mice and pigs DNA was incorporated into male germ cells and, in mice, 7-13% of epididymal spermatozoa were confirmed to contain foreign DNA. No offspring have been reported to be produced by this method as yet, however it may become a powerful technique for producing transgenic livestock in the future given the current absence of ES cells for livestock species.

1.5 Nuclear Transfer and 'Reprogramming'

Over the last decade nuclear transfer techniques have been developed for a number of mammalian species including sheep, cows, pigs, rabbits, mice, and, more recently, Rhesus monkeys (Meng *et al.*, 1997) and goats (Yong and Yuqiang, 1998). The technique involves transferring a nucleus to an enucleated (chromosome-free) unfertilised egg and using electrical pulses to fuse the two together. As a consequence of this process the donor nuclei is 'reprogrammed' by factors present in the cytoplasm of the egg to behave as a zygotic nucleus. The

exact molecular mechanism involved is not known but is speculated to involve DNA modifications such as methylation, involved in the regulation of gene expression (Kono, 1997).

Willadsen (1986), was the first to demonstrate this technique in sheep and later in cows (1989), and to date, it is in sheep and cows that the method remains the most successful. Recently Campbell *et al*, (1996) showed that lambs could be produced from an established cell line of sheep foetal derived cells. Previously, only early embryo derived cells have been used successfully as the nuclear donors in other species. Campbell's technique involved serum starvation of the donor nuclei cells before transfer to induce quiescence, which may help in the nuclear reprogramming. Hence, the significance of this report was that an already differentiated cell could be reprogrammed, essentially back in time, to early embryo status. More significantly, Wilmut and colleagues (1997) used the same method and demonstrated nuclear transfer using an adult mammary gland cell resulting in the birth of the famed 'Dolly'. Although at present time a rather inefficient procedure (1 in 277 fusions), it remains a scientific breakthrough in the area of cloning. Wakayama and coworkers (1998) have recently demonstrated the creation of cloned mice (full-term: 2-2.8%) by introducing the nuclei from adult somatic cells (cumulus cells) into enucleated oocytes, and following closely behind, Kato's research team (1998) reported the successful cloning of cows using adult somatic cells. Both reports illustrate that the technique of nuclear transfer, using terminally differentiated somatic cells, is likely to be possible in other mammals, including pigs.

Despite the success of nuclear transfer in many livestock species using non-terminally differentiated cells, the efficiency remains low, especially for the pig. In cows and sheep the efficiency is around 1% of fusions, and even lower than this for pigs (pers. comm. Dr. M. Nottle, BresaGen, Australia). With further research a greater understanding of the processes involved in reprogramming will be gained, hopefully leading to greatly improved efficiencies.

1.5.1 Homologous recombination: Gene targeting in cultured cells

Schnieke and coworkers (1997) have demonstrated the feasibility of using homologous recombination in primary (untransformed) somatic cells (foetal fibroblasts) to introduce a specific gene, human factor IX, into the germline of sheep. Furthermore, they demonstrated the production of transgenic sheep by nuclear transfer from foetal fibroblasts identified as containing the correctly targeted transgene.

One can envisage a time in the future where perhaps, by firstly utilising the technique of homologous recombination with an established somatic cell line, and followed by nuclear transfer, one could produce clonal transgenic pigs carrying a precisely located, single-copy genetic change. The advantages of this approach are, as for gene targeting in ES cells, there is a single-copy insertion and the cells can be characterised prior to the nuclear transfer step. Therefore, although a frequency of 1-2% for the nuclear transfer efficiency (Schnieke *et al.*, 1997) may seem low, in comparison to ~5% for microinjection [number of transgenic animals per total animals born (Damak *et al.*, 1996)] in fact 100% of the offspring are transgenic making this approach more economical than microinjection as there is no gestation of non-transgenic animals. However, one disadvantage with using homologous recombination to insert genes into livestock species is the time consuming task of characterising and cloning isogenic DNA (from the same strain; te Riele *et al.*, 1992) suitable for constructing a targeting vector containing extensive regions of homology. Furthermore, every new transgene would require another homologous recombination step.

1.6 Site–Specific Recombination Systems

1.6.1 Introduction

Genetic recombination can be defined as the production of a new DNA molecule from two parental DNA molecules involving the physical reorganisation of the

DNA sequences involved. Genetic recombination can be further subdivided into homologous recombination and site-specific recombination. In homologous recombination there is a requirement for extensive homology between the two reacting DNA substrates (Thaler and Stahl 1988; Smith, 1988). Homologous recombination occurs naturally during meiosis and mitosis of eukaryotic cells, and in the last decade it has been exploited, with great success, to achieve targeted gene disruption in mouse ES cells (Craig *et al.*, 1988; reviewed in Joyner, 1991).

Site-specific recombination involves specific DNA sequences on at least one of the two interacting DNA molecules. Important to my thesis research are site-specific recombinase systems requiring specific DNA sequences on both DNA molecules (Sadowski, 1986; Gellert and Nash, 1987) and I will not discuss the other (transposition) class. The recombination is achieved by phosphoryl transfer involving precise strand cleavage and religation in the absence of any high-energy cofactors (Kikuchi, Nash 1979). Instead there is conservation of the bond energy in a covalent phosphotyrosine (integrases) (Argos *et al.*, 1986), or phosphoserine (resolvase/invertases) (Grindley, 1993) linkage.

1.6.2 The Cre and FLP site-specific recombination systems

FLP and Cre belong to the integrase family of recombinases. The enzymes recognise specific nucleotide sequences and function through a transient DNA-protein covalent linkage. The FLP protein is encoded by the 2-μm circle plasmid of *Saccharomyces cerevisiae*. Its *in vivo* function is to maintain equal segregation of plasmid molecules during mitosis (Volkert *et al.*, 1986). The FLP recombinase target site (FRT) has been defined to a 48 bp region within each of two 599 bp long inverted repeats on the plasmid. Each FRT site region encompasses three 13 bp repeats with the first inverted relative to the other two and separated from them by an 8-bp spacer (overlap) region (Jayaram, 1985; Cox, 1988). Deletion analysis has shown that the minimal, fully functional FRT site consists of 28 bp which includes the 8 bp spacer and 10 bp of each flanking repeat (Senecoff *et al.*, 1985; Gronostajski *et al.*, 1985). Cre is a bacteriophage P1-encoded recombinase that

catalyses recombination between two 34 bp repeats, called *loxP* sites. Its *in vivo* function is to dissociate bacteriophage dimers following replication (Ambreski and Hoess, 1983 and 1984). In both systems, intermolecular recombination between sites on different DNA molecules results in insertion, generating one DNA molecule. Intramolecular recombination of a DNA molecule with directly repeated sites results in excision, and between two inverted sites results in inversion of the DNA between the two sites. The non-palindromic nature of the overlap regions in *loxP* and FRT determines the alignment of the recombination sites (Hoess *et al.*, 1986; Senecoff *et al.*, 1988).

1.6.2.i The development and utilisation of Cre and FLP

Since the advent of homologous recombination 10 years ago many genes have been knocked out in mouse ES cells, however some of these are null mutations that result in developmental defects causing embryonic lethality. Recently it has been possible to mutate these genes and analyse the effect at a specific developmental stage or in a particular cell type or tissue. Rajewsky and coworkers (1993) introduced DNA, flanked by two *loxP* sites, into mouse ES cells via homologous recombination. Transiently expressing Cre they efficiently deleted the DNA between the *loxP* sites. Later (Rajewsky *et al.*, 1994) they produced a 'conditional' knockout by having Cre expression under the control of a T-cell-specific promoter and specifically deleted genomic DNA in the T-cell lineage. More recently Feil (1996) demonstrated inducible gene inactivation, and Tsurushita (1996) the engineering of synthetic antibody libraries, using the Cre/*loxP* system.

Although FLP has been shown to work in a number of eukaryotic systems [*Drosophila* (Golic and Lindquist, 1989), mosquitoes (Morris *et al.*, 1991), mammalian cell lines (O' Gorman *et al.*, 1991) and in murine ES cells (Jung *et al.*, 1993; Dymecki, 1996)], its usefulness has lagged behind Cre significantly due to its different thermostability. While Cre is optimally efficient at 37°C, FLP has an optimum near to 35°C, and unknown to many researchers until 1996 a

commercially available FLP expression plasmid had a mutation making the FLP protein thermolabile at temperatures above 30°C, (Buchholz *et al.*, 1996) and thus made its use in many cell lines, including ES cells very difficult. Recently, Buchholtz (1998) reported the generation of an improved FLP recombinase (FLPe), generated by cycling mutagenesis, with improved thermostable properties in *E. coli*, human 293 and mouse ES cells.

1.6.2.ii Excision versus integration

The major drawback with the FLP and Cre systems is the difficulty in controlling the direction of recombination. Cre and FLP can mediate excisive (intramolecular) and integrative (intermolecular) recombination. However, the integration reaction is very inefficient because the integrated DNA, which has target sites (FRT or *loxP*) at both ends, consequently has a tendency to undergo excision soon after it integrates. That is to say, the excision and integration reactions are reversible (Figure 1.1) and the excision reaction is highly favoured over insertion as it is intramolecular as opposed to intermolecular. Experiments in bacteria using FLP have shown stabilisation of an integrant by delivery of a pulse of FLP activity (Huang *et al.*, 1991). However, for the reason mentioned above (*1.6.2.i*), very little progress has been made with the FLP system in recent times. Encouragingly, Araki and coworkers (1997) used mutant *loxP* sites (Albert *et al.*, 1995) to stably integrate DNA into mouse ES cells and achieved a frequency of 16% integration compared to less than 0.5% for the wild-type *loxP* sites. Albert (1995) identified three sets of mutant *lox* sites that favour integration over the excision reaction and later made nucleotide changes into the left 13 bp element (LE mutant *lox* site, ◀▭) or the right 13 bp element (RE mutant *lox* site, ◇▬). Recombination between a LE mutant site and a RE mutant site produces a wild-type *lox*P site at one end of the integrated DNA and a double (LE+RE) mutant site at the other end. The LE+RE mutant site has a reduced binding affinity for Cre recombinase and therefore the inserted DNA is relatively stable. This added flexibility of the Cre system gives it greater attraction for future use in stably integrating transgenes, although if one could avoid using a reversible system in the

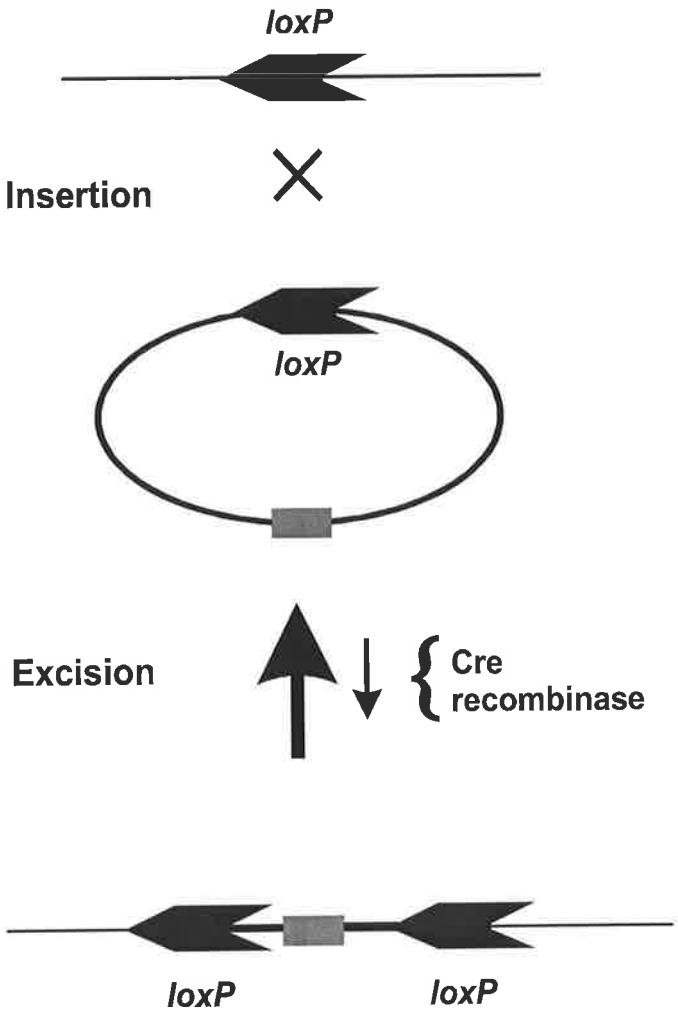


Figure 1.1 Recombination between *lox*P sites
The vertical arrows indicate the frequency and direction of the reaction. Integrated DNA is easily excised in the presence of Cre recombinase, hence the insertion product is unstable.

first place the frequency of integration events achieved may be even higher. Today in mouse ES cells, integration via homologous recombination, where the frequency of insertion events achieved, with selection, can be up to 50% or greater, remains the system of choice.

The major application of these systems is clearly in the area of excision. For instance, when used alone, homologous recombination often results in the inclusion of a selectable gene like *neo*^{R}, but when complemented with a subsequent Cre or FLP mediated excision reaction the bacterial antibiotic resistance gene can be removed. In the future this may permit the generation of transgenic livestock species able to be approved by regulatory bodies such as GMAC. Further, both Cre and FLP have been demonstrated to work in fertilised mouse eggs (Araki *et al.*, 1995: Ludwig *et al.*, 1996: Sunaga *et al.*, 1997; de Wit *et al.*, 1998) which suggests they may be exploited to manipulate the genome of animals for which ES cells are currently unavailable. However, the target site must first be inserted into the animal's genome either randomly by microinjection, or via homologous recombination. To date, there have been no reports of homologous recombination in fertilised eggs of any livestock species (pers. comm. Dr. I. Lyons, BresaGen, Australia) and only two literature reports of using homologous recombination in fertilised mouse eggs (Brinster *et al.*, 1989: Susulic *et al.*, 1995) in which the success rate was very low, 1 in approximately 500 transgenic mice and moderate, 1 in 23 transgenic mice respectively, and furthermore many reports of other unsuccessful attempts have been made (pers. comm. Dr. I. Lyons, BresaGen, Australia). Consequently the use of homologous recombination in fertilised eggs to insert a target site at a specific genomic location is not an available option at present.

1.6.3 The temperate coliphage 'directional' site-specific recombination systems

The temperate coliphages like λ, P2 and 186 (Bertani and Bertani, 1971) can enter either a lytic or lysogenic lifecycle upon infection of their *Escherichia coli* host

(Figure 1.2). Following adsorption to the cell's surface, the phage injects its DNA into the host where it circularises and, if the lytic pathway is selected, early transcription begins, leading to middle gene expression. This results in expression of the phage DNA replicative genes and the phage genome is replicated. In the final stage of lytic development the expression of the late genes allows synthesis of phage morphogenic proteins, proteins required for phage assembly and proteins needed for host cell lysis. The phage, once assembled, lyse the host cell releasing about 100-200 progeny phage per lytically infected cell. Occasionally the phage genome becomes integrated into the host cell DNA (prophage), the lysogenic state, allowing automatic replication each time the bacterium divides. The prophage state is extremely stable until repression of lytic transcription is removed (derepression), and the phage DNA is excised from the host genome where it can re-enter the lytic pathway. The integration and excision reactions of the phage DNA involve site-specific recombination between sequences present on the bacterial and phage genomes.

1.6.3.i Excision versus integration

The best characterised directional site-specific recombination system is that of λ (Figure 1.3). The integration reaction requires two proteins: the phage-encoded integrase (Int) and the host-encoded integration host factor (IHF). Int catalyses strand exchange between the phage and bacterial attachment sites, called *att*P and *att*B respectively. This produces the composite sites *att*L and *att*R flanking the prophage. Excision requires Int, IHF and the phage-encoded excisionase (Xis) protein (for reviews: Craig, 1988; Landy, 1989). In addition, the presence of Xis inhibits integration while Fis, a host protein, stimulates excision when the concentration of Xis is limiting.

As different combinations of proteins mediate the excisive and integrative pathways these systems have the directionality lacking in the one protein systems like Cre and FLP. In the case of λ, in the absence of any Xis protein the integrative reaction is essentially irreversible (Ambreski and Gottesman, 1981). It

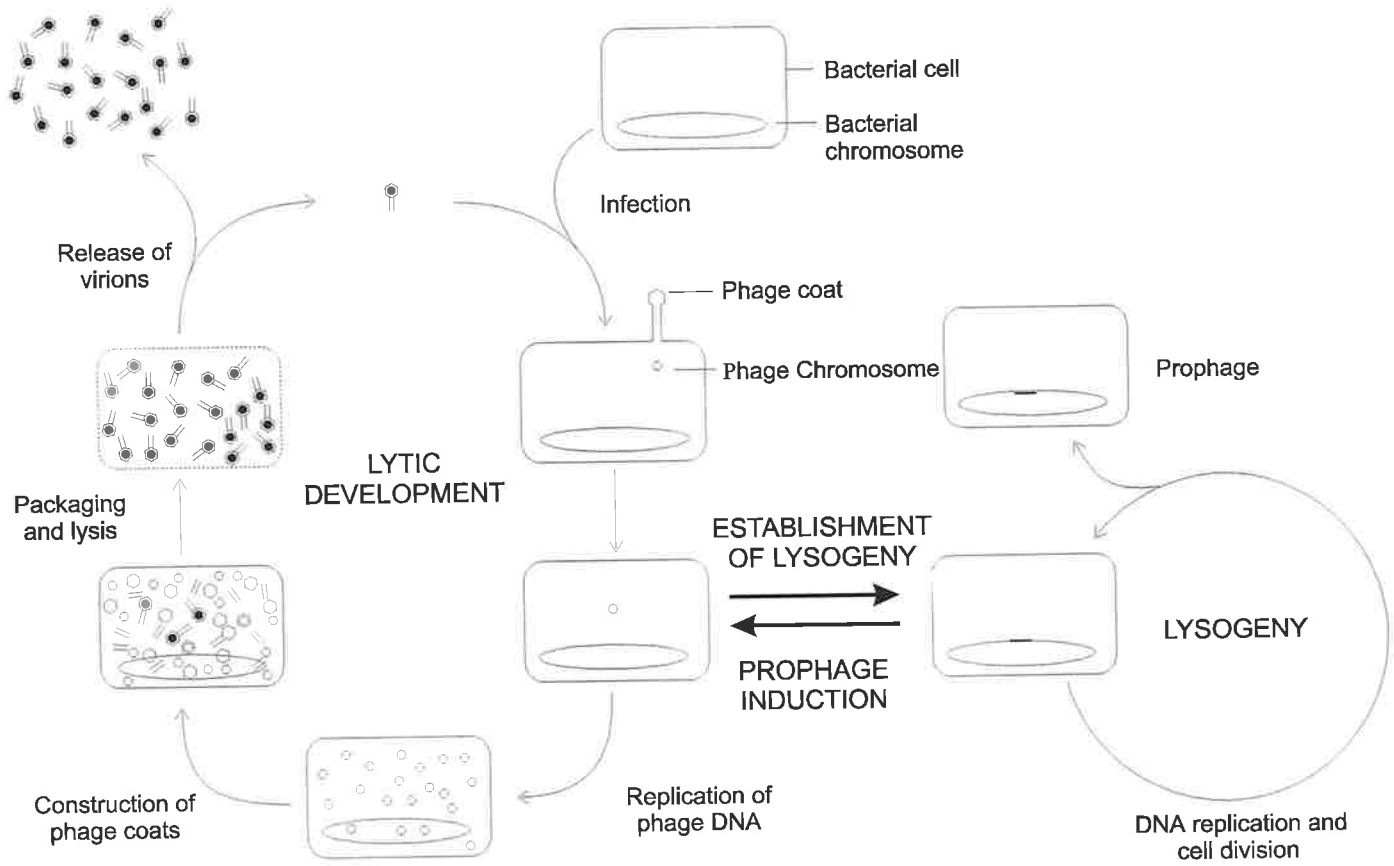


Figure 1.2 Life cycle of temperate, integrating bacteriophages

Following adsorption of the phage to the host cell, the phage DNA is injected into the host where it circularises and early transcription begins. If the phage is to develop lytically, a coordinated series of events then occurs: the phage DNA is replicated, the structural proteins are synthesised and assembled into phage coats, the replicated DNA is packaged into these coats, the host cell is lysed and the phage are released. If the phage is to enter lysogeny, lytic functions are repressed and the phage DNA is integrated into the host chromosome and becomes a prophage. The prophage is passively replicated with the host DNA and remains stably integrated until it is induced. During induction, lytic repression is removed (derepression) and the prophage is excised from the host chromosome, and it re-enters the lytic cycle.

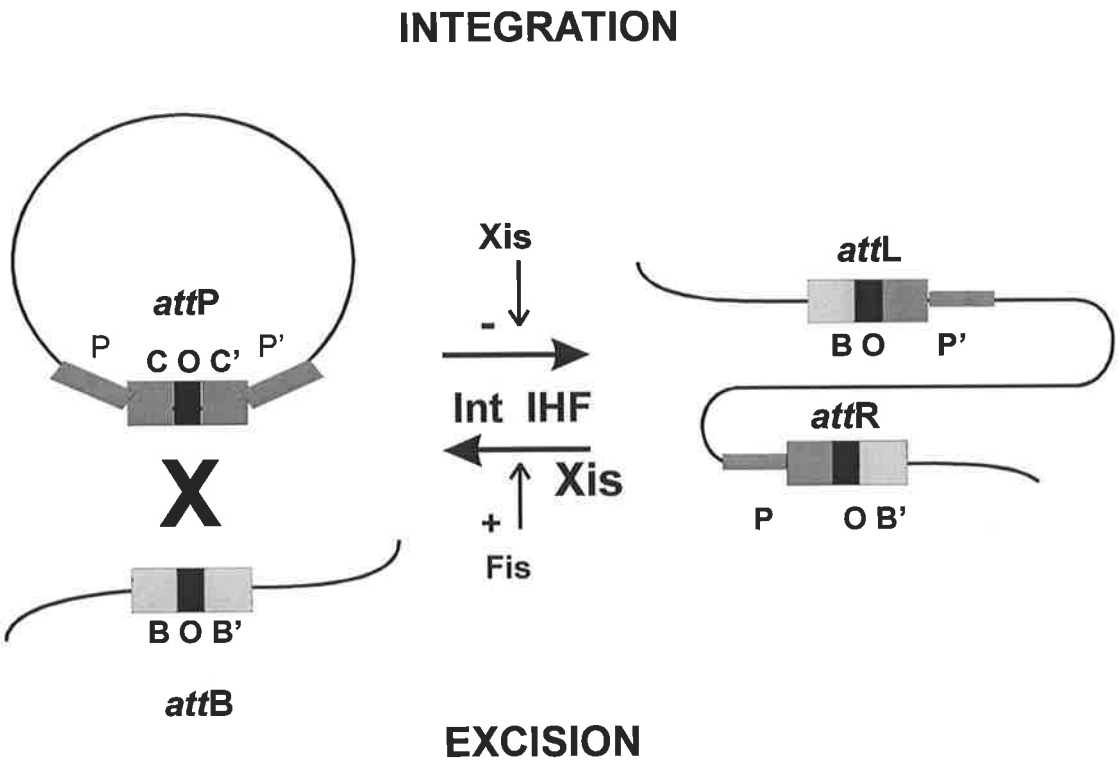


Figure 1.3 Integrative and Excisive Recombination of Phage λ
Site-specific recombination between the *att*P site on the phage genome and the *att*B site on the bacterial chromosome generates the composite sites *att*L and *att*R. The phage-encoded Int and *E coli* IHF proteins are required for excision and integration. In addition, Xis is required for excision and is also inhibitory to integration. Fis can stimulate excision when the levels of Xis are limiting.

is possible that the extreme directionality of these systems (eg. λ, P2 and 186) may be exploited for the development of targeted transgenesis systems. However, to date none have been shown to operate in a eukaryotic context.

1.7 The λ Site-Specific Recombination System (Integration)

The best studied system is that of λ, which will be described in more detail as the prototype system (review, see Landy 1989).

1.7.1 The attachment sites *att*B and *att*P

The *att*B and *att*P attachment sites share a region of homology of 19/21 base pairs (bp), the core sequence, within which recombination occurs. The sites of top and bottom strand exchange are separated by a 7 basepair 'overlap' region, designated 'O'. The sequence of the overlap region is not critical, with the exception of one basepair (Kitts and Nash, 1987), provided homology is maintained between the two reacting partners (Mizuuchi *et al.*, 1981; Craig and Nash 1983; Weisberg *et al.*, 1983). The common core sequence (COC') is flanked by non-homologous DNA designated P and P' in *att*P and B and B' in *att*B. The minimal *att*P and *att*B sites have been determined as 234 bp and 21 bp, respectively (Mizuuchi and Mizuuchi; 1980 and 1985). The P and P' arms contain the binding sites for Int (arm-type sites), IHF, Xis and Fis, while the core contains two core-type Int binding sites C and C', present as two imperfect inverted repeats at the junction of the core and arm sequences. The *att*B site consists of two core-type Int binding sites in the short flanking arm sequences. A striking feature of the *att*P sequence is its high AT content (73%), including several stretches of up to six consecutive adenines known to cause intrinsic curvature of DNA (Landy and Ross, 1977; Wu and Crothers, 1984).

1.7.2 Int

The λ Int protein is a 40 kD protein with two discrete DNA binding domains identified by limited proteolysis (Moitoso de Vargas *et al.*, 1988; Tirumalai and Landy, 1997). The amino-terminal domain (amino acid residues 1-64) specifically binds the arm-type sites with high affinity while the region of residues 65-169 binds the core-type sites, at the points of strand exchange, with low affinity. The remaining residues, 170-356, contain the topoisomerase activity of the Int protein. The accessory proteins, IHF, Xis and Fis, share one common feature, the ability to dramatically bend DNA upon binding (Craig and Nash, 1984; Yin *et al.*, 1985; Thompson *et al.*, 1987; Robertson and Nash, 1988; Thompson and Landy, 1988; Kosturko *et al.*, 1989). Their sites of DNA binding lie between the arm and core-type Int binding sites leading to the proposal that λ Int forms DNA loops by linking arm and core sites (Moisoto de Vargas *et al.*, 1989; Kim and Landy, 1992).

λ Int belongs to the Int family of recombinases, which includes Cre and FLP, that share a group of invariant residues, the arginine-histidine-arginine (λ: Arg212, His308, Arg311) catalytic triad (Argos *et al.*, 1986; Ambreski and Hoess, 1992; Han *et al.*, 1994) and the tyrosine nucleophile (λ: Tyr342). In all cases the conserved Tyr residue becomes covalently linked to the DNA strands during the recombination event (Pargellis *et al.*, 1988; Hoess and Ambreski, 1985; Gronostajski and Sadowski, 1985; Evans *et al.*, 1990). Two Int mutants, Int-h (E174K) and IntF (Y342F), have been characterised in detail. Int-h can catalyse recombination in the absence of IHF (Miller *et al.*, 1980) and IntF, in which Tyr342 is replaced by a Phe, retains its DNA binding properties but lacks the topoisomerase and recombinase activities (Pargellis *et al.*, 1988; Burgin and Nash, 1992). The recently determined crystal structure of λ Int revealed Tyr342 to lie on a flexible loop of 17 amino acid residues, placing it in reach of the RHR catalytic triad in either the *cis* or *trans* cleavage mode (Kwon *et al.*, 1997). For *cis* cleavage, a single protomer provides both the catalytic tyrosine and the RHR triad for cleavage of one DNA strand (Landy, 1993; Nunes-Düby *et al.*, 1994), whereas for *trans* cleavage the tyrosine from one protomer cleaves a DNA strand that is

bound and activated by the RHR triad of a neighbouring protomer (Chen *et al.*, 1992). The crystal structure results helped rationalise two conflicting reports on whether λ Int worked in *cis* (Nunes-Düby *et al.*, 1994) or *trans* (Han *et al.*, 1993), as the flexible loop, containing Tyr342, can structurally accommodate both mechanisms. Cre was thought to work in *trans* (Shaikh and Sadowski, 1997), however the recently determined crystal structure, which captured recombination in the act, shows unequivocally that Cre acts in *cis* (Guo *et al.*, 1997). HP1 integrase, a relative of the P2-186 family of temperate phages (Eposito *et al* 1996) has had its structure determined by crystallography (Hickman *et al.*, 1997) also revealing a *cis* mechanism of action. Much evidence supports a *trans* mechanism for FLP (Lee *et al.*, 1992; Lee *et al.*, 1994; Qian and Cox, 1995) but a crystal structure for clarification of the *trans* mechanism is required.

1.7.3 IHF

Integration host factor (IHF) belongs to the class of histone-like proteins of *Escherichia coli.* First discovered by Miller *et al* (1979) and Nash (1981) as being essential for the integration of λ DNA into the chromosome, it is now known to influence many cellular processes (review, see Friedman, 1988), including expression of the λ cII gene (Hoyt *et al.*, 1982), DNA replication (Gamas, 1986; Hwang and Kornberg, 1992), λ packaging (Kosturko *et al.*, 1986), and activation of transcription in *E. coli* (review, see Freundlich, 1992).

IHF is a small basic protein consisting of two subunits, the α-subunit (11.2 kD) encoded by the *himA* gene, and the β-subunit (10.6 kD) encoded by the *himD* (or *hip*) gene. The two proteins show 30% amino acid sequence identity, and share amino acid sequence homology to HU, another histone-like protein from *E. coli* (Drlica and Rouviere-Yaniv, 1987). Unlike HU, which binds DNA non-specifically, IHF recognises a specific sequence with the consensus WATCAANNNNTTR, where W is A or T, R is A or G and N is any deoxynucleotide (Craig and Nash, 1984), and CA are the only two bases absolutely conserved in all known IHF binding sites. An expanded binding site,

which includes A/T rich elements upstream (5') of the core consensus element has been determined by Goodrich *et al* (1990). Hales and coworkers (1994 and 1996), determined that the A/T rich region, while essential for one binding site in λ *att*P, is not essential for another, although it does enhance IHF binding if present.

Binding of IHF to λ *attP* occurs at three sites, and can occur in the absence of Int, whereas there are no IHF binding sites in *attB* (Craig and Nash, 1984). Protection and interference studies indicate that IHF contacts its binding site in the minor groove of DNA (Yang and Nash, 1989; Wang *et al.*, 1995). IHF's primary role is to bend DNA, facilitating the formation of other protein-DNA interactions (Prentki, 1987; Robertson and Nash, 1988; Thompson and Landy, 1988; Snyder *et al.*, 1989). Further, a recent report implicates IHF as a major contributor to chromosomal DNA condensation (Murtin *et al.*, 1998). It has been demonstrated that IHF can partially be replaced by other DNA bending proteins, such as HU and CRP (cyclic AMP receptor protein), and by sequence directed bends (Goodman and Nash, 1989 and 1992; Segall *et al.*, 1994; Perez-Martin *et al.*, 1994; Parekh and Hatfield 1996). The crystal structure of an IHF-DNA (35 bp) complex has been determined (Rice *et al.*, 1996) showing that the DNA wraps around the protein, bending the DNA by >160° and causing a 'U-turn' in the DNA structure. This result confirmed that IHF facilitates the simultaneous binding of λ Int to two distal sites.

1.7.4 The intasome

The first step in the integrative pathway involves the binding of Int and IHF proteins to their specific recognition sites in *att*P, forming a nucleoprotein structure termed the intasome. Formation of a functional intasome is dependent on the supercoiling of *att*P (Richet *et al.*, 1986). Although the core-type Int-binding sites are where Int catalyses strand exchange, specific Int binding to *att*B core sites is undetectable *in vitro* and hence, it is proposed that *att*B does not obtain its Int from solution, but rather that a naked *att*B obtains its Int from the *att*P-intasome (Ross and Landy, 1983; Richet *et al.*, 1988; Nash, 1990). How the intasome and

*att*B actually synapse remains undetermined; one model favours random collision and another the sliding of the *att*P-intasome along the host chromosome (Gellert and Nash, 1987).

1.7.5 Intermediates and the order of strand exchange

The use of substrates that trap the intermediates in the λ recombination pathway has established that recombination proceeds via a sequential mechanism, which produces a partially single-stranded exchange intermediate known as a Holliday junction (Holliday, 1964 and 1974). During the normal recombination reaction the intermediate forms have been too transient to be easily detected. Two approaches have been used to produce the intermediates. In the first, a nick was placed in the target DNA, either in the overlap region central to the two cleavage sites (Pargellis *et al.*, 1988), or at the top or bottom strand cleavage positions (Nunes-Düby *et al.*, 1987). A centrally placed nick identified Tyr342 as the specific residue of λ Int which covalently links with DNA. A nick placed at the bottom strand cleavage position produced a Holliday intermediate known as the α-structure, whereas a nick placed at the top strand cleavage position produced no intermediates or recombination products (Figure 1.4). This indicated that strand exchange proceeds in a defined order: first the top strand is cleaved and exchanged followed by cleavage and exchange of the bottom strand. The presence of the nicks at the cleavage sites affects the chemistry of the recombination reaction as it stops Int from using the energy from the phosphodiester bond. The second approach to producing intermediates used phosphorothioate nucleotides (dNTPαs) (Kitts and Nash, 1988). Again, when located at the positions of strand exchange the presence of a large sulphur atom in the dNTPαs, which replaces a non-bridging oxygen atom, affects the chemistry of the reaction, confirming that recombination proceeds via a Holliday intermediate and that there is a set order of strand exchange. Furthermore, a group of synthetic Holliday junctions in which the branchpoint was constrained to the centre of the λ 7 bp overlap region gave strong evidence to support a strand-swapping model whereby a central isomerisation step shifts the Int-DNA complex from top strands swapped to bottom strands swapped

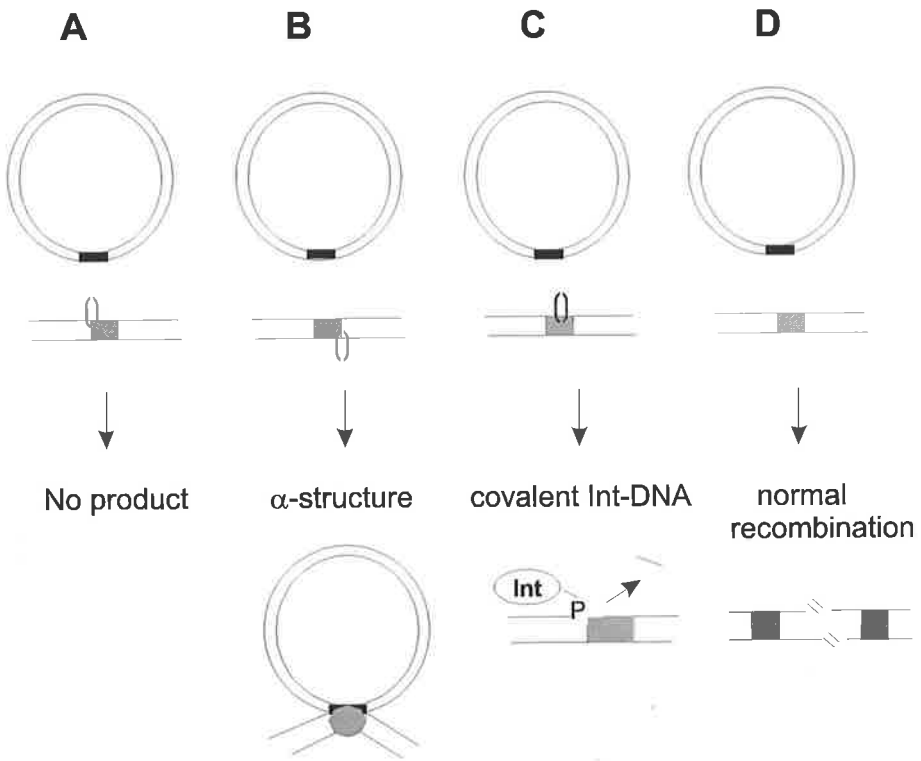


Figure 1.4 Suicide *att*B substrates and their products

In the supercoiled *att*P substrates, the black boxes represent the overlap region of the common sequence; in *att*B the overlap region is represented by the grey box. **A**. A nick placed in the top strand of *att*B at the first Int cleavage site produces no product. **B**. A nick placed in the bottom strand of *att*B at the second Int cleavage site produces a Holliday intermediate termed the α-structure. **C**. A medially placed nick in the *att*B overlap region produces covalent-Int DNA molecules. **D**. Normal recombination between *att*B and supercoiled *att*P proceeds in a defined order. First the top strands are cleaved succeeded by branch migration and strand exchange. Following this, a second round of strand cleavage, branch migration and strand exchange of the bottom strands takes place.

thereby favouring cleavage/exchange and resolution respectively (Nunes-Düby *et al.*, 1995; Azaro and Landy, 1997; Nunes-Düby *et al.*, 1997).

1.8 The 186 Site-Specific Recombination System

Coliphage 186 is a temperate phage, and belongs to a group of non-lambdoid phages including coliphage P2 and satellite phage P4 (Bertani and Six, 1988). At the commencement of my project the 186 site-specific recombination system was only partially characterised. The 186 Int gene is required for both integration and excision (Bradley *et al.*, 1975) and belongs to the integrase family of site-specific recombinases (Argos *et al.*, 1986). The excision reaction requires the *apl* gene product, IHF was demonstrated to be required for integration and Fis was not required for either reaction (M. Reed, Ph.D thesis, 1994; University of Adelaide, Dept. of Biochemistry). The 186 bacterial attachment site (Appendix, Figure 1) has been mapped between the *phe*A and *nal*B loci at the 57 min region of the *E. coli* chromosome (Woods and Egan, 1972; Reed, 1994). However, *in vitro* studies, using purified components for integration, 186 Int and IHF, had not yet been undertaken. Also, although the 186 *att*B site was thought to be positioned somewhere in the common sequence region of *att*P and *att*B, according to the Campbell model (1992), its exact location, sites of cleavage and length of overlap sequence was unknown. By comparison of *att*R and *att*L sequences, the common sequence has been located to position 681-716 of the phage sequence [numbered from the *Pst*I site at 65.5% of the 186 genome (Kalionas *et al.*, 1986)] giving a common sequence of 36 bp. In addition, the sequence identity between *att*R and *att*L continues leftward for a further 12 bp after a single base pair mismatch giving a 48/49 bp DNA sequence homology over the whole region (Appendix, Figure 1), and placing the *att*P site in the lysogenic operon.

1.9 Targeted Livestock Transgenesis

1.9.1 Current technology: Pro's and con's

Pronuclear microinjection of foreign DNA is the current technique used to produce transgenic livestock animals (Gordon *et al.*, 1980), and involves microinjection of several hundred copies of a specific DNA fragment into one pronucleus of each one-cell zygote. Normally linearised DNA is used for the microinjection step. Most of the microinjected DNA is rapidly degraded by cellular nucleases but in some cells a few molecules successfully integrate into one, or more rarely, multiple sites of the host chromosome. Inherent problems are that the integration site is random and transgene expression is influenced by the surrounding DNA context of the integration site. Further, the transgene can insert to form tandem or inverted arrays from one to several hundred copies per site and there is no correlation between transgene copy number and the expression level of the transgene.

The Cre/*lox* and FLP/FRT recombinase systems may still hold a place in transgenic livestock production either via microinjection into fertilised eggs or in a somatic cell line. As with all the recombinase systems, including λ. P2 and 186, their target sites could be positioned into a somatic cell line via homologous recombination and thereafter, once in place, the target site could be used to insert various transgenes using the appropriate site-specific recombination system. Using mutant *lox*P sites Albert *et al.*, (1995) achieved a 16% integration efficiency using mouse ES cells, and utilising a promoter trap targeting vector system (1.3.3) Kolb *et al.*, (1997 and 1999) achieved similar integration rates with Cre. If any of the directional recombinase systems can be shown to operate in a mammalian environment, either via microinjection into fertilised eggs or in a somatic cell line, it is likely the integration frequencies achieved would be significantly higher due to their directionality. Furthermore, the recombinase systems may offer an alternative route to homologous recombination for the initial introduction of the target site, and that is by using closely matching naturally occurring target sites in combination with an altered integrase, one could bypass the homologous recombination step altogether.

This would offer a significant saving in both cost and time as the construction of isogenic targeting constructs for different livestock strains is a major impediment to the homologous recombination route.

Such a strategy would be ideal for human gene whereby gene targeting in cultured somatic cells (or haematopoietic stem cells) is potentially a very powerful tool for treating human genetic diseases permitting one to make specific modifications in endogenous genes to correct mutations. Recently, as an antiviral therapy Lee and Park (1998) have postulated altering Cre recombinase to recognise the closest matching '*lox*P' sites in the 5' and 3' long terminal repeats (LTRs) which surround human immunodeficiency virus type 1 (HIV-1) proviral DNA to enable its excision.

1.9.2 Towards targeted transgenesis using directional recombination systems

1.9.2.1 Current applications for directional recombination systems

The recombinases (integrases) of temperate phages such as λ, P2 and 186 have a significant advantage over the use of Cre or FLP: they can insert DNA stably, as all depend on additional proteins for excision. One aim of this project is to develop a directional recombinase system for use in higher eukaryotes for either the precise, single copy placement of transgenes, or inactivation of a gene, using the wild-type 186 attachment sites and Int/IHF proteins. However, apart from mice and cultured cells, utilisation of this technology in higher eukaryotes relies on the concomitant development of methods to firstly introduce the wild-type *att*B target site into the genome either via homologous recombination in fertilised eggs, or in ES cells, or perhaps even in somatic cells, followed by nuclear transfer and reprogramming. To date the first two methods do not exist for any livestock species, although with future research it is likely one or both may well do. So far homologous recombination in somatic cells followed by nuclear transfer and reprogramming has only been successfully shown for sheep.

1.9.2.ii Development of 'modified' directional recombination systems

Until ES cell technology and homologous recombination in livestock animals, especially in the case of pigs, is successfully developed, another approach is to use existing 'natural' sites for site-specific recombination. We propose that mammalian genomes have sequences that are closely related to the target sequences of different integrases and that we can modify the specificity of the integrases to recognise a native mammalian target sequence (mtBOB') and to insert a transgene carrying an equivalently modified mtPOP' site. This would provide us with a 'stable' of integrases capable of operating on different, naturally occurring *att* sites. Precedents exist for the modification of DNA binding proteins to identify mutated binding sites. Jamieson and coworkers (1994) demonstrated that a DNA binding protein (Zif268) could be randomly mutated to recognise an altered DNA binding site. They generated and selected altered Zif268 proteins that could bind to an altered DNA target site. Furthermore, it has been demonstrated that although the Int proteins encoded by the bacteriophages λ and HK022 are 70% identical, and neither recombines the attachment sites of the other phage, mutants of λ integrase could be selected for increased ability to recombine HK022 sites (Dorgai *et al.*, 1995) and *vice versa* (Yagil *et al.*, 1995). Recently, Hartung and Kisters-Woike (1998) constructed a Cre mutant which had reduced recombination efficiency with the wild-type *lox*P target but was able to recombine an altered *lox*P target sequence, in contrast, the wild-type Cre was incapable of recombining with the altered *lox*P site.

For our work, we could undertake computer analysis of the pig GalT gene (see 1.1.1) and select potential target sites (mtBOB's) for a particular integrase, in which we have already characterised which bases in the native BOB' site are essential for Int action. In λ, for the 21 bp native BOB' site, not all the bases contribute to the λ Int binding determinants (Nagaraja and Weisberg, 1990; Dorgai *et al.*, 1998), thus allowing us to select closely matching sequences in the GalT locus to the actual integrase binding determinants, and then to modify λ Int to recognise these. It has been shown in λ that only one basepair of the 7 bp overlap

region is essential and cannot be altered even if homology is maintained between POP' and BOB' therefore providing an extra 6 bp with which to match to GalT sequences.

Theoretically, we could use the pig GalT gene sequence to dictate the sequence of the 7 bp overlap region in mtPOP'. Potentially such technology could be used to inactivate the pig GalT gene by microinjection of Int, IHF and the engineered mtPOP' into fertilised pig oocytes, and more widely, to insert transgenes, or inactivate genes in any livestock species. A further potential application of this technology is the insertion of genes into human cultured cells for somatic gene therapy (for reviews see Verma and Somia, 1997; Choo, 1997).

However, the *att* target site would need to be unique within a particular mammalian genome. The pig genome is approximately $3x10^9$ bp. A natural site would require 16 bp to be a unique *att*B site ($4^{16}=4x10^9$). Since we can use the 'O' region sequence present in a natural target site to dictate the 'O' region sequence in mtPOP' (in λ only one basepair of the 7 bp overlap region is an essential Int determinant and cannot be altered even if homology is maintained between POP' and BOB') then we require a maximum of 16-6=10 bp ($4^{10}=1x10^6$) for Int binding determinants. If a close match to the maximum number of essential bases, for instance 6-8 bp of the 10 bp required, can be found in a target gene then the Int protein can be modified through successive rounds of mutagenesis (each time changing a base in the wild-type *att* site toward the natural site) to recognise the chosen genomic *att* site. A match of 6 bp would mean almost every gene, or $4^6=4096$ bp, of a higher eukaryotic genome would contain a potential natural target site. However, remembering that further specificity can be dictated by the 6 bp of the overlap region ($4^{6+6}=4^{12}$) we therefore theoretically need only modify the recognition of 4 basepairs to obtain a unique (4^{16}) *att*B site. This number could be reduced further if a 7-8 bp match to the 10 bp required is found. We are optimistic at achieving a level of modification of 3 to 4 basepairs, especially since Jamieson and coworkers (1994) randomly mutated amino acids in the zinc finger protein Zif268, and selected altered proteins that bound to a DNA site altered by 3

basepairs (eg. GCG TGG ATA, rather than the native site GCG TGG GCG). An alternative to mutagenesis would be to have available a stable of 256 (4^4) different integrases each with 10 bp determinants to target specifically a gene of 4^6 bp, with 6 bp of the overlap region dictated by the target.

Considering an integrase with fewer determinants, eg. 8 bp, then we would require a stable of 16 integrases, or optionally, modification of the integrase to accommodate 2 bp, in order to target a genomic gene of 4096 (4^6) bp of coding sequence, with 6 bp of the overlap region dictated by the target. On average we would find such a site 11 ($3x10^9 \div 4^{14}$) times in the genome, necessitating a screen of 11 cells to isolate the correctly targeted event. Furthermore, one could combine the approach and use mutagenesis to accommodate a single bp in the 8 bp target sequence of four integrases to gain the same result.

The following Table lists the calculations for the range of integrases with determinants from 6 to 10 bp.

N° DETERMINANTS	N° BASES INVOLVED FOR MODIFIED INT	N° INTEGRASES IN STABLE	N° SITES IN $3x10^9$ bp
6	–	1	180
7	1	4	45
8	2	16	11
9	3	64	3
10	4	256	1

1.10 Scope Of This Thesis

The λ site-specific recombinase system is the best characterised system at this point in time. However we plan to characterise a number of phage systems to

provide a 'stable' of integrases with greater flexibility for gene insertion. The temperate bacteriophage 186 is one such system. At the beginning of my project the 186 site-specific recombinase system was only partially characterised and *in vitro* studies, using purified components for integration, 186 Int and IHF, had not yet been undertaken. Also, although the 186 *att*B site was thought to be positioned somewhere in the common sequence region of *att*P and *att*B, according to the Campbell model (1992), its exact location, sites of cleavage and length of overlap sequence was unknown. Moreover, at the start of this project it was not known if any of the integrases, such as λ, P2 and 186, could work in a eukaryotic system, in the context of chromatin. While a colleague (B. Hopwood) was investigating the λ system, to determine if a directional recombination system could direct gene insertion in eukaryotic cells, I commenced characterisation of the 186 system.

To this end, *in vitro* studies were conducted with cell-free and purified components. These results are presented in Chapter Two, and demonstrate that 186 integration requires only Int and IHF proteins. Other characteristics of the reaction were also investigated. Chapter Three describes a detailed analysis of the 186 *att*B site. I defined the location of *att*B within the common sequence, using deletion analysis, and precisely located the points of strand exchange using phosphorothioate-containing oligonucleotides representing the BOB' site. Furthermore, some initial mutagenesis studies were done to ascertain the Int binding determinants and therefore the potential number of bases in BOB' which would contribute to the uniqueness of a mammalian target site, mtBOB'. Chapter Four describes some initial investigations to test whether the 186 system can work in a eukaryotic cell. I first created lines of mice transgenic for the 186 *att*B site, and later addressed whether microinjection of purified Int, IHF and an *att*P-containing donor substrate into mouse fertilised eggs could result in the precise and stable insertion of DNA into a eukaryotic chromosome.

CHAPTER 2

In vitro studies of the 186 site-specific integration system

Chapter 2

In vitro studies of the 186 site-specific integration system

2.1 Introduction

The bacteriophage 186 site specific recombination system is analogous to those of λ and P2. In the P2 system the P2 Int protein and *E. coli* IHF are required for integration, while the Cox protein provides directionality as the excisionase and, in contrast to λ, *E. coli* Fis protein is not required, and does not enhance excision (Yu and Haggård-Ljungquist, 1993a; Yu and Haggård-Ljungquist, 1993b). For 186 the phage-encoded Int and *E. coli* IHF proteins catalyse insertion of the 186 genome into a precise location in the *E. coli* chromosome. Insertion is dependent upon proteins binding to two attachment sites, *att*P and *att*B. Excision of the prophage only occurs in the presence of Int and another phage-encoded protein, Apl. Hence inserted DNA remains stably integrated in the absence of Apl (Reed, 1994).

For use in a site-specific recombination assay system, based on that of Leong *et al.*, (1985), Michael Reed (1994) from our laboratory constructed a plasmid, pFLIP. This plasmid contains the phage and bacterial attachment sites orientated such that recombination between the two sites inverts the DNA segment between them (Figure 2.1). Two *Xho*I restriction sites, one within the invertible fragment and the other in the uninvertible region of the plasmid, permit monitoring of the recombination reaction by determining the levels of the integrative form pFLIP(I) and the excisive form pFLIP(X). The substrate pFLIP allowed *in vivo* analysis of the directionality determinants of the 186 site-specific recombination system. Reed (1994) showed that *E. coli* IHF^+/Fis^- cells carrying a source of Int protein were capable of displaying intra-molecular recombination between POP' and BOB'

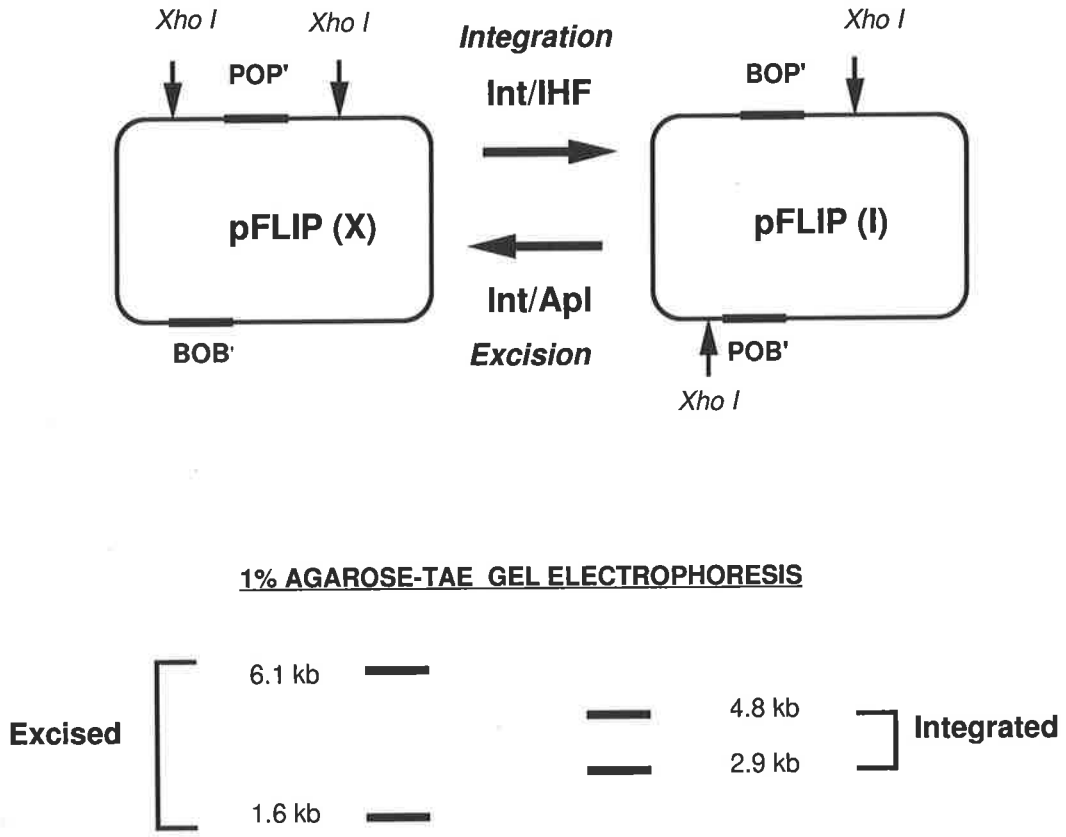


Figure 2.1 Schematic representation of the *cis* recombination assay. Shown are the excised, pFLIP(X), and integrated, pFLIP(I), forms of pFLIP. Digestion of pFLIP(X) with *Xho*I generates the 6.1 kb and 1.6 kb fragments. Digestion of pFLIP(I) with *Xho*I generates the 4.8 kb and 2.9 kb fragments.

sites, carried on a plasmid such as pFLIP. Such an assay in which both sites are on the same DNA molecule is known as a *cis* recombination assay.

pFLIP is a 7.7 kb derivative of pUHE24-2 (Lanzer & Bujard, 1988) carrying 307 bp *att*P and 333 *att*B fragments in inverted orientation with respect to each other. *Xho*I digestion of the pFLIP(X) form generates fragments of 6.1 kb and 1.6 kb, whereas in its integrated form, pFLIP(I), fragments 4.8 kb and 2.9 kb are generated (Figure 2.1). The plasmid also contains the 186 *int* gene (a 1.77 kb *Sal*I-*Bgl*I fragment) under an IPTG inducible promoter-operator sequence.

In this chapter I have used pFLIP to study *in vitro* the requirements for 186 integrative site-specific recombination firstly using cell-free extracts containing a source of 186 Int and IHF proteins, and later using purified components. The results of this work permitted me to confidently conduct recombination experiments in fertilised mouse eggs (Chapter 4) knowing the exact protein requirements and some important characteristics of the recombination reaction.

2.2 Results

2.2.1 *In vitro* assays for integrative recombination: Preparation of cell-free extracts and plasmid substrates

I have examined the ability of cell-free extracts containing 186 Int and IHF to promote integrative recombination *in vitro* using pFLIP(X) as the substrate. Further, several different buffer conditions, in which the salt or pH was varied, were examined for their ability to support or inhibit recombination.

2.2.1.i Preparation of cell-free extracts

Preparation of the cell-free extracts will be described here briefly, however a more detailed description can be found in 6.7.3. *E. coli* C600 (IHF+) and C600Δ*himA* (IHF-) cells were transfected with plasmid pFLIP(X). The cells, containing the

186 *int* gene on pFLIP(X) under the control of an IPTG inducible promoter, were at mid log phase induced with IPTG for 2 hours. The induced cells were harvested, resuspended in cell sonication buffer and subjected to disruption by sonication (3 x 1 min bursts, 5 min rest on ice between bursts). The two lysates, IHF+/Int+ and IHF-/Int+, were cleared by centrifugation (18000 rpm, 30 min, 4°C, SS34 rotor) and aliquots frozen in dry ice and stored at –80°C. Samples of the cell-free extracts were diluted 1/5, 1/10 and 1/25 in TE, and 5 µl aliquots run on a 1% agarose-TAE gel to determine if any pFLIP plasmid DNA was visible, and which may have contaminated the *in vitro* recombination assays performed later. Results indicated no evidence of plasmid contamination in either of the cell-free extracts at all dilutions investigated (data not shown). Similarly a control extract from C600 (IHF+/Int-) was also prepared in the above manner, without IPTG induction.

2.2.1.ii Preparation of pFLIP(X) and pFLIP(I) DNA

pFLIP(I) and pFLIP(X) DNA was prepared from overnight cultures of *E. coli* C600 (IHF+) and C600Δ*himA* (IHF-) cells transfected with plasmid pFLIP(X) respectively. pFLIP(X), containing the 186 *int* gene inducible with IPTG, has a sufficient amount of leaky expression of Int protein to convert 100% of pFLIP(X) *in vivo* to pFLIP(I), in the presence of IHF. High quality DNA was prepared according to 6.6.8.

2.2.2 *In vitro* assays for integrative recombination

Initial investigations demonstrated that a cell-free extract, IHF+/Int+, can mediate site-specific recombination *in vitro* (*6.6.12.i*). Incubation of the substrate pFLIP(X) (300 ng) with an aliquot (2 µl) of the cell-free extract, in a total volume of 50 µl (30 mM Tris-HCl pH7.5, 5 mM NaCl, 90 mM KCl, 2 mM EDTA, 1% glycerol) resulted in approximately 60% conversion of the substrate to the integrative form (Figure 2.2(a); Track A). Moreover, incubation of pFLIP(X) with the cell-free extract IHF-/Int+, showed integrative recombination to be dependent

Figure 2.2(a)

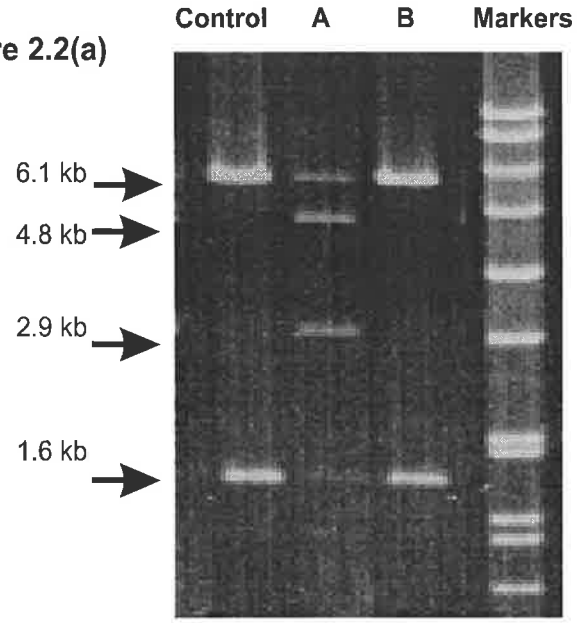


Figure 2.2(b)

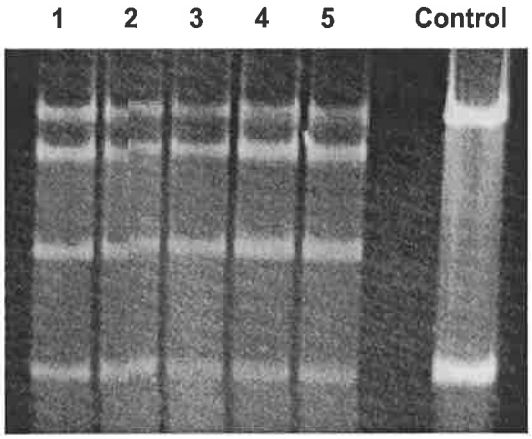


Figure 2.2(a) An ethidium bromide stained 1% agarose gel of the *cis in vitro* recombination assay (*Xho*I digested) using pFLIP(X) as substrate. The control track contains pFLIP(X) and extract from C600 (IHF+/Int-). In track A pFLIP(X) was reacted with cell-free extract containing 186 Int and IHF (from C600:pFLIP(I)) resulting in approximately 60% conversion of pFLIP(X) to pFLIP(I) as judged by the appearance of bands 4.8 kb and 2.9 kb, indicative of integrative recombination. In track B pFLIP(X) was reacted with cell-free extract containing Int only (from C600Δ*him*A:pFLIP(X)).

Figure 2.2(b) An ethidium bromide stained 1% agarose gel of the *cis in vitro* recombination assay (*Xho*I digested) using pFLIP(X) as substrate, cell-free extract (Int+,IHF+) and the following buffer conditions: **1.** 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$. **2.** 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl. **3.** 10 mM Tris-HCl pH 7.5,10 mM $MgCl_2$, 100 mM NaCl. **4.** 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 100 mM KCl 5. 10 mM Tris-HCl pH 8.5, 10 mM $MgCl_2$, 100 mM KCl . The control track contains pFLIP(X)/*Xho*I.

on the presence of IHF (Track B). The control track was pFLIP(X) incubated with IHF+/Int- cell-free extract and indicated integrative recombination to be dependent on Int, as evidenced from the lack of conversion of pFLIP(X) to (I). In track A there is a noticeable decrease in the overall amount of DNA present and it is likely to be due to losses incurred after the recombination reaction, such as during phenol/chloroform extraction and ethanol precipitation, prior to *Xho*I digestion. In the many reactions performed during the course of this project, and not necessarily only those containing Int and IHF CE's, this effect was sometimes observed to occur, in a random fashion. The *in vitro* reaction demonstrated intramolecular recombination between the *att*P and *att*B sites carried on the same plasmid, the *cis* reaction.

Further, by investigating the % conversion from the excised to integrative form of the substrate pFLIP, in the presence of various pH and salt conditions, I found, under the conditions tested, that there was approximately 50-60% substrate conversion from the excised form to the integrative form in 30 minutes at 37°C (Figure2.2(b)). Although we cannot alter the prevailing buffer conditions inside a mammalian cell, it was reassuring to find that the recombination reaction worked under a variety of conditions that varied with respect to the salt and pH conditions. The buffer conditions in lane 4 (10 mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 100 mM KCl, 2 mM EDTA) were chosen as the recombination buffer for all further *in vitro* recombination assays as they most closely satisfied the observations made for optimal λ *in vitro* recombination by Nash and Robertson (1981).

2.2.3 *In vitro* kinetics assay

A *cis* recombination reaction was terminated at various time points ranging from 1 minute to 2 hours to examine the kinetics of the reaction (Figure 2.3). To permit analysis the gel was transferred to a Zeta-probe membrane via the alkali transfer method (6.6.21) and the filter probed with random primed ^{32}P-pFLIP(X) DNA (6.6.20). The extent of recombination in each lane was quantified as the % of pFLIP(I) to the total of pFLIP(I)+pFLIP(X) present using a phosphorimager and

Figure 2.3(a)

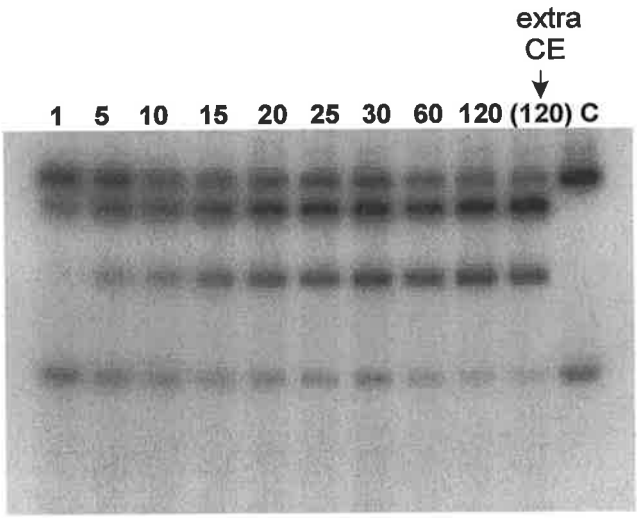


Figure 2.3(b)

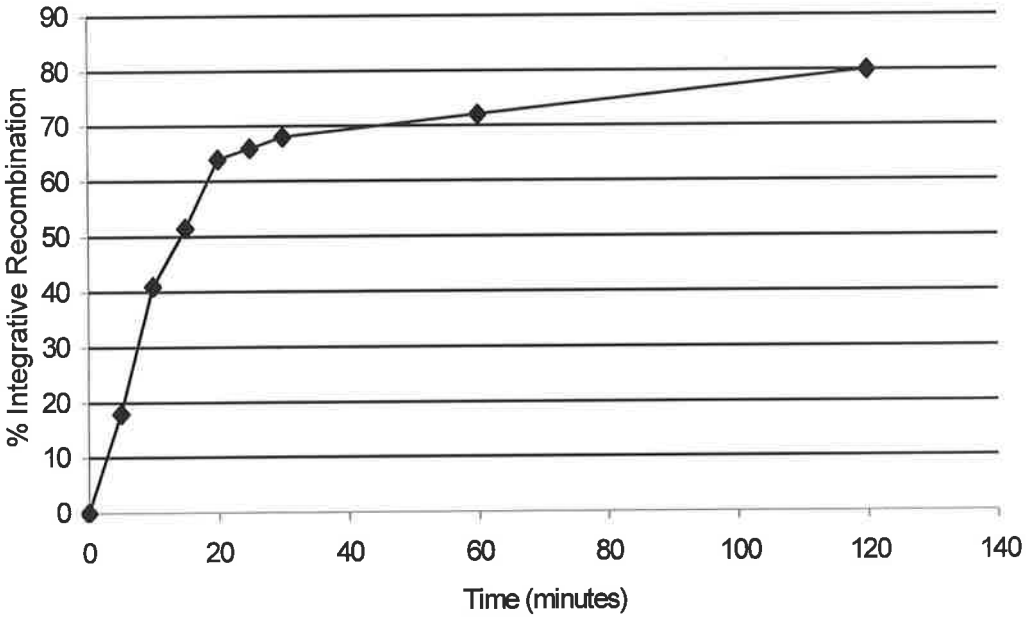


Figure 2.3(a) An *in vitro* kinetics assay of the conversion of pFLIP(X) to pFLIP(I). The recombination reaction was terminated at the time points indicated. In the lane marked (120) the reaction had extra (2x) cell-free extract (Int+, IHF+) added at 5' and the reaction allowed to proceed until termination at 2 hours.

Figure 2.3(b) The phosphorimage above was quantified using ImageQuant software as the % integrative recombination ie. pFLIP(I)/(pFLIP(I) + pFLIP(X)), plotted against time (minutes).

ImageQuant software. The initial peak of recombination (63%) was reached at about 20 minutes after which the recombination rate slowed reaching a level of 80% after 2 hours.

The reaction in the lane marked (120) had, at the 5 minute time point, extra cell-free extract added (2x) and the reaction allowed to proceed until termination at 2 hours. The results, 83% integrative recombination, indicated that the addition of more cell-free extract at 5 minutes did not increase the % conversion significantly above 80% indicating that the levels of Int and IHF proteins were not limiting.. Further to support this, the addition of more substrate pFLIP(X) at 30 minutes also resulted in a % conversion of approximately 80%, again indicating that the Int/IHF activity in the cell-free extract is not exhausted (data not shown). Although not shown, it is unlikely the reaction would reach 100% conversion to the integrative form, if left for a longer period of time, as a nominal percentage of the substrate does not exist in a supercoiled form. It is known for λ and P2 that topologically *att*P needs to be supercoiled for recombination with *att*B (Mizuuchi and Mizuuchi, 1978; Yu and Haggård-Ljungquist,1993). Another feasible reason to account for the lack of 100% conversion is the possible presence of other *E. coli* DNA binding proteins in the cell-free extracts interfering with the binding of IHF and Int to their respective sites.

Theoretically is was possible that the *in vitro* integrative recombination assay using cell-free extracts could be dependent upon host proteins in addition to IHF. To investigate this possibility, *in vitro* site-specific recombination reactions using purified Int and IHF proteins were pursued.

2.3 Construction of an *E. coli* IHF Over-Expressing Strain and Purification of IHF

2.3.1 Introduction

IHF is a small, basic DNA-binding protein consisting of the α-subunit (11.2 kD) encoded by the *himA* gene at 38 min in the *E coli* chromosome (Miller, 1981), and a β-subunit (10.6 kD) encoded by the *him*D gene at 25 min (Flamm and Weisberg, 1985). Intracellular levels of IHF are low, making purification difficult (Nash and Robertson, 1981). Even though the levels are known to increase in stationary-phase cells (Bushman and Thompson, 1986), in order to obtain substantial levels of protein it was necessary to construct an *E. coli* strain over-expressing both subunits simultaneously. Although individually the HimA peptide can be over-expressed and purified (Zulianello *et al.*, 1994), depending on the expression system and *E. coli* strain used, the over-expression of the HimD peptide alone results in the accumulation of insoluble protein. This suggests that the presence of HimA peptide helps to solubilise the HimD peptide (Nash *et al.*, 1987; Zulianello *et al.*, 1994).

2.3.2 Choice of expression system

The goal of this part of the project was to create an IHF over-expressing strain. A number of promoter-plasmid combinations have proved suitable for the expression of recombinant proteins in *E. coli*. The vectors can be of two major classes: 1) the ColE1-type plasmids such as pBR322 and pUC derivatives, which are maintained at high copy number, 2) the R1-type plasmids which are more tightly regulated and exist as 1 or 2 copies per cell. The high-copy number plasmids frequently use the *lac* or *trp-lac* fusion promoters, whereas transcriptional control by λ phage promoters, with heat-sensitive repressors, are often the choice for the low-copy number plasmids.

To over-express IHF I chose the expression vector pEC611 (T. Brumby, 1986; B.Sc Honours Thesis, University of Adelaide, Dept. Biochem.) as it has been used

successfully by this research group to overexpress growth hormone (GH) from various species (human, canine, cameline and equine GH; unpublished data). The expression vector pEC611 contains the strong *trc* promoter (Brosius *et al.*, 1985), a fusion between the promoters from the lactose and tryptophan operons, that is controlled by a repressor protein, the product of the *lac*I gene.

pEC611 is a derivative of pKT52, a high copy-number expression vector based on pBR322 and contains the *trc* promoter with a modified Shine-Dalgarno sequence (AGGAGG), and a *Nde*I site overlapping the ATG start codon, CA$\downarrow$TATG. The *Nde*I site close to the *rep* gene was removed by blunting, which both increased copy number and made the *Nde*I site overlapping the ATG start codon unique (T. Brumby, 1986; B.Sc (Hons) Thesis, University of Adelaide, Dept. Biochem.). The vector has strong transcriptional terminators (T_1T_2) from the *rrnB* ribosomal RNA operon (Brosius *et al.*, 1985) which are necessary to prevent transcription from the *trc* promoter extending around the plasmid. With high copy number plasmids carrying a *trp-lac* fusion promoter, a relatively high number of lac repressor molecules are required to maintain control of expression as low levels may result in constitutive expression of the protein due to inadequate repression. Hence, *E. coli* carrying the pEC611 vector also carry a vector pACYC:*lac*I^q, a derivative of pACYC177, carrying the *lac* repressor gene (*lac*I) expressed from the I^q promoter (Lanzer and Bunjard, 1988), an up-regulated promoter resulting from a single base pair change, resulting in high levels of lac repressor molecules.

2.3.3 Results

2.3.3.i Cloning of the himA and himD genes

The genes for *him*A and *him*D were PCR amplified using the listed primers below and standard PCR conditions (6.6.11), using *E. coli* genomic DNA as the template. In addition to appropriate restriction sites, a second ribosome binding site (AGGAGG), incorporated into one of the PCR primers (HimD-1), was included between the genes to enable translation of a dicistronic transcript.

Oligonucleotides used to clone IHF

Primer	Sequence (5'-3')
HimA-1	TCAC**CATatg**GCGCTTACAAAAGCTGAAATGTCA
HimA-2	ACTC**CTGCAG**ttaCTCGTCTTTGGGCGAAGCGTT
HimD-1	TCAC**CTGCAG (AGGAGG)** TAA**CATatg**ACCAAGTCAGAATTGATAGAA
HimD-2	CTAACTC**AAGCTT**ttaACCGTAAATATTGCGATCGCGCAGTTCTTT

The restriction sites *Nde*1 (with overlapping start site), *Hind*III and *Pst*I, are shown (underlined). The second ribosome binding site is bracketed while the two start sites (ATG) and termination sites (ATT) are in lowercase.

The resultant *him*A and *him*D DNA fragments were digested (6.6.2) with *Nde*I/*Pst*I and *Pst*I/*Hind*III respectively. The vector pEC611 was digested with *Nde*I and *Hind*III and treated with calf intestinal phosphatase (*6.6.4.i*). Following this the purified vector, *him*A and *him*D gene fragments (6.6.3) were then 3-way ligated (6.6.5) and the reaction purified (6.6.6). An aliquot (1/20) of the mix was electroporated (6.5.4) into competent MM294 (pACYC:*lac*I^q) cells (6.5.3) and selected on L+Amp (100ug/ml) + Kn (30ug/ml). After overnight incubation several colonies were picked and inoculated into liquid media and after overnight growth small scale isolation of plasmid DNA was performed (6.6.7). The DNA was analysed by restriction endonuclease digestion with *Hind*III and clones containing both genes identified. Further confirmation was made by dideoxy sequencing (*6.6.9.i*) across both subunits, confirming the correct sequence of both subunits and the second ribosome binding site. The IHF expressing plasmid has been designated pIHF (Figure 2.4).

2.3.3.ii Small-scale expression analysis of pIHF

Cells (MM294:pACYC(*lac*I^q):pIHF) were grown to A_{600} = 0.6, induced with IPTG (0.5 mM) and grown for a further 2 hours. SDS-PAGE analysis *(6.7.2.i)* of extracts with and without IPTG showed high levels of two peptides running close together at approximately 11 Kd and 10 Kd in the induced sample, while the uninduced sample showed low levels of the two peptides indicating some

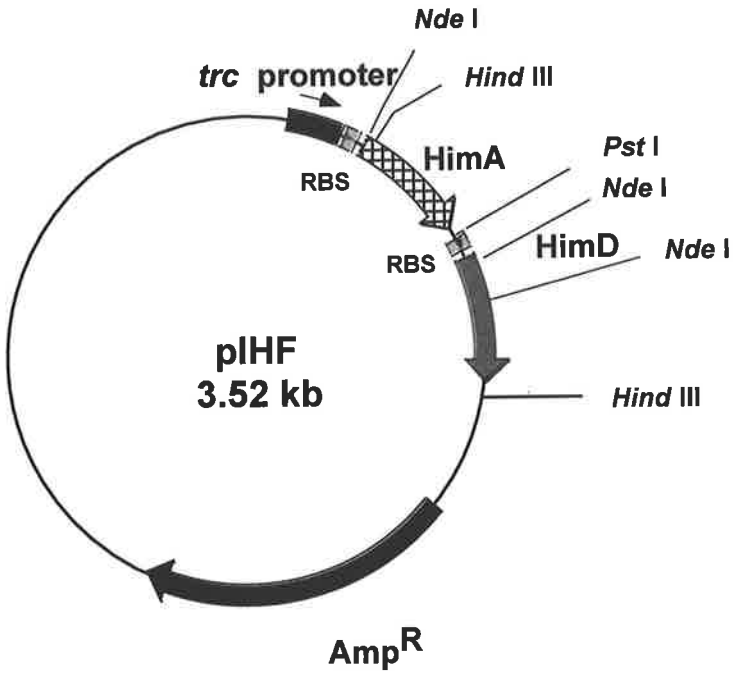


Figure 2.4 Diagrammatic representation of the IHF over-expressing plasmid pIHF. The *him*A and *him*D genes were PCR amplified (6.6.11) from *E. coli* genomic DNA, purified and digested with *Nde*I, *Pst*I and *Pst*I, *Hind*III respectively, and cloned in tandem into pEC611 vector (*Nde*I, *Hind*III). A second ribosome binding site was included between the genes to permit translation of a dicistronic transcript.

leakiness in control of the *trc* promoter (Figure 2.5(a)). The size of the peptides corresponds to the published molecular weights for HimA and HimD, and they seemed to be present in stoichiometric amounts. Hence the pIHF construct appeared to be synthesising abundant IHF protein and the second ribosome binding site was indeed functional, permitting translation of the dicistronic transcript.

2.3.3.iii Activity analysis of cell-free extracts of pIHF

To assess the activity levels of recombinant IHF, cell-free extracts (CE) were made from IPTG induced MM294:pACYC(*lac*I^q):pIHF and also MM294, so as to provide a source of wild-type levels of IHF for comparison. To maintain constant levels of Int in each assay, CE (IHF-) from cells containing pFLIP(X) was used as the source of Int. The pIHF and MM294 CE's were serially diluted from 1/10 to 1/5000 in TEG buffer containing 200µg/ml BSA and 1 ul assayed in the *cis* recombination reaction mix containing purified pFLIP(X) as substrate and Int from the pFLIP(X) CE. After 30 min incubation at 37°C the samples were purified by phenol:chloroform extraction and ethanol precipitation. Samples were *Xho*I digested and run on a 1% agarose gel which was transferred to Zeta-probe membrane via the alkali transfer method and the filter probed with random primed ^{32}P-pFLIP(X) DNA (6.6.22). An autoradiograph of the results is presented in Figure 2.5(b).

Results indicated that the levels of integrative recombination activity of pIHF at 1/500 dilution approximated that of the endogenous levels in MM294 at 1/10 dilution, and pIHF at 1/1000 dilution approximated the activity levels in MM294 at 1/20. In lane A the control pFLIP(X) CE (Int+/IHF-) shows no conversion of the excised form of substrate to the integrative form. In lane B CE from pFLIP(I) (Int+/IHF+) shows conversion of the substrate to the integrative form as evidenced by the presence of the DNA restriction fragments 4.8 and 2.9 kb, which align with the integrative bands (if present) in other lanes. Alone MM294 CE (Int-/IHF+) does not permit conversion of the substrate pFLIP(X) to pFLIP(I) as expected, and

Figure 2.5(a)

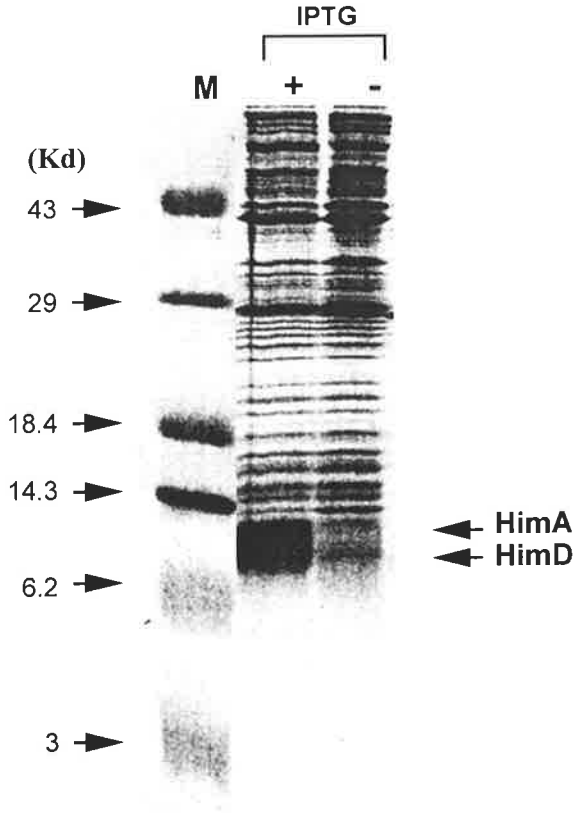


Figure 2.5(b)

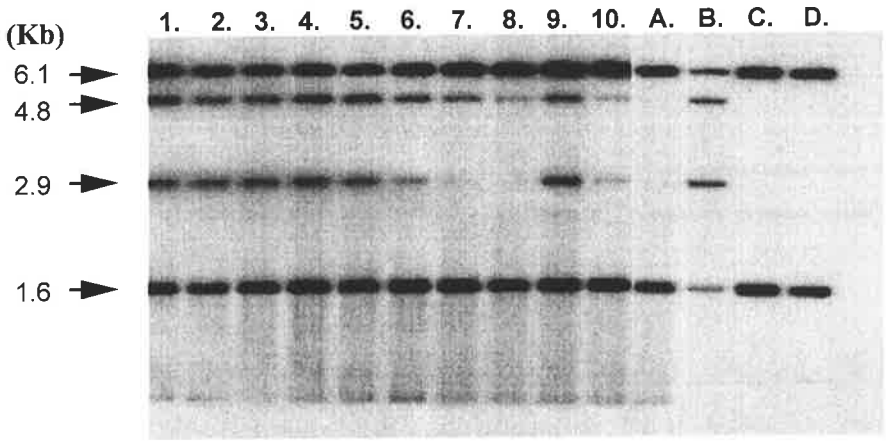


Figure 2.5(a) 15% SDS-PAGE analysis of pIHF expression in the presence and absence of IPTG. Cells from 1 ml of (+/-) IPTG cultures were resuspended in 50 ul of water and 50 ul 2x SDS loading buffer. Each track contains 20 ul of heat denatured sample.

Figure 2.5(b) An autoradiograph of *cis* recombination assays using pFLIP(X) as substrate. pIHF CE was diluted (Tracks 1-8) 1/10, 1/20, 1/100, 1/250, 1/500, 1/750, 1/1000 and 1/2000. MM294 CE (+Int CE) (wild-type levels of IHF) was diluted (Tracks 9 and 10) 1/10 and 1/20. The controls are: A. pFLIP(X) CE (Int+/IHF-). B. p FLIP(I) CE (Int+/IHF+). C. MM294 (Int-/IHF+). D. No CE (Int-/IHF-).

similarly the control in lane D, no CE (Int-/IHF-) also fails to exhibit integrative recombination activity indicating the starting substrate pFLIP(X) was indeed 100% the excised form. Therefore the recombinant IHF levels of my clone are in the order of 50 fold higher than the normal endogenous level in *E. coli,* indicating that this clone will be suitable for large scale purification of IHF.

2.3.3.iv Purification of IHF

Purification of IHF was based on the method of Filutowitz *et al.*, (1994) with modification:

Cell extract:- *E. coli* strain MM294 containing pIHF and pACYC:*lacI*^{q} was grown in 2 litres of LB medium supplemented with 100μg/ml ampicillin and 30μg/ml kanamycin at 37°C. At A_{600} = 0.6, IPTG was added to 0.5 mM and incubation continued for a further 2 hours. All operations from this point were conducted at 4°C. Cells were harvested by centrifugation at 6000 rpm for 20 min. and resuspended in 80 ml of Buffer A (25 mM Tris-HCl pH7.4, 1 mM EDTA, 3 mM DTT, 50 mM NaCl, 5% glycerol) containing 20μg/ml PMSF and 23μg/ml DNaseI. The wet weight of cells was approximately 12 g. The cells were passed once through a French pressure cell at 1500 psi and the cell debris was removed by centrifugation at 17000 rpm for 1hour. The supernatant was ultra centrifuged in a Ti70 rotor using a Beckman J2-21 centrifuge for 2.5 hours at 42000 rpm.

Ammonium sulphate precipitation:- The supernatant (85 ml) underwent a two step ammonium sulphate precipitation. Initially ammonium sulphate was added at 0.334 g per ml of supernatant and slowly stirred for 2 hours, following which the mix was centrifuged at 17000 rpm for 20 min. To the supernatant was added a further 0.23 g of ammonium sulphate per ml. The mixture was allowed to stir slowly for 1 hour after which it was centrifuged at 17000 rpm for 20 min. The pellet containing IHF was resuspended in 5 ml of Buffer A and dialysed overnight with gentle agitation against 4 litres of Buffer A. The dialysate was clarified by centrifugation at 10000 rpm for 10 min and the supernatant filtered through a 0.2 μm Millipore filter.

Heparin-sepharose chromatography:- A 0.9 x 10 cm Amicon column of heparin-sepharose CL6B (Pharmacia) was prepared according to the manufacturer's protocol and equilibrated in Buffer A. Using a Pharmacia FPLC the clarified extract containing IHF was loaded onto the column at a flowrate of 0.5 ml/min using a 5 ml bypass loop. The column was then washed with Buffer A for 2 hours using the same flowrate. A 120 ml salt gradient of 50 mM NaCl to 1.65 M NaCl in Buffer A was applied to the column. Twenty four 5 ml fractions were collected.

2.3.3.v Phastgel and activity analysis of fractions

Phastgel (8-25%) gradient gels with SDS buffer strips (*6.7.2.ii*) were used to analyse eluted fractions from the heparin-sepharose column. It is important to note at this point that the two subunits of IHF, α (11.6 kD) and β (10.2 kD), are not resolved into two distinct bands using the Phastgel system employed for this section of work. Results indicated presumptive IHF protein to be present in fractions 13, 14, 15 and 16 (Figure 2.6(a)), although fractions 13 and 14 appeared to contain some higher molecular weight contaminants. Fractions were assayed for integrative recombination ability using pFLIP(X) as substrate and CE (Int+/IHF-) prepared from cells carrying pFLIP(X). Samples of the *Xho*I digests were run on a 1% agarose gel which indicated that fractions 14, 15 and 16 contained IHF activity (Figure 2.6(b)). Fractions 15 and 16 were each dialysed against 4 litres of Buffer A overnight at 4°C. The next day 500µl aliquots were removed, and to the remainder, an equal volume of glycerol was added and all the fractions (+/- glycerol) stored at –80°C.

2.3.3.vi Purity and concentration

Purified human growth hormone (hGH) at 250, 100 and 50 ng per track was used as a standard on a 8-25 Phastgel (Figure 2.7(a)) to measure the concentration of IHF in fractions 15 (data not shown) and 16 (both no glycerol). For analysis 10 µl of the two fractions were mixed with 10 µl loading buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA, 3.3 % SDS, 10 mM DTT, 0.01 % Bromophenol blue) and the

Figure 2.6 (a)

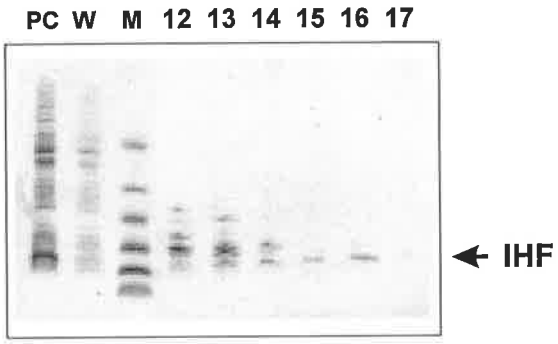


Figure 2.6 (b)

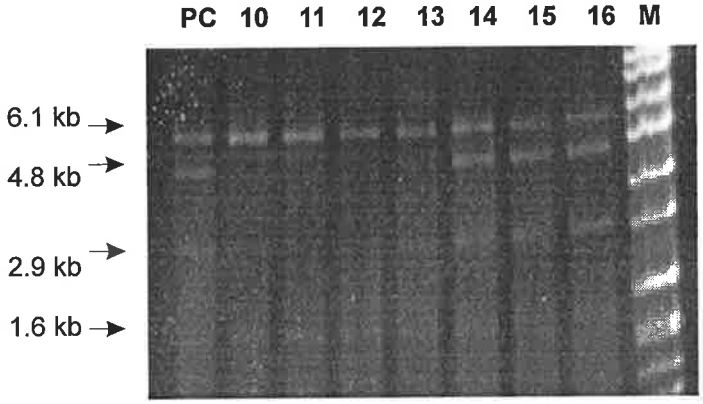


Figure 2.6 (a) 8-25 Phastgel of IHF elution off the Heparin–Sepharose column. PC refers to the pre-column sample, W the wash fraction and 12-17, fractions eluting off the column when the salt gradient was applied. The marker sizes are 43 Kd, 29 Kd, 18.4 Kd, 14.3 Kd, 6.2 Kd and 3 Kd.

Figure 2.6 (b) 1% agarose gel of the *cis* recombination assay showing IHF activity in fractions PC, 14, 15 and 16.

Figure 2.7(a)

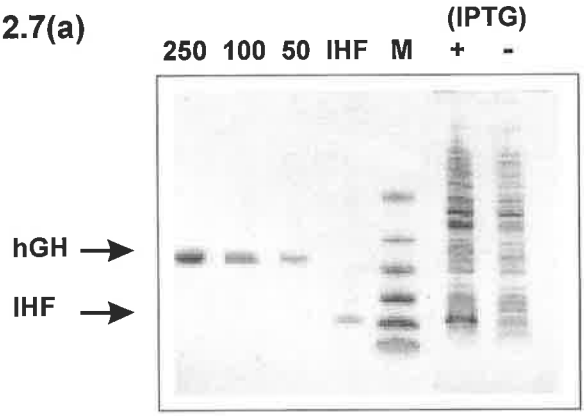


Figure 2.7(b)

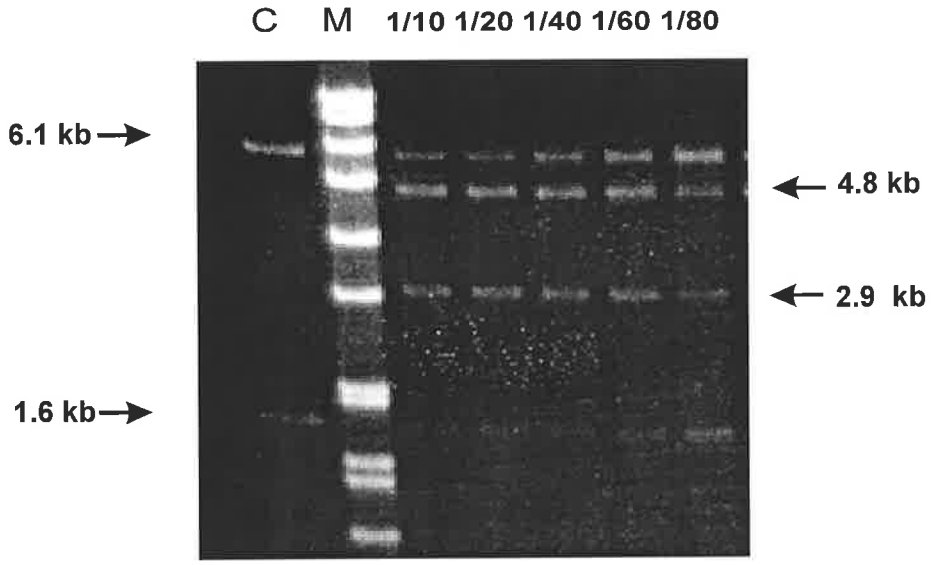


Figure 2.7 (a) 8-25 Phastgel of purified IHF run against 250 ng, 100 ng and 50 ng pure hGH. The dried gel was scanned and using ImageQuant software the calculated concentration of IHF was 40 ng/µl. The marker sizes are 43 Kd, 29 Kd, 18.4 Kd, 14.3 Kd, 6.2 Kd and 3 Kd. Aliquots of cells +/- IPTG are shown to indicate the relative position of induced IHF.

Figure 2.7 (b) 1% agarose gel of the *cis* recombination assay using the purified IHF which was serially diluted from 1/10 to 1/80. A conservative estimate of the unit activity of IHF is 40 U/µl.

samples heated to 100°C for 5 min and microfuged for 10 seconds to bring down condensation formed on the lid of the eppendorf tubes. Then 1 ul samples were run on a Phastgel (8-25) gradient gel and a Molecular Dynamics Personal Densitometer was used to scan the gel, and finally analysis of concentration was performed using ImageQuant software (Table 2.1). The calculated IHF concentration for fractions 15 and 16 was 37 ng/µl and 40 ng/µl respectively. As each track contained a 1:1 ratio of protein to loading buffer, each sample is essentially diluted 50% on the gel, hence the actual concentration would be double in the samples without glycerol but remains at 37ng/µl and 40ng/µl for the bulk of the IHF fractions which contain 50% glycerol. Further, the purity of the fractions, combining the area of the peaks for both HimA and HimD and comparing the % area to that of other minor peaks (combined), was 80% and 86% for fractions 15 (data not shown) and 16 respectively.

Table 2.1

Peak	Area	Percent	Description
1	1.576	100	250ng hGH
2	0.701	100	100ng hGH
3	0.327	100	50ng hGH
5	0.259	100	16 conc.
1	0.113	14.383	minor peaks
2	0.622	85.617	16 purity

2.3.3.vii Activity analysis

Fraction 16, containing IHF at a calculated purity of 86%, was serially diluted with Buffer A and assayed for its integrative recombination ability using pFLIP(X) as substrate. One unit of IHF activity was defined as that amount of IHF required for maximal intramolecular integrative recombination in the presence of non-limiting amounts of 186 Int in 30 min at 37°C using 300 ng of the substrate pFLIP(X). In figure 2.7(b) the highest dilution of IHF to still exhibit maximal conversion of pFLIP(X) to (I) is the 1 in 40 dilution, whereas the 1 in 60 and 1 in 80 dilutions show a decrease in activity with higher levels of the 6.1 kb and 1.6 kb bands, representing non-conversion of pFLIP(X) to pFLIP(I), and a corresponding

decrease in the 4.8 kb and 2.9 kb bands. The IHF activity in fraction 16 was conservatively estimated to be at least 40 U/µl.

2.4 Purification of 186 Int From an Overproducing Strain

2.4.1 The expression system

The product of the 186 *int* gene is a member of the integrase family of site-specific recombinases (Argos *et al.*, 1986) required for both integration and excision. An Int expression plasmid (pMRR18) was constructed (M. Reed, 1994) by inserting a *Nde*I-*Bam*HI fragment of a 186 Int PCR product into the *Nde*I and *Bam*HI sites of the expression plasmid pET3a. The expression vector is harboured in *E. coli* BL21(DE3) (Studier and Moffat, 1986) which has the phage T7 RNA polymerase gene controlled by the *lac*UV5 promoter incorporated into the host chromosome. IPTG induction of T7 RNA polymerase subsequently controls translation from the strong Ø10 T7 RNA promoter and s10 translation site (Rosenberg *et al.*, 1987) of pET3a.

2.4.2 Results

2.4.2.i Small-scale expression analysis

Cells (BL21(DE3)pMRR18) were grown to A_{600} = 0.6 and induced with IPTG to a final concentration of 0.5 mM and grown for a further 2 hours. SDS-PAGE analysis (Figure 2.8(a)) of extracts, with and without induction, showed high levels of a protein running at approximately 39 kD in the induced sample. The uninduced sample showed low levels of the protein probably due to a small amount of basal expression of T7 RNA polymerase from the *lac*UV5 promoter.

Figure 2.8(a)

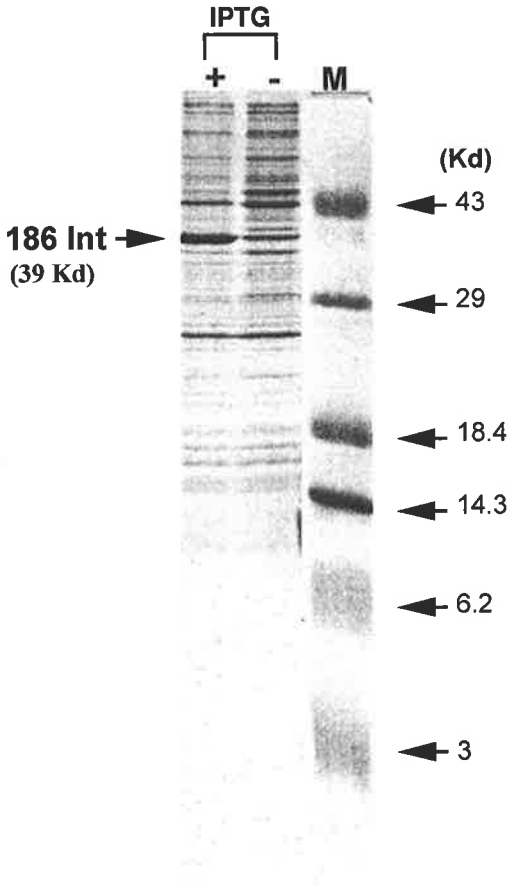


Figure 2.8 (b)

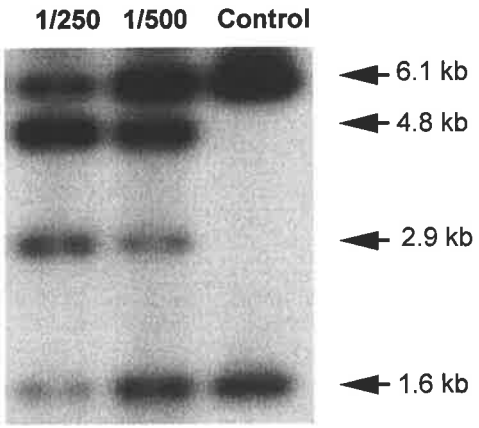


Figure 2.8(a) 15% SDS-PAGE analysis of MRR18 expression in the presence and absence of IPTG. Cells from 1 ml of (+/-) IPTG cultures were resuspended in 50 ul water and 50 ul 2x SDS loading buffer. Each track contains 20 ul of heat denatured sample.

Figure 2.8(b) An autoradiograph of *cis* recombination assays using pFLIP(X) as substrate. MRR18 CE was diluted 1/250 and 1/500. The control was DH5 α CE.

2.4.2.ii Activity analysis of cell-free extracts of MRR18

Cell-free extracts were made from IPTG-induced BL21(DE3)pMRR18, and also from DH5α in order to maintain a constant level of IHF in each *cis* recombination assay using pFLIP(X) as substrate. The CE from MRR18 was diluted 1/250 and 1/500 in TEG buffer containing 200 μg/ml BSA and 1μl assayed in the recombination reaction mix containing purified pFLIP(X) as substrate and IHF from the DH5α CE (2 μl). After 30 min incubation at 37°C the samples were purified by phenol:chloroform extraction and ethanol precipitation. Samples were *Xho*I digested and run on a 1% agarose gel which was transferred to Zeta-probe membrane via the alkali transfer method and the filter probed with random primed ^{32}P-pFLIP(X) DNA. An autoradiograph of the results is presented in Figure 2.8(b).

Results indicated that the levels of Int activity obtained from the pMRR18 expression vector would be suitable for purification of the protein, as the 1/250 dilution of CE gave approximately 70% conversion of pFLIP(X) to pFLIP(I) while the 1/500 dilution was still able to convert 40% of the substrate to the integrated form. The control lane contained DH5α CE (Int-/IHF+) and showed no conversion of the excisive form of the substrate to the integrative form. Hence, the results suggested high levels of soluble and active 186 Int protein could be obtained from pMRR18.

2.4.2.iii Purification of 186 Int

Heparin-sepharose is ideal for purifying DNA binding proteins and due to the success of using heparin sepharose chromatography for purifying IHF, I designed a similar protocol for Int. The main alterations are the use of a different buffer system and salt gradient for elution of protein from the column and the absence of a ultra-centrifugation step. The salt gradient was changed from a NaCl gradient, used in the IHF purification, to a KCl gradient (100 mM to 1.8 M KCl) as a trial preparation indicated Int absorbed to the heparin-sepharose strongly (data not

shown), necessitating the use of a stronger ionic salt for elution to avoid using even higher concentrations of NaCl.

Cell extract:- *E. coli* strain BL21(DE3) containing pMRR18 was grown in 1 litre of LB medium supplemented with 100μg/ml ampicillin at 37°C. At A_{600} = 0.6, IPTG was added to 0.5 mM and incubation continued for a further 2 hours. All subsequent operations were conducted at 4°C. Cells were harvested by centrifugation at 6000 rpm for 20 min and resuspended in 15 ml of Buffer AI (25 mM Potassium phosphate pH7.5, 1 mM EDTA, 3 mM DTT, 100 mM KCl, 5% glycerol) containing 20 μg/ml PMSF. The wet weight of cells was approximately 4.5 g. The cells were passed once through a French pressure cell at 1500 psi. The cell debris was removed by centrifugation at 17000 rpm for 30 min.

Ammonium sulphate precipitation:- The supernatant (17 ml) was subjected to an ammonium sulphate cut. Ammonium sulphate was added to 40% (22.9 g/100 ml) slowly stirred for 1 hour, and the mix then centrifuged at 12000 rpm for 20 min. To the supernatant (16 ml) was added further ammonium sulphate to 75% (22.5 g/100 ml). The mixture was stirred slowly for 2 hours and then centrifuged at 15000 rpm for 20 min. The pellet containing Int was resuspended in 2 ml of Buffer AI and dialysed overnight with slow stirring against 4 litres of Buffer AI. The dialysate was clarified by centrifugation at 10000 rpm for 10 min, following which the supernatant was filtered through a 0.2 μm Millipore filter.

Heparin-sepharose chromatography:- A 0.9 x 10 cm Amicon column of heparin-sepharose CL6B (Pharmacia) was prepared according to the manufacturer's protocol and equilibrated in Buffer AI. Using a Pharmacia FPLC the clarified extract containing Int was loaded onto the column at a flowrate of 0.5 ml/min using a 2 ml bypass loop. The column was then washed with Buffer AI for 2 hours using the same flowrate. A 100 ml salt gradient of 100 mM KCl to 1.8 M KCl in Buffer AI was applied to the column. Twenty 5 ml fractions were collected.

2.4.2.iv Phastgel and activity analysis of fractions

8-25 Phastgel analysis of fractions indicated presumptive 186 Int protein to be present in fractions 8, 9 and 10, with the majority of protein present in fraction 9 (Figure 2.9(a)) and, more importantly, without visible levels of contaminants. Fraction 9 was assayed for integrative recombination ability using pFLIP(X) as substrate and CE prepared from DH5α cells (IHF+). Samples of the *Xho*I digests were run on a 1% agarose gel, which indicated fraction 9 to contain high levels of Int activity (Figure 2.9(b)). Fraction 9 was dialysed against 4 litres of Buffer AI overnight at 4°C. The next day a 500ul aliquot was removed and to the remainder, an equal volume of glycerol was added, and the stock stored at –80°C.

2.4.2.v Purity and concentration

Purified human growth hormone (hGH) at 250, 100 and 50 ng per track was used as a standard on a 8-25 Phastgel (Figure 2.10(a)) to measure the concentration of Int in fraction 9 (no glycerol). For analysis 10 µl of the fraction was mixed with 10 µl loading buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA, 3.3 % SDS, 10 mM DTT, 0.01 % Bromophenol blue) and the samples heated to 100°C for 5 min and microfuged for 10 seconds to bring down condensation formed on the lid of the eppendorf tubes. Then a 1 µl sample was run on a Phastgel (8-25) gradient gel and a Molecular Dynamics Personal Densitometer was used to scan the gel, and finally analysis of concentration was performed using ImageQuant software (Table 2.2).

Table 2.2

Peak	Area	Percent	Description
1	1.386	100	250ng hGH
2	0.589	100	100ng hGH
3	0.258	100	50ng hGH
4	1.592	100	9 conc.
1	0.0326	4.346	minor peaks
2	0.699	95.654	9 purity

Figure 2.9 (a)

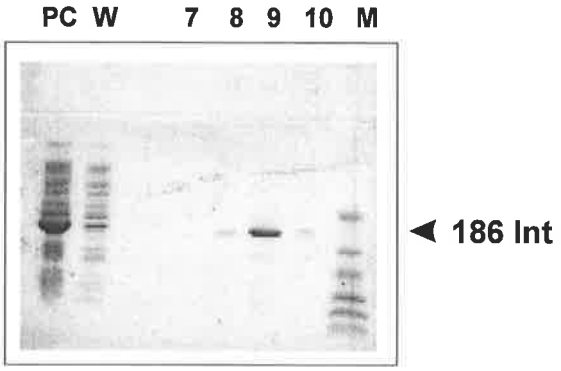


Figure 2.9(b)

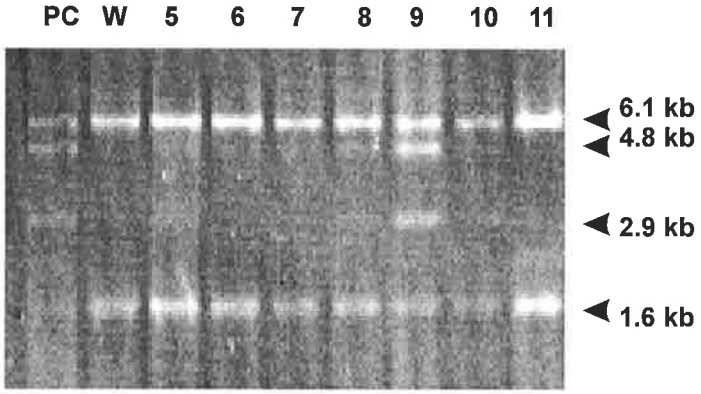


Figure 2.9(a) 8-25 Phastgel of 186 Int elution off the Heparin-Sepharose column. PC refers to the pre-column sample, W the wash fraction and 7-10, fractions eluting off the column when the salt gradient was applied.

Figure 2.9(b) 1% agarose gel of the *cis* recombination assay showing 186 Int activity predominantly in fraction 9.

Figure 2.10 (a)

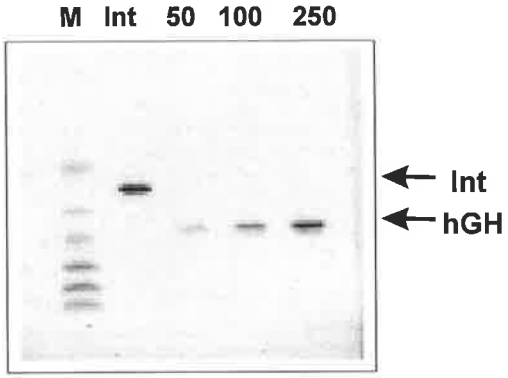


Figure 2.10 (b)

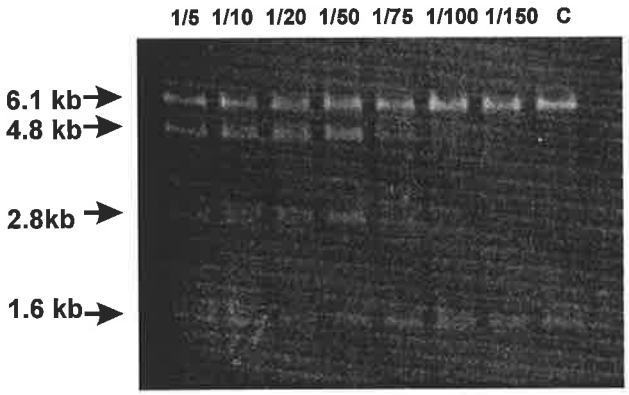


Figure 2.10(a) 8-25 Phastgel of purified Int run against 250ng/µl, 100ng/µl and 50ng/µl pure hGH. The dried gel was scanned and using ImageQuant software the calculated purity of Int was 288ng/µl. The marker sizes are 43 Kd, 29 Kd, 18.4 Kd, 14.3 Kd, 6.2 Kd and 3 Kd.

Figure 2.10(b) 1% agarose gel of the cis recombination assay using the purified Int which was serially diluted from 1/5 to 1/150. A conservative estimate of the unit activity of 186 Int is 50U/µl.

The calculated Int concentration for fraction 9 was 288 ng/µl. Since the track contained a 1:1 ratio of protein to loading buffer, the sample is in essence diluted 50% on the gel, hence the actual concentration would be double in the sample without glycerol but remains at 288 ng/µl for the bulk of the purified Int protein which contains 50% glycerol. Further, the purity of the fraction, taking the area of the peak for Int and comparing the % area to that of other minor peaks (combined), was 96% for fraction 9.

2.4.2.vi Activity analysis

Fraction 9 containing 186 Int, at a calculated purity of 96%, was serially diluted and assayed for its integrative recombination ability using pFLIP(X) as substrate. One unit of Int activity was defined as that amount of Int required for maximal intramolecular integrative recombination in the presence of non-limiting amounts of *E. coli* IHF in 30 min at 37°C using 300 ng of the substrate pFLIP(X). In figure 2.10(b) maximal activity is maintained up to the 1 in 50 dilution, whereas the 1 in 75 dilution shows a decrease in activity with higher levels of the 6.1 kb and 1.6 kb bands representing non-conversion of pFLIP(X) to pFLIP(I), and a corresponding decrease in the 4.8 kb and 2.9 kb bands. The Int activity in fraction 9 was conservatively estimated to be at least 50 U/µl.

2.5 *In Vitro* '*trans*' recombination assays using purified 186 Int and IHF

2.5.1 Introduction

The 186 integrative recombination reaction has so far been studied using the *cis* reaction, where the phage and bacterial attachment sites are located on the same DNA molecule. This section describes a new set of substrates which were utilised for studying the *in vitro* integrative recombination assay of phage 186 in the *trans* mode. For *trans* recombination the *att*P site is contained on a circular (supercoiled) plasmid while the *att*B, used in this section of work, is contained on

a separate small linear molecule, an oligonucleotide (^{32}P-labelled). *Trans* recombination produces a linear molecule, containing both *att*L and *att*R, of a length equal to the sum of the two substrates (Figure 2.11). The *in vitro trans* recombination assays using purified Int and IHF proteins allowed me to determine the requirements for efficient integrative recombination.

2.5.2 The *att*P and *att*B substrates

A plasmid, pPgkP'OPNeo (5.25 kb), was created (Dr. P. Wigley, BresaGen, Adelaide, Australia) containing the 186 *att*P site and the Neomycin resistance gene under the constitutive control of the murine phosphoglycerate kinase promoter and polyadenylation cassette. This would permit me to select for Neo^R mammalian cell lines in future experiments, and thereby determine if the insertion was a random event or a 186 directed event at a specific target site. The 186 *att*P site used was the same 307 bp fragment utilised in the construction of pFLIP substrates (M. Reed, 1994). The *att*B substrate used was a 45 bp linear molecule created by annealing (6.6.18) two oligonucleotides (WT-T and WT-B) representing the top and bottom strands of wild-type *att*B:

WT-T 5'-**GCGACCAAGCGATTATGAGT**CGCCTGCTCTAACCACTGAGCTAAA-3'
WT-B 5'-TTTAGCTCAGTGGTTAGAGCAGGCG**ACTCATAATCGCTTGGTCGC**-3'

The bases in bold represent those falling within the 48/49 bp identity sequence (see 1.8 and Appendix, Figure 1) between *att*P and *att*B. In Chapter 3, I show that the *att*B site lies within these 45 base pairs. Utilising this information allowed me to develop a simple *trans* recombination assay, used in the following section of work, once the purified proteins became available.

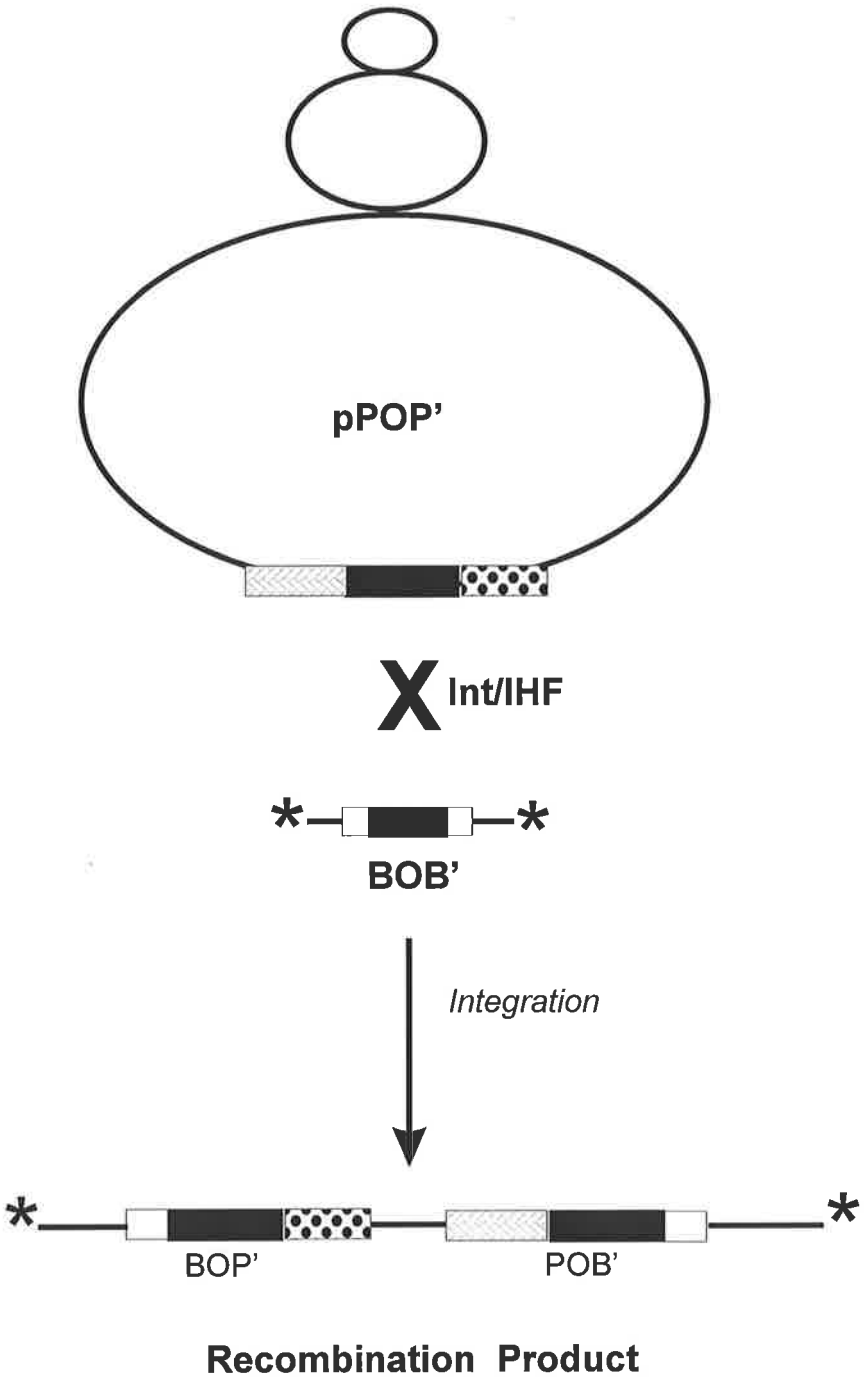


Figure 2.11 Schematic representation of the *trans* recombination assay. Shown is a supercoiled POP'- containing plasmid (pPOP') and a linear ^{32}P-end labelled (*) duplex oligonucleotide containing BOB'. In the presence of Int and IHF integration can occur resulting in a ^{32}P-labelled linear molecule containing BOP' and POB', and of a length equal to the sum of the two substrates. On a 1% agarose/TAE gel the recombination product runs with slower mobility than the linear ^{32}P-end labelled duplex oligonucleotide containing BOB'.

2.5.3 Results

2.5.3.i Units of Int and IHF required for trans recombination

The duplex *att*B linear fragment was kinased with γ-^{32}P-ATP (6.6.19) and reacted with supercoiled pPgkP'OPNeo in recombination buffer in the presence of varying amounts of Int and IHF proteins at 37°C for 60 min (*6.6.12.ii*). The resultant recombination product is a linear molecule of approximately 5.3 kb detected by agarose gel electrophoresis and autoradiography. For *trans* recombination one unit of Int or IHF is defined as the minimal amount of protein required to achieve maximal recombination in the presence of non-limiting amounts of the other protein. One *trans* unit of IHF was approximately quantified as a 1 in 20 dilution of the purified IHF stock, giving an activity of 20 U/µl (Figure 2.12(a)). One *trans* unit of Int was approximately quantified as a 1 in 5 dilution of the purified stock, resulting in an activity of 5 U/µl (Figure 2.12(b)).

The results strongly indicate that integrative recombination exhibits no requirement for additional proteins or factors other than Int and IHF. In the control track of figure 2.12(a) only IHF is present, whereas in the control track of figure 2.12(b) only Int is present, and neither reaction gives rise to the recombination product. Importantly too, these results indicate that an excess of either IHF (100 U) or Int (25 U) protein is not inhibitory to the reaction. In addition to the expected recombination product band, one or two bands with slower mobility are evident. These are likely to be Holliday intermediates, such as the α-structure, which results when only the first strand exchange occurs (1.7.5).

2.5.3.ii Kinetics of integrative recombination

In order to examine the kinetics of integrative recombination, I used 1U (*trans*) of both Int and IHF, and terminated the reaction at various time points from 5 minutes to 75 minutes (Figure 2.13(a) and (b)). At 30 minutes the level of *trans* recombination was approximately 22% and rose to 24% around 45 minutes (where %Recombination=RP/total radioactivity), after which the appearance of

Figure 2.12(a)

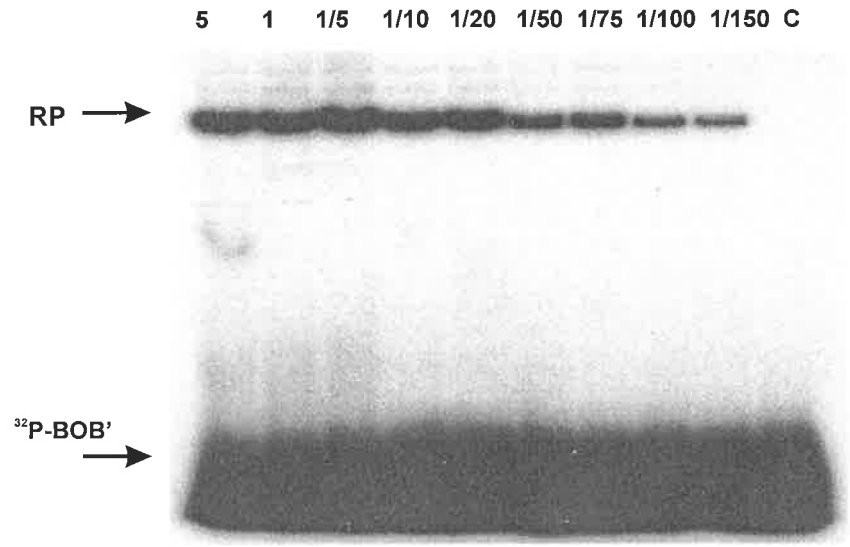


Figure 2.12(b)

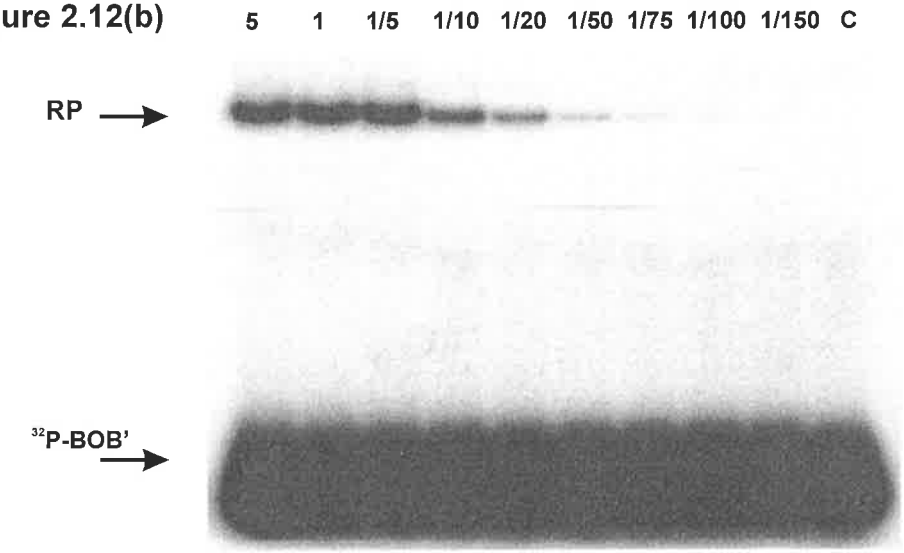


Figure 2.12(a) Autoradiograph of the *trans* recombination assay in which IHF activity was titrated. The positions of the recombination product band (RP) and free ^{32}P-BOB' are indicated. Purified IHF protein was assayed from 5ul down to 1ul of the 1/5 to 1/150 dilutions and 10 ul of assay (50 ul) run on a 1% agarose-TAE gel, transferred to Zeta-probe membrane and autoradiographed. An excess of purified Int (20 U) was used in all assays. The control lane (C) contains Int protein only.

Figure 2.12(b) Autoradiograph of the *trans* recombination assay in which Int activity was titrated. The positions of the recombination product band (RP) and free ^{32}P-BOB' are indicated. Purified Int protein was assayed from 5ul down to 1ul of the 1/5 to 1/150 dilutions and 10 ul of assay (50 ul) run on a 1% agarose-TAE gel, transferred to Zeta-probe membrane and autoradiographed. An excess of purified IHF (80 U) was used in all assays. The control lane (C) contains IHF protein only.

Figure 2.13(a)

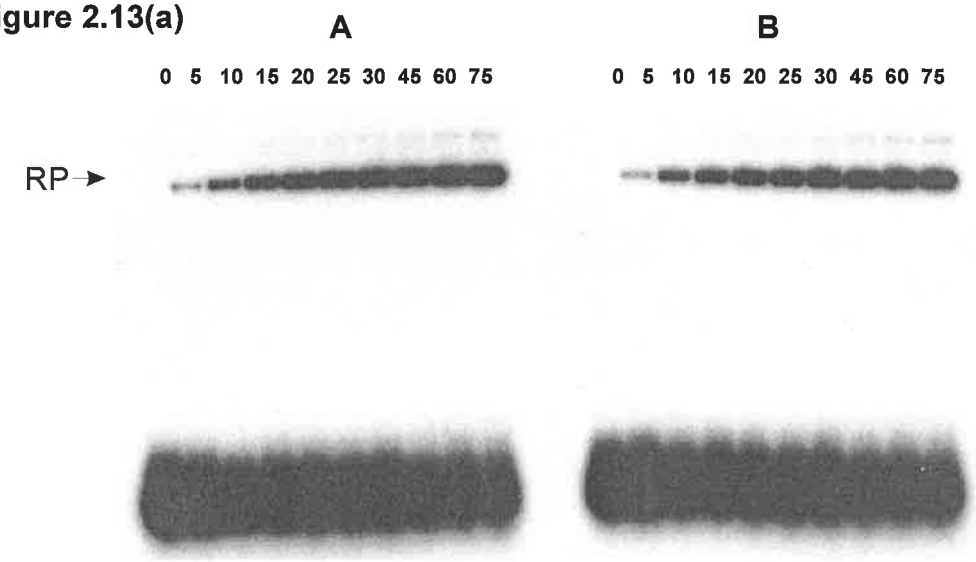


Figure 2.13(b)

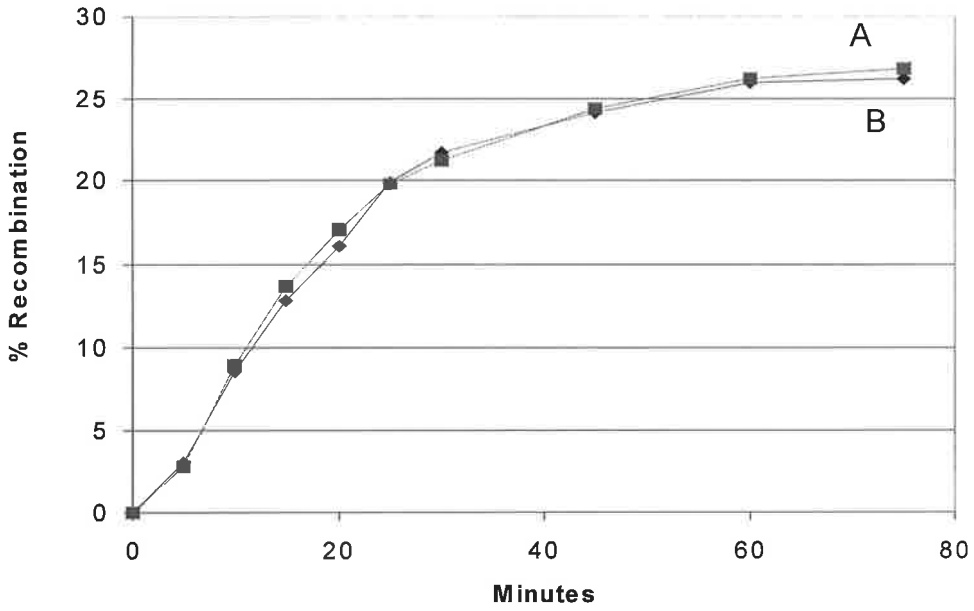


Figure 2.13(a) A kinetics assay of trans recombination using 1U of Int and 1U of IHF proteins. The reaction was performed in duplicate (A and B) and aliquots of the assay terminated at the indicated time points. The autoradiograph was developed after 18 hr exposure at room temperature. RP is the recombination product band.

Figure 2.13(b) Using a phosphorimager, a shorter exposure (5 hr) of the above results was analysed using ImageQuant software, and the results depicted as a graph of the % recombination (RP/total radioactivity) against time (minutes).

recombination products did not significantly increase beyond 25-26%. Using approximately 2-5ng (0.07-0.17 pmoles) of end-labelled *att*B and 300 ng (0.09 pmoles) of pPgkP'OPNeo I could achieve a level of *trans* recombination of around 25% routinely.

2.6 Discussion

The 186 site-specific recombination system is analogous to the λ and P2 systems and is involved in the switch between lytic and lysogenic lifecycles. My aim is to exploit the 186 recombinase system for the stable insertion of DNA at specific locations in the genomes of higher eukaryotes. The major advantage of the recombination systems like λ and P2 is their directionality. This directionality is due to different combinations of proteins mediating integrative and excisive recombination. Reed (1994) demonstrated this directionality for the 186 system using the pFLIP *cis* recombination assay *in vivo*. Transformation of the plasmid in the excised form into an *ihf*+ *E. coli* cell, carrying a source of Int, led to the integrated form of the plasmid, while the reverse reaction was dependent also on a source of Apl (the excisionase), but independent of IHF. However, either of these two reactions may have been dependent upon other host factors. In order to investigate this possibility it was necessary to conduct *in vitro* experiments using purified proteins.

A good starting point for the project was to assess whether I could use *in vitro* the same *cis* recombination assay, using the pFLIP(X) substrate and cell-free extracts containing Int and IHF proteins. Indeed this was the case, and the assay system permitted assessment of some requirements for efficient recombination *in vitro*. Results indicated that different buffers, in which the salt and pH conditions were varied, could all support efficient *cis* recombination. I did not assess the difference in the rates of the reactions only the final extent of recombination. Nevertheless, the results showed 186 recombination to operate in various salt and pH conditions and in the absence of spermidine. Spermidine was not used in the

assays as it has been in others (Kikuchi and Nash, 1978, Yu and Haggård-Ljungquist, 1993a) since it has been shown not to be an absolute requirement but as an enhancer of recombination (Nash and Robertson, 1981), and I wished to assess the efficiency of reactions *in vitro* under conditions more likely to exist in eukaryotic cells. There was no significant difference in the final extent of recombination achieved in all five buffers used. Even in the simplest of buffers (10 mM Tris-HCl pH7.5, 10 mM $MgCl_2$), approximately 50% recombination could be achieved in the absence of an accompanying salt. In the *cis* assay, using cell-free extracts, we could routinely achieve integrative recombination rates of 50 to 80%, over a 30-60 minute incubation time, a result similar to that obtained by Mizuuchi and Nash (1976).

The establishment of a *cis* recombination assay *in vitro* permitted me to assess whether the cloned proteins were functional before purification and remained so after the purification procedure. Elution of IHF, off the heparin-Sepharose column, occurred in the range of 0.95 M to 1.08 M NaCl, in agreement with the results obtained by Filutowitz (1994). The calculated purity was 86% with a specific activity of 1000 U/μg of protein.. The 186 Int protein eluted off the heparin-Sepharose column in the range of 0.76 M to 0.95 M KCl. The calculated purity was 96%, with a specific activity of 172 U/μg of protein. The amount of protein required to convert 300 ng of pFLIP(X) to at least 50–60% pFLIP(I), in 30 minutes at 37°C, was 1 ng of IHF and 5.76 ng of Int. Both IHF and 186 Int, at 86% and 96% purity respectively, were deemed suitable for all further investigations involved in this project.

The *trans* recombination assays, demonstrated here to work with reasonable efficiency *in vitro*, permitted me to evaluate the potential of the 186 site-specific recombination system to promote recombination in the cellular environment of a higher eukaryote (Chapter 3). In the *trans* recombination assays, using 1U of Int and 1U of IHF, I could achieve levels of integrative recombination of approximately 25% after 45 minutes, a result akin to that obtained by Yu *et al.*, (1993b) for P2. In the P2 results (*att*B x *att*P), using 1U of IHF and 1U of P2 Int,

they obtained levels of *trans* recombination of about 35% at 90 minutes (data for the 45 min time point was not shown). For 186 the minimal amount of Int and IHF protein needed to carry out *trans* recombination to at least 20-25%, using 300 ng of *att*P substrate (pP'OPNeo) and 50 ng of *att*B, in 30 minutes and 37°C, was 2 ng of IHF and 57.6 ng of Int. The levels of IHF required for maximal *cis* and *trans* integrative recombination are similar. In stark contrast, the amounts of Int protein required for *cis* and *trans* recombination were significantly different (Table 2.3). This may reflect the ease with which *cis* recombination can occur by comparison to the *trans* reaction, where the *att*P intasome needs to collide with *att*B molecules that are at a greater distance away in solution, and effectively the *att*B concentration is diluted in the *trans* mode. It has been estimated that intermolecular (*trans*) recombination in the λ system occurs 25-fold less frequently than intramolecular (*cis*) recombination (Nash, 1975).

Table 2.3

Protein Requirements	***cis* assay**	***trans* assay**
Int	50 U/μl (or 5.76 ng)	5 U/μl (or 57.6 ng)
IHF	40 U/ μl (or 1 ng)	20 U/μl (or 2 ng)

186 integrase absolutely requires IHF for the integration reaction, as judged in both the *cis* and *trans in vitro* assay systems. The same is true for the λ (Miller *et al.*, 1979; Nash, 1981) and P2 systems (Yu and Haggård-Ljungquist, 1993a) whereas for HP1, a phage related to P2 and 186 (Esposito *et al.*, 1996), IHF is not essential however its presence stimulates recombination (Goodman and Scocca, 1989; Astumian *et al.*, 1989).

In the P2 system it was found that the integration reaction was extremely sensitive to the levels of P2 Int protein with inhibition dramatic when 2-3 U of Int was used instead of the optimal 1 U (Yu and Haggård-Ljungquist, 1993b). My finding that excess levels of Int and IHF proteins are not inhibitory to the 186 integrative recombination assay will be advantageous to using this system in a higher

eukaryotic cell, as we may not need to fine tune the exact amount of protein to deliver to a cell.

2.7 Conclusions

In order to better understand 186 integrative recombination, two *in vitro* assays, the *cis* and *trans* systems, were developed, and together with the purified proteins, Int and IHF, established that these are the only proteins required for efficient *in vitro* integrative recombination. Further, the elucidation of the levels of Int and IHF proteins needed for efficient *trans in vitro* recombination will permit optimisation of intasome formation prior to delivery into a eukaryotic cell system, and more generally, permit a more detailed analysis of 186 recombination to be pursued. For instance, undertaking characterisation of the phage attachment site, and relevant to this study, characterisation of the bacterial attachment site.

CHAPTER 3

Localisation and characterisation of the 186 *att*B site

Chapter 3

Localisation and characterisation of the 186 *att*B site

3.1 Introduction

In the well studied λ site-specific recombination system the minimal *att*B sequence required is 21 bp, sharing a region of identity to *att*P of 15 bp (Mizuuchi and Mizuuchi, 1985). Recombination is executed by the phage-encoded integrase protein (Int), a type 1 topoisomerase with DNA nicking and rejoining activity. Both processes do not require any high-energy cofactors (Kikuchi and Nash, 1979), instead Int conserves the bond energy in a covalent phosphotyrosine linkage with the 3' end of the cut strand (Craig and Nash, 1983, Pargellis *et al.*, 1988). In λ, the overlap region is comprised of 7 base pairs within the region of identity between *att*P and *att*B. The cleavage, exchange and religation steps occur at the edges of the overlap sequence.

Following assembly of the intasome and the capture of a protein-free *att*B, the reaction begins with the cleavage of one strand of each of *att*B and *att*P (Figure 3.1). The cleavage reaction is performed by the conserved tyrosine residue (λ: Tyr342) which nucleophilically attacks the phosphate backbone resulting in the formation of a covalent bond between the Int protein and the 3' phosphoryl group at the site of strand cleavage, leaving a free 5' hydroxyl group. The cleaved strands from *att*P and *att*B then exchange positions, and the free 5' hydroxyl groups of the broken DNA attack the covalent bond between Int and the 3' phosphoryl groups releasing Int and religating the cleaved strands. The outcome of this sequence of events is the formation of a Holliday junction, which is subsequently resolved when the cleavage, exchange and religation steps are repeated with the remaining two stands.

Figure 3.1

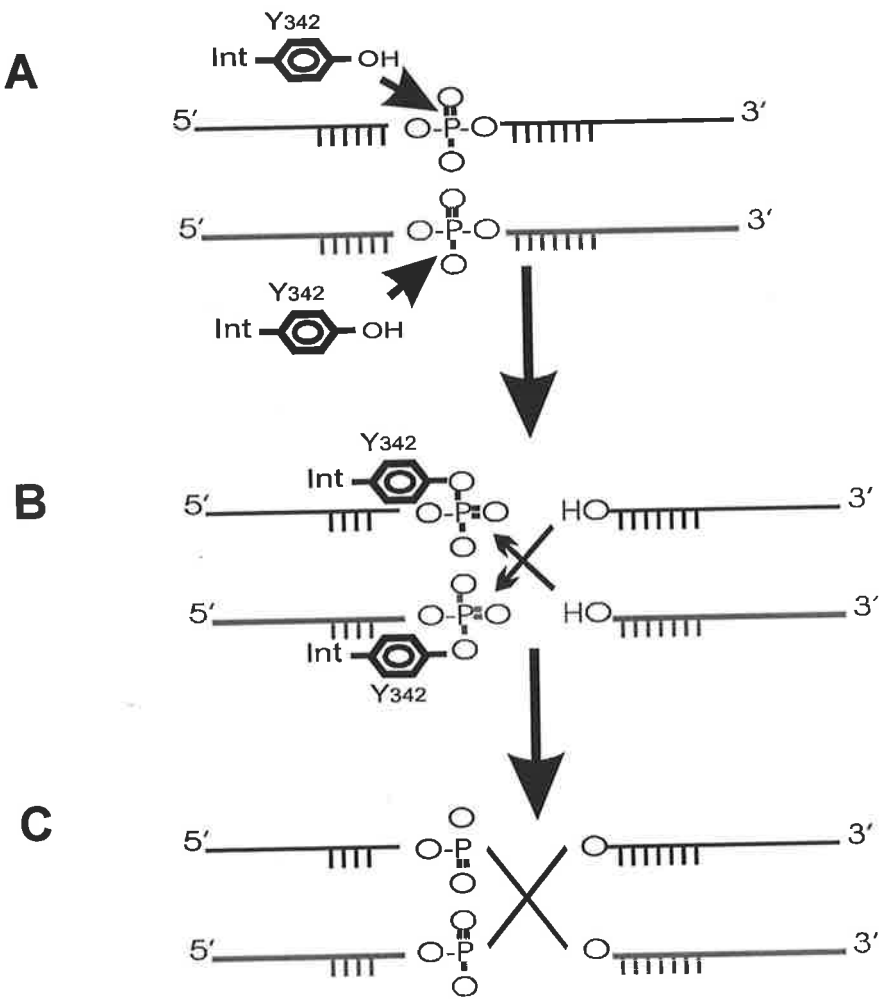


Figure 3.1 A schematic outline of λ Int action. **A.** Following assembly of the intasome and the capture of a protein-free *att*B, the reaction begins with the cleavage of one strand of each of *att*B and *att*P. The cleavage reaction is performed by the conserved tyrosine residue (λ; Tyr342) which nucleophilically attacks the phosphate backbone. **B.** This results in the formation of a covalent bond between the Int protein and the 3' phosphoryl group at the site of strand cleavage, leaving a free 5' hydroxyl group. **C.** The cleaved strands from *att*P and *att*B then exchange positions, and the free 5' hydroxyl groups of the broken DNA attack the covalent bond between Int and the 3' phosphoryl groups releasing Int and religating the cleaved strands. The outcome of this sequence of events is the formation of a Holliday junction, which is subsequently resolved when the cleavage, exchange and religation steps are repeated with the remaining two stands.

λ Int has two DNA binding domains which recognise two different DNA sequences. The amino-terminal domain of Int bind the "arm-type" sites of *att*P with high affinity while the carboxyl-terminal domain binds to the "core-type" sites of *att*P with low affinity (Moisoto de Vargas *et al.*, 1988). Integration host factor (IHF) is required for both excision and integration, and functions as a sequence-specific DNA bending protein, stimulating the formation of Int protein bridges (Gardner and Nash, 1986; Robertson and Nash, 1988; Goodman and Nash, 1989; Snyder *et al.*, 1989; Goodman *et al.*, 1992) . The *att*B site does not bind any proteins directly, instead it acquires them from the *att*P intasome which contains two Int molecules with empty "core-type" binding sites (Richet *et al.*, 1988). Although the four core binding sites for λ Int have slightly different sequences from each other, the consensus dyad symmetry associated with C, C', B and B' is thought to permit positioning of the topoisomerase activity of Int at either end of the overlap region (Ross and Landy, 1982; Leong *et al.*, 1985 and 1986; Nash and Robertson 1989).

The 186 Int protein is a member of a large family of integrases that share a group of invariant residues, the arginine-histidine-arginine catalytic triad (Argos *et al.*, 1986; Ambreski and Hoess, 1992; Han *et al.*, 1994) and the tyrosine nucleophile (Pargellis *et al.*, 1988). Within the family of integrases the C-terminal domain is the most highly conserved, particularly in a region of 40 residues containing two of the three invariant residues of the catalytic triad and the active tyrosine residue (Argos *et al.*, 1986; Ambreski and Hoess, 1992). It is likely that many members of the Int family will share structural properties similar to λ Int, such as possessing two DNA binding domains which recognise two different DNA sequences. Indeed, Yu (1993) has demonstrated for P2 Int that it also contains two DNA binding domains with different sequence specificities. P2 and 186 are highly related (Bertani and Bertani, 1971), and although no limited proteolysis of the 186 Int protein and DNA footprint analysis has been undertaken at this stage, by comparison with the location of the P2 Int sites in *att*P with the 186 *att*P sequence, Reed (1994) identified five highly conserved Int arm-type sequences. If we are to develop the 186 system for chromosome manipulation in higher eukaryotes we

will need to identify precisely the position of the *att*B site and the bases that are important for 186 integrase action. Once we know the binding determinants for 186 Int in BOB' and have demonstrated 186 integrative recombination in a higher eukaryote, we can search for a closely related sequence in a mammalian genome, and later via mutagenesis of the Int protein, "evolve" a modified Int that will mediate recombination between a pre-existing mammalian BOB' site and an engineered POP' site.

In this Chapter, I describe deletion analysis to locate the BOB' site, and further, using duplex oligonucleotides representing the BOB' sequence, containing single phosphorothioate nucleotide substitutions, the precise locations of the top and bottom strand crossovers are identified. Finally, mutagenesis of the B and B' Int core-type binding sites revealed base pairs important for recombination. Surprisingly, within the dyad, the important bases in B are different to those of B', indicating asymmetry in the action of Int at the junctions of the overlap.

3.2 Results

3.2.1 Localisation of BOB' within the common sequence as determined by deletion analysis

The attachment sites *att*P and *att*B for 186 share a common sequence of 36 bp. However, after a single base pair mismatch, the sequence identity continues leftward for a further 12 bp, giving a 48/49 bp match over the region of homology (Appendix, Figure 1.). Following the Campbell model of integration, it is within this region of homology that the crossover must take place during the recombination event. To locate BOB' a series of PCR primers (Table 1(b)) were designed to enable the construction of a series of fragments progressively deleted across the common sequence in 10 bp increments from the 5' and 3' ends (Figure 3.2). The digested (*EcoR*I/*Hind*III) and purified PCR products (size range; 25-177bp) were cloned into a Bluescript vector, pBS SKII$^+$, and a *Pvu*II fragment

Figure 3.2

BOB' DELETIONS

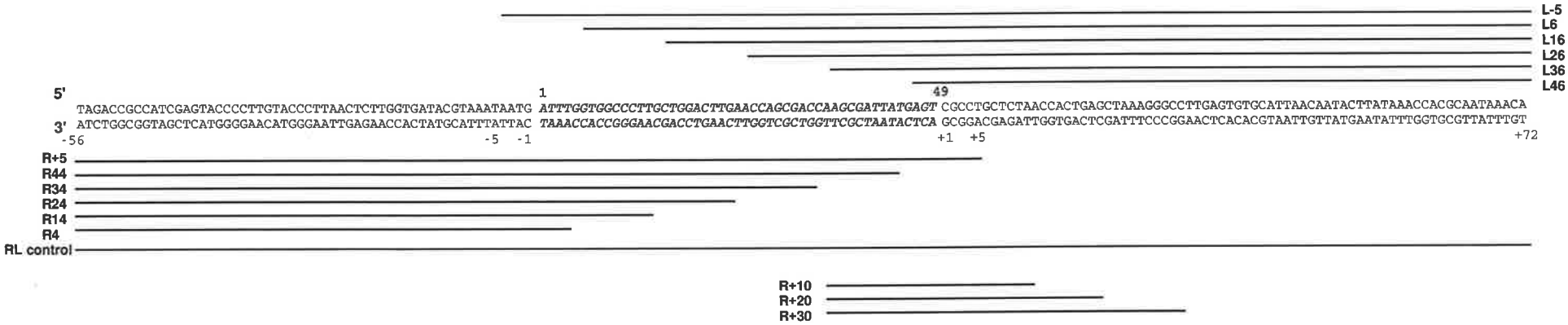


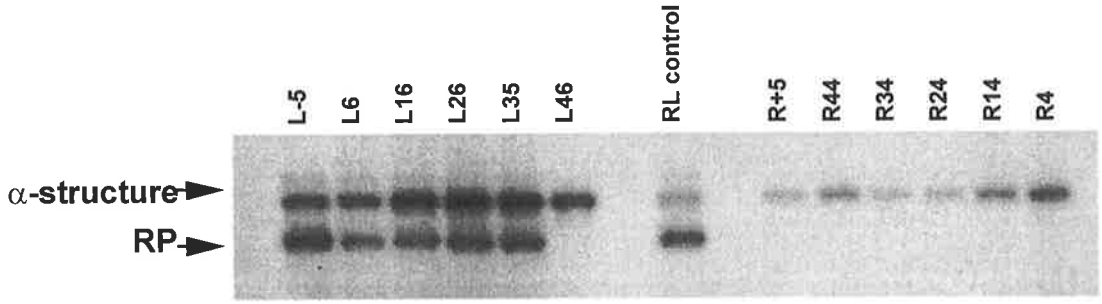
Figure 3.2 Diagrammatic representation of the BOB' deletion fragments generated by PCR and cloned into pBS SKII+ (6.6.4) The horizontal lines represent the sequences not deleted. The DNA sequence of the primers used are shown in 6.4.7 (Table 1(b)).

containing the specific deletion (size range; 461-613bp) was isolated and 5' end-labelled (6.6.19). In the *trans in vitro* recombination assay linear end-labelled deletion fragments were incubated with plasmid pP'OPNeo and IHF and 186 Int proteins. I was only interested in whether each deletion fragment could recombine or not, as such, no attempt was made to exactly quantify the amount of 5' end-labelled product recovered after purification to remove unincorporated label, hence there is variation in the intensity of bands seen due to the variation of losses incurred through the purification procedure.

The expected linear recombination product resulting from reciprocal exchange between pP'OPNeo (5.25 kb) and a labelled linear (*Pvu*II) BOB' deletion fragment would be a labelled molecule migrating with the mobility of a 5.78 kb (± 76 bp) linear DNA fragment. In Figure 3.3 a contaminating band, present in all tracks, running with slower mobility than the correct recombination product band, is a result of vector sequence present in the *Pvu*II fragment of pBS SKII$^+$ having significant homology to the BOB' sequence, with incomplete recombination resulting in the formation of a Holliday intermediate, an α-structure (discussed later).

All 5' deletions from L-5 to L35 (taking into account that the leftward cloning into a *EcoR*1 site reconstructed the wild-type bp at position 35) of the common region gave rise to the correct recombination product (Figure 3.3(a)), However, 5' deletion of sequence beyond this point, as in L46, resulted in the absence of the correct recombination product band. Therefore, I deduced that sequence between L35-L46 was needed for a functional *att*B site. All 3' deletions from R+5 to R4 failed to give rise to the recombination product band, hence it was deduced that the full sequence of BOB' lies further 3' than the R+5 deletion start point. The deletion results indicate the *att*B site for 186 lies toward the extreme 3' end of the common sequence and that 3' sequence, greater than 5 bp beyond the common sequence, is required to form a functional *att*B site.

Figure 3.3(a)



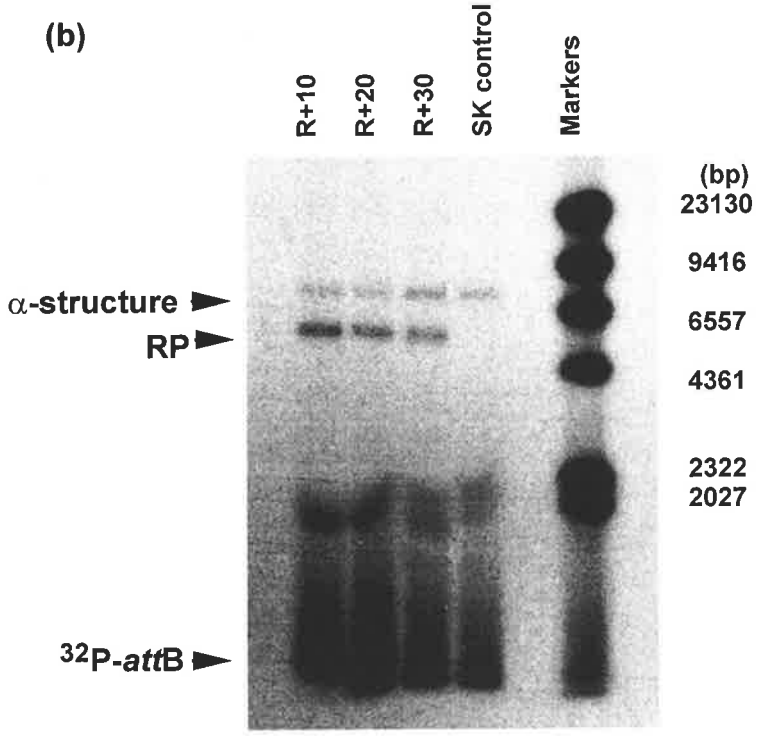


Figure 3.3 Autoradiographs showing results from *trans in vitro* integrative recombination assays. Supercoiled *att*P and linear, ^{32}P-5'-end-labelled *Pvu*II *att*B deletion fragments were incubated as described in Materials and Methods (*6.6.12.ii*). **(a)** *att*B deletion fragments L-5 to L35 generate the correct recombination product (RP) band. *att*B deletion fragments R+5 to R4 failed to generate the RP. A slower migrating band (a-structure) is present in all tracks arising from vector sequence (discussed in text). **(b)** Additional *in vitro* integrative assays using *att*B deletion fragments R+10, R+20 and R+30 locate *att*B between L35 and R+10. The vector control track shows the contaminating band. Free ^{32}P-*att*B is unreacted substrate.

Additional deletion clones were constructed containing sequence starting at L35 of the common sequence and ending at either R+10, R+20 or R+30 (Figure 3.2). In the *in vitro* recombination assay all three clones gave rise to the recombination product band indicating that sequence between L35 and R+10 (outside of the common region) is required to form a functional BOB' site (Figure 3.3(b)).

The contaminating band, present in all tracks, running with slower mobility than the correct recombination product band, is a result of vector sequence present in the *Pvu*II fragment of pBS SKII$^+$ having significant homology to the BOB' sequence. A *Pvu*II fragment isolated from pBS SKII$^+$ gave rise to the slower migrating band seen in the above results (Figure 3.3(b), Track 4). A homology search between the pBS SKII$^+$ *Pvu*II fragment sequence and the BOB' sequence, L35 to R+10, revealed two sequences with significant homology to BOB', either to the right or left (Figure 3.4), which may have resulted in the formation of the α-structure (1.7.5). One of these homologous regions, starting at position 552 in the pBS SKII+ vector, contains a high region of homology if the extra 'G' in the 5' region is ignored. At the present time we do not know if 186 Int would tolerate an extra basepair insertion at this position thus it seems appropriate to include it as a possibility.

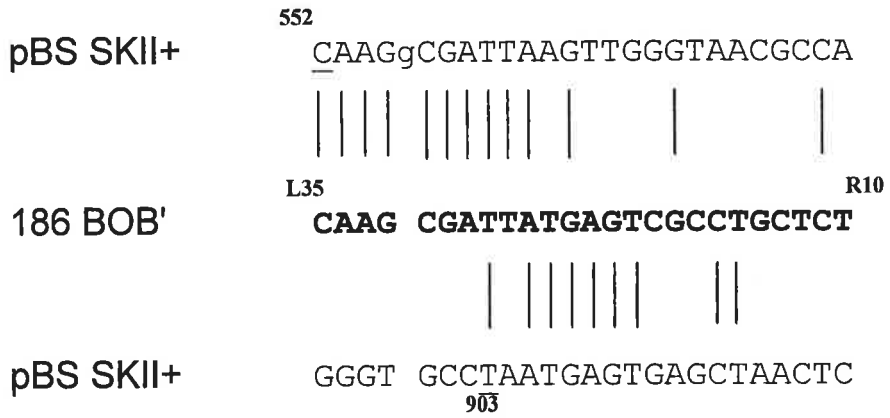


Figure 3.4 The *Pvu*II fragment of pBS SKII+ contains two sequences with significant homology to the 186 *E. coli attB* site (in bold print). The start site of the region of homology is indicated by an underline and the location in pBS SKII+ by the site number. The lowercase 'g' represents the extra base; if this is ignored homology continues on further rightward. All sequences are shown 5' to 3'.

Confirmation of which homologous region in the vector contributed to the α-structure formation was made by determining the cleavage sites (3.2.2) and the order of strand exchange (3.2.6). From the data generated I concluded that sequence at the 903 start site of the region of homology was responsible for the formation of the α-structure. Specifically, since the 903 region of homology only contains the bottom strand cleavage position, the site of first strand exchange for 186 Int, and no second strand cleavage position, an aborted product, a Holliday intermediate is produced known as the α-structure.

3.2.2 Precise location of the crossover sites using phosphorothioate containing oligonucleotides.

A phosphorothioate nucleotide differs from a normal nucleotide by having one of the thioester oxygen atoms on the phosphate replaced by an atom of sulphur. In *att*B-λ (Kitts *et al.*, 1984; Kitts and Nash, 1988), *att*B-HK022 (Kolot *et al.*, 1994) and *att*B-L5 (Péna *et al.*, 1996) the presence of an α-phosphorothioate at the Int cleavage positions inhibits *in vitro* integrative recombination, whereas α-phosphorothioate substitutions at other positions are not inhibitory. Initially I showed that two annealed 45mer oligonucleotides (L30-R+25) representing the *att*B-186 site could be used as a substrate for the *in vitro trans* recombination reaction (data not shown). Oligonucleotides (Table 1(b)) were synthesised to contain a single α-phosphorothioate nucleotide (dNTPαs) substitution in the top strand at position 40, 41, 42 or 43 of the common sequence and annealed to a bottom strand oligonucleotide containing normal dNTPs. Similarly, oligonucleotides were synthesised to contain a single dNTPαs in the bottom strand at position 46, 47, 48 or 49 and annealed to a top strand oligonucleotide containing normal dNTPs.

In Figure 3.5, the reaction products resulting from *in vitro* integrative recombination using the single dNTPαs substituted oligonucleotides are shown. The expected linear recombination product (RP) of 5.29 kb was obtained in all reactions but was markedly inhibited where a dTTPαs replaced a dTTP at position

Figure 3.5

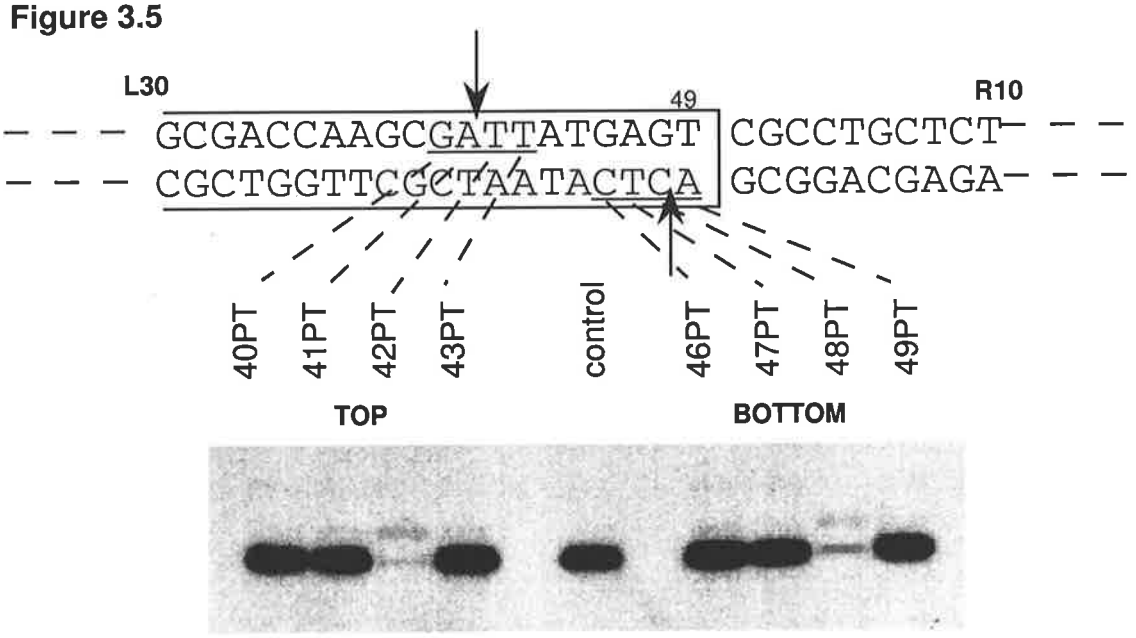


Figure 3.5 The sequence of the two 45mer oligonucleotides used as substrate for the *trans in vitro* assay is partly shown. The open-ended box indicates sequence present in the common region. The position of nucleotides, 40 to 43 in the top strand, and 46 to 49 in the bottom strand, singly synthesised as α-phosphorothioate nucleotides (dNTPαs) are underlined (6.4.7, Table 1(b)) and indicated by the dashed lines. The autoradiiograph shows the results of the *in vitro* recombination reaction using the dNTPαs substituted oligonucleotides, and the arrows depict the deduced cleavage sites.

42 in the top strand, and where a dCTPαs replaced a dCTP at position 48 in the bottom strand. I concluded that top strand exchange results from cleavage between positions 41 and 42 and bottom strand cleavage occurs between positions 48 and 49 of the common sequence. I was not able to definitively determine the order of strand exchange as neither showed a marked level of α-structure band.

3.2.3 BOB and B'OB' can substitute for BOB'

The core-type sites, B and B', are imperfect dyads with a single basepair mismatch (Figure 3.6). We asked the question, can BOB or B'OB' substitute for the wild-type BOB' site?

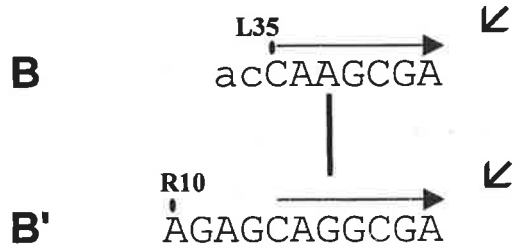
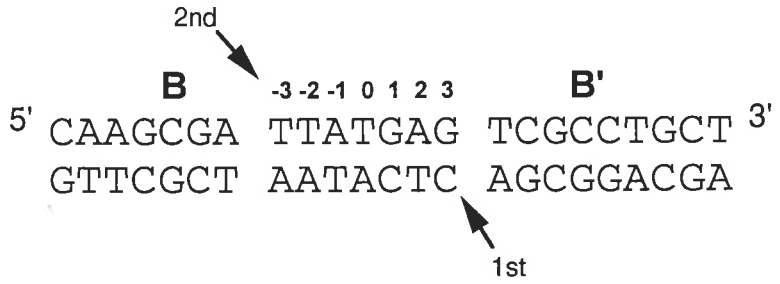
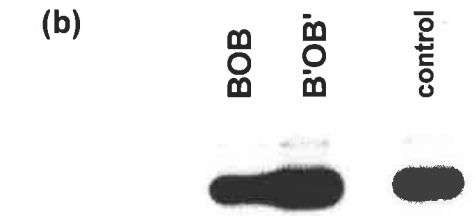


Figure 3.6 The vertical line indicates the basepair mismatch in B and B'. The horizontal arrows represent the imperfect dyad. In B, the bases in lowercase are those that lie outside of the known sequence requirements (as determined from the deletion data results). Sequences are 5' to 3'.

The results, presented in Figure 3.7, demonstrated that both BOB and B'OB' can substitute equally well for BOB'. Therefore, the data implied that this basepair is not an important binding determinant for 186 Int, since an A/T to G/C basepair change is well tolerated when B' behaves as B in B'OB'. One can observe from the autoradiograph that B'OB' appears to function better than both BOB and BOB', which function equally well as each other. The experiment has been repeated three times (data not shown) and it appears to be a reproducible result. Therefore the inference is that although an A/T to G/C change is tolerated at ±8, it would appear that the G/C basepair actually enhances the recombination competency of the site.

Figure 3.7(a)





BOB 5' GCGACCAAGCGATTATGAGTCGCTTGGTCGC 3'
CGCTGGTTCGCTAATACTCAGCGAACCAGCG

B'OB' 5' TAGAGCAGGCGATTATGAGTCGCCTGCTCTA 3'
ATCTCGTCCGCTAATACTCAGCGGACGAGAT

Figure 3.7(a) The *att*B site and the locations of the 1st and 2nd 186 Int cleavage sites; the positions of the -3 to +3 bases of the overlap region and the relative positions of B and B'. **(b).** The autoradiograph shows the results of the trans *in vitro* integrative recombination assay using the synthetic substrates BOB, B'OB' and the control BOB', and the sequence of the oligonucleotides used.

3.2.4 Mutagenesis of bases in B and B'

The C and C' segments of *att*P-186 and the B and B' segments of *att*B-186 presumably provide the binding sites for 186 integrase for strand cleavage and transfer reactions. In this section of work some initial progress has been made in determining the bases in B and B' that are important for 186 recombination. Earlier, the deletion data results indicated that sequence L35-R+10 was required to form a functional *att*B site, therefore it was likely that duplex BOB' oligonucleotides representing this sequence range would be sufficient for this section of work. Initially, comparisons of the *trans in vitro* recombination ability of three duplex BOB'-oligonucleotides of varying length, a 30mer (L31-R+11), 25mer (L35-R+10) and the control 45mer (L30-R+25) were undertaken. Results indicated (data not shown) that a 30mer was as proficient as the 45mer, while the 25mer failed to generate the recombination product.

For an initial mutational analysis of BOB'-186, I made single basepair changes (A→C, T→G, G→T, C→A) in the 30mer oligonucleotide representing the 186 BOB' site (6.4.7; Table 1(c)) and analysed the effect the substitution had on the recombinational competency of BOB' in the *trans in vitro* assay. The base pairs to the left of the 7bp overlap sequence were designated –7 to –1, and to the right, +1 to +12, with the dyad 7bp either side of the overlap region. Examples of autoradiographs obtained are shown in Figure 3.8(a), (b) and (c), and are representative of four independent experiments.

The most striking feature of the results is the asymmetry of bases (apparently) important for recombination in the dyad. The bases, in B, at positions –7, -6, -5, -4, and –1, were very highly important (a bp change results in >80% decrease in recombination). In B', by comparison, the only bases highly important were at +1, +6 and +9, while +4 (conserved in all four half-sites) and +7 were only moderately important (50-70% decrease in recombination). This data was consistent with the deletion results which indicated sequence between L35 and R+10, outside of the common region, to be important. Furthermore, a G/C to T/A basepair change at position +3 appears to enhance the recombinational competency of the site,

Figure 3.8

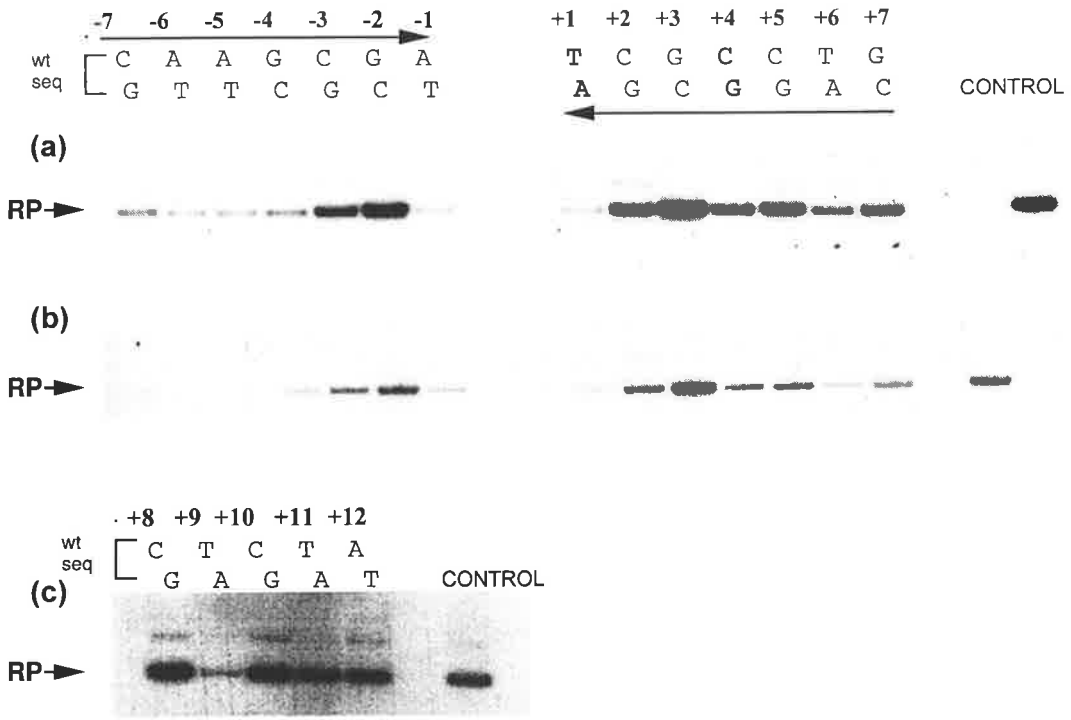


Figure 3.8(a), (b) and (c). Mutagenesis of bases in *att*B. The wild-type (wt) *att*B sequence is shown and the numbers above, -7 to -1 and +1 to +12, correspond to the bases to the left and right of the 7bp overlap sequence respectively. The autoradiograph shows the results of the *in vitro* integrative recombination assay when mutagenesis of the wt sequence (A to C, G to T, C to A and T to G) was performed singly at each position. See 6.4.7, Table 1(c) for the synthetic oligonucleotides used for B and B' mutagenesis. The two basepairs conserved in B, B', C and C' are indicated in bold print and the arrows indicate the dyad in B and B'. Panels (a) and (b) represent the same experiment performed in duplicate. The control is wt *att*B duplex oligonucleotides, used as substrate in the *trans in vitro* recombination assay. RP is the recombination product band.

probably by intensifying the bending of the site by increasing the A/T base content. A summary of the important base pairs in B and B' is given below (Figure 3.8(d));

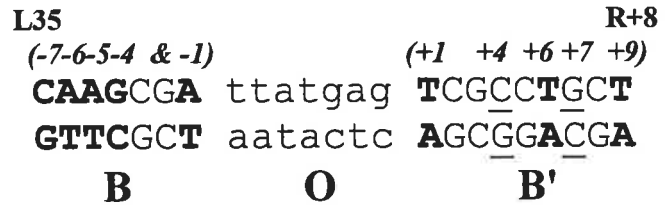


Figure 3.8(d) The base pairs showing a very high degree of importance (a bp change results in >80% decrease in recombination) are shown in bold, +4 (the conserved basepair in all four half-sites) and +7, showing a moderate level of importance (50-70% decrease in recombination) are indicated by an underline.

It is interesting to note that the important T/A basepair at the 3' end in B' (+9) aligns with an extended BOB' dyad sequence in B (-9) i.e.

-9 +9
TGGTTCGCTaatactcAGCGGACGA

where the bases of the dyad are indicated in bold type.

3.2.5 Minimal DNA sequence required to form a bacterial attachment site (BOB')

The two outermost important bases lie at –7 (position 35 of the common sequence and to +9 (8bp outside the common sequence). This corresponds to the deletion data as sequence L35 to R+10 was required for a functional 186 BOB' fragment, taking into account that the leftward cloning into a *EcoR*1 site reconstructed the wild-type sequence from position 35. Hence the minimal DNA sequence needed for a fully functional 186 bacterial attachment site (BOB') is 23 bp.

However, 25mer annealed oligonucleotides (-7 to +9 or L35 to R+10) containing the BOB' site (CAAGCGATTATGAGTCGCCTGCTCT) cannot function as a substrate for the *in vitro* recombination assay whereas 30mers are able to. This suggests that the 186 Int recombinase requires the extra length of DNA, either side of the minimal sequence, for steric reasons only.

3.2.6 Mutagenesis of bases in 'O'

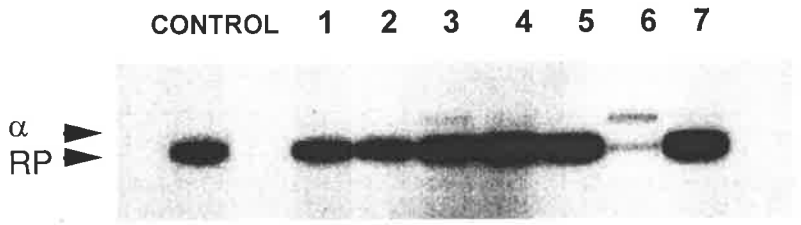
To ascertain if any bases in the overlap region are important it was necessary to mutagenise individually each of the seven base pairs of 'O' in BOB' and pP'OPNeo to maintain homology. Oligonucleotides were used for BOB' (6.4.7 Table 2(a)). Site-directed mutagenesis (6.6.23) was performed on pP'OPNeo (6.4.7 Table 2(b)) and clones representing the appropriate change in each of the seven bases in 'O' were confirmed by sequencing (*6.6.9.ii*). The *trans in vitro* recombination assay showed the basepair at position 6 in the overlap region (TTATGAG) was important for recombination (Figure 3.9(a)).

Furthermore, recombination between the seven BOB' overlap mutants and the wild-type POP' site showed, as with λ (Kitts and Nash, 1988) and HK022 (Kolot, 1994), that strand exchange proceeds in a defined order (Figure 3.9(b)). Non-homology in the right of the overlap region blocks the formation of all products indicating this to be the point of first strand exchange. In contrast, non-homology in the left of the overlap only blocks formation of completed recombination products, but permits accumulation of α-structures, indicating this point to be the location of the second strand exchange. By determining the order of strand exchange, it allowed me to choose between the two alternatives of the pBS SKII+ vector sequences (Figure 3.4) contributing to the α-structure band in the deletion data results. It would appear 903 (pBS SKII+) with homology at the right end is the one as the presence of the α-structure relies upon successful first strand exchange and aborted second strand exchange. The 903 sequence only contains significant homology around the location of first strand exchange.

3.2.7 Insertion of an extra helical turn (10 bp) of DNA into the POP' and BOB' overlap sequence

The results in 3.2.3 demonstrated that both BOB and B'OB' can substitute for BOB'. Mutagenesis data indicated that of the 23 bp required for a functional BOB' site, we had at most 14 bp which could be altered from the wild-type sequence. However, there is evidence of some sequence degeneracy in the system because at

Figure 3.9(a)



(b)

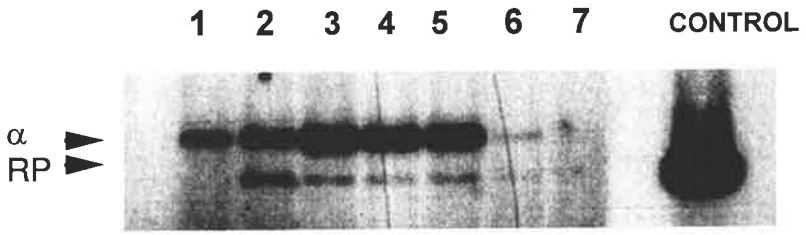


Figure 3.9 Tracks 1-7 correspond to bases in the overlap region (5'- TTATGAG -3') of *att*B. The autoradiograph shows the results of the *trans in vitro* recombination assay when mutagenesis of the wt sequence (A to C, G to T, C to A and T to G.) is performed singly at each position and the mutant *att*B sites are reacted **(a)** with homologously mutant *att*P supercoiled substrates in the presence of Int and IHF **(b)** with wild-type *att*P supercoiled substrate in the presence of Int and IHF.

least at one position (-5) an A→ G change was well tolerated in B'OB' (Figure 3.6), while an A→C was not tolerated in the mutagenesis data (Figure 3.8(d)). The implication of this result is that we would need to test all three base changes at each position in BOB' in order to fully analyse which alterations are not permitted. At present, of the 23 bp for the minimal *att*B-186 site, we have at 11 sites which cannot have a particular bp at a given site, and consequently calculations relying upon ¼ probability at a given site are in doubt until all three base substitutions are checked. One way of increasing the chances of finding an existing and unique mammalian target BOB' site would be to insert an extra helical turn (10 bp) of DNA into the overlap regions of BOB' and POP'. If the system permitted such a change our chances of finding a unique site in a mammalian genome would be highly probable as the 7 bp overlap region, carrying one integrase determinant, would give us the 16 bp required for uniqueness.

The insertion of 10 bp, an extra helical turn, into the two overlap sequences may preserve the correct phasing of the integrase protein action at the four core-binding sites B, B', C and C'. A DNA sequence of 10 bp was inserted into both overlap regions 5' to the central basepair [TTA(gatacatcat)TGAG]. The 5' position, rather than the 3' position, was chosen as the A/T basepair at position +6 of the overlap had been demonstrated to be important for recombinational competency (this work) and therefore, it was thought that the greater the distance between this site and the site of first strand exchange, and the insertion of 10 bp, the more likely it was that the insertion might be tolerated by the system. The actual sequence chosen for the insertion was essentially random, except that a sequence low in homology to existing core-type sequences and with a high A/T content was preferentially selected. Duplex oligonucleotides were used for *att*B while site-directed mutagenesis was performed on pP'OPNeo to insert the identical 10 bp sequence (6.4.7 Table 3). The correct insertion of the 10 bp sequence for *att*P was confirmed by sequencing.

Initial results showed that for the sequence used, and the position of insertion of the extra 10 bp, two homologously reacting *att*B∇^{10} X *att*P∇^{10} were not

recombinationally competent (Figure 3.10). No correct recombination product resulted from the *trans in vitro* recombination assay (repeated twice). Furthermore, no Holliday intermediates (α-structures), indicative of first but no second strand exchange, were evident indicating that 186 Int cannot function on the *att*BV^{10} X *att*PV^{10} substrates at all. This was somewhat a surprising result given that the 903-pBS SKII homologous region to 186-*att*B does give rise to the α-structure having 5 bp of the right side of the overlap region in common, while *att*BV^{10} having 4 bp does not. Therefore it is likely that the insertion of this particular extra 10 bp sequence in the overlap caused a steric distortion of Int action at the site of first strand cleavage. If the 186 recombination system can be shown to operate in a eukaryotic environment *per se*, at a later date a more comprehensive series of mutational analysis would be undertaken, varying the position, sequence and number of base pairs inserted.

3.3 Discussion

3.3.1 The minimal sequence and location of BOB'

Early experiments showed that a DNA segment of 333 bp containing the *att*B 36 bp common core was a good substrate for recombination *in vitro* (this work). However it was unlikely that the entire segment was needed for integration. Indeed for λ, only 21 bp of *att*B DNA is required for λ integrase-mediated recombination (Mizuuchi and Mizuuchi, 1985), and for the temperate bacteriophage L5, only 29 bp is needed (Peña et al., 1996). To determine the precise location of *att*B and the minimum amount of DNA sequence necessary for *att*B function, I performed deletion and mutagenesis analysis and tested the resulting substrates in the *trans in vitro* recombination assay. Taken together, the data indicated that a 23 bp sequence of *att*B DNA (L35-R+8) was required for proficient integrative recombination, which included DNA 8 bp to the right of the common sequence.

Figure 3.10

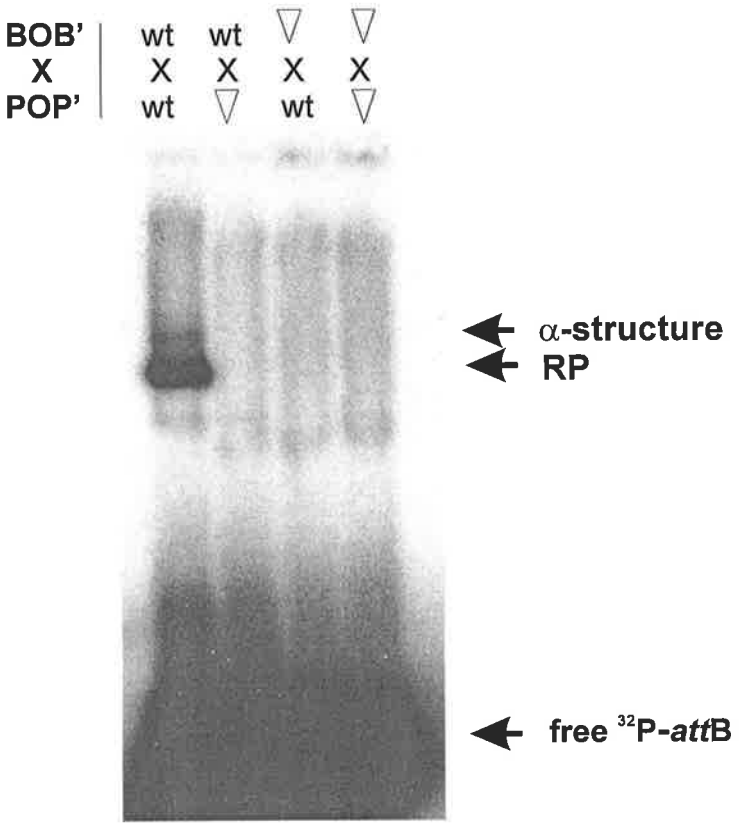


Figure 3.10 An autoradiograph showing results from a *trans in vitro* integrative recombination assay using supercoiled wild-type (wt) or ∇^{10} *att*P, and ^{32}P-end-labelled wild-type or ∇^{10} *att*B. The positions of recombination product and α-structures is shown.

3.3.1.i A consensus sequence for the core-type sites cannot be derived

A surprising result from this work was that a consensus sequence could not be derived for the core-type sites B, B', C and C' (Figure 3.11).

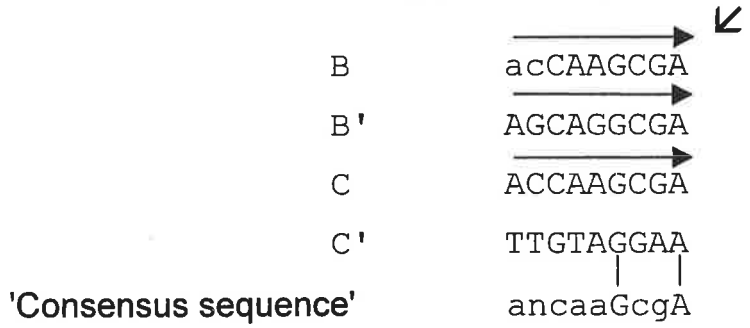


Figure 3.11 The core-type sequences B, B', C and C' are shown. The lower case letters in the B core-type site represent the bases that, from the deletion data, are known to be non-essential for a fully functional BOB' site. The horizontal arrow shows the region of dyad symmetry. The diagonal arrow shows the position of 186 Int cleavage for all four sites. The vertical lines show the two base pairs conserved in all four sites, while the lowercase letters in the consensus indicate a ¾ consensus for that site. All sequences are shown in the 5' to 3' direction.

Alignment, by the points of strand cleavage, of the C, C', B and B' sites does not give a good consensus sequence for 186 integrase action at the four cleavage positions. Surprisingly, there are only two conserved basepairs for all four sites. Notably, C' has almost no similarity, apart from these two bp, to the other three sites and does not contain a consensus dyad sequence. In the phage L5-*att*B site, much of B and C also lie outside of the common region like 186 B' and C' (Peña *et al.*, 1997), and the consensus sequence (anCtAgT$^{\downarrow}$) is only marginally better than the 186 consensus for all four sites. The requirement for dyad symmetry was thought to be a highly conserved aspect of bacteriophage recombination systems and to permit recognition by the core-type binding domain of the integrase at the junctions of the overlap sequence. Furthermore, in contrast to λ, it is likely that COC'-186 would not be able to substitute for BOB' (Mizuuchi and Mizuuchi, 1980). The prediction is easily testable, but has not undertaken as part of this project.

Direct identification of the P2 Int core sites has not been undertaken, however Campbell (1992) has suggested the position of a 7 basepair overlap (CGTGTAA)

region within the common sequence of P2 which gives C, C', B and B' (*loc*I) almost perfect dyad symmetry. Reed (1994) predicted the 186 Int consensus core binding site to be 5' YYGCTGG 3' (C, C', B and B'), present as inverted repeats, separated by 7 base pairs and positioned at the left end of the common sequence (Appendix, Figure 1). In disagreement with Reed's (1994) prediction for the location of the *att*B site, the data presented in this work locates the *att*B toward the extreme right end of the common sequence, and that sequence (8bp) beyond the common region is required to form a functional *att*B site.

Although only 186 and L5 appear to be the exceptions to the rule at this stage, perhaps others should be wary of predicting *att*B sites based on the notion of a consensus dyad symmetry for core-binding sites, until the actual sites of cleavage are confirmed.

3.3.2 Determination of the order and positions of strand exchange

The positions of strand exchange in the 186 *att*B site were determined using a similar experimental strategy employed for *att*B of λ (Kitts *et al.*, 1984; Kitts and Nash, 1988), *att*B-HK022 (Kolot *et al.*, 1994) and *att*B of L5 (Péna *et al.*, 1996). The strategy is based on the observation from the λ recombination system that a phosphorothioate-substituted nucleotide (dNTPαs) at the position of strand cleavage inhibits recombination, whereas a dNTPαs at other positions does not. The system employed for the 186 recombination system was a variation on that used by others. In my system a series of oligonucleotides were used which had a single phosphorothioate nucleotide substitution, either in the top or bottom strand of *att*B, where as others have synthesised *att*B substrates using the incorporation of a particular dNTPαs resulting in multiple sites containing the dNTPαs. One assumption is made that because 186 belongs to the Int family of recombinases then it too is likely to share similar chemistry for cleavage, thus the results indicated that 186 integrase cleaves DNA to produce 7 bp 5'staggered cuts analogous to those produced by λ, HK022, HP1 and P22 integrases (Kitts *et al.*, 1984; Kitts and Nash, 1988; Kolot *et al.*, 1994; Hauser and Scocca, 1992; Péna *et*

al., 1996). Taken together the results show, as for other members of the Integrase family λ, HK022 and L5, that 186 has a 7 basepair overlap sequence resulting from a 186 Int-induced 5'-staggered cut. Furthermore, analysis of the *att*B site revealed that 186 DNA integrates into the anticodon (Figure 3.12) of a putative novel isoleucine tRNA gene closely related to *ile*X. The use of tRNA genes as sites of insertion is a phenomenon common to many lysogenic phage (Reiter *et al.*, 1989; Hauser and Scocca, 1992; Brown *et al.*, 1994; Smith-Mungo *et al.*, 1994).

3.3.2.i 186 att*B lies in a putative novel isoleucine tRNA gene*

The *att*B site had been previously mapped between the *phe*A and *nal*B loci (57 min) on the *E. coli* chromosome (Woods and Egan, 1972). Reed (1994) undertook a DNA sequence data base search with the *att*B sequence which revealed striking similarity (74/76 bp identity) with the *E. coli ile*X gene for $tRNA^{Ile}$. The identification of the 186 Int cleavage sites locates them to the anticodon loop (this work). The sequence in common with *att*P is located in the 3' half of the putative tRNA gene. Reed (1994) identified a potential weak promoter sequence some 200 bp upstream of the gene, but no likely terminator sequence following the gene was found.

Expression studies of tRNA genes *in E. coli* (Komine *et at.*, 1990) did not identify a tRNA gene in the locus described here. Given the high degree of conservation between the *ile*X gene and the putative gene identified by Reed (1994) it is possible that the *ile*X gene could also function as a 186 attachment site as the A to G change in *ile*X relative to 186 *att*B is outside of the minimal 23 bp 186 *att*B site identified (this work). The two single base pair mismatches between this gene and the *ile*X gene replace a G-C pair with an A-U pair in the acceptor stem of the tRNA, while in a lysogen the sequence of the tRNA gene is changed again such that the A-U pair is replaced by a G-U pair (tolerated in RNA secondary structures). Both changes may not alter the functionality of the molecule. However, at least eight independent 186 lysogens have mapped to the 57 min region of the *E. coli* chromosome (Woods and Egan, 1972; Reed, 1994), whereas

Figure 3.12

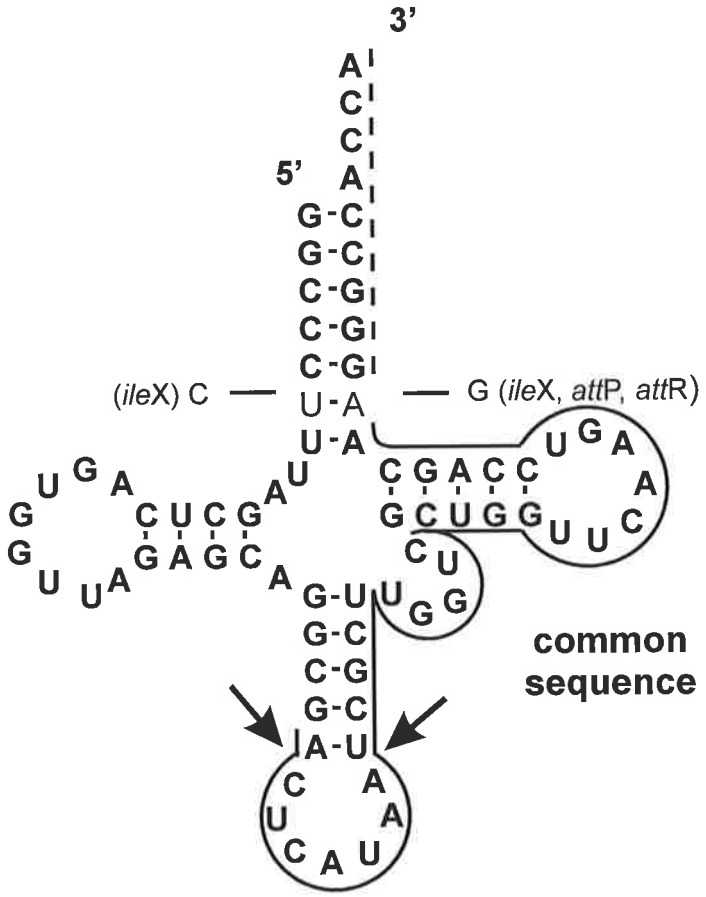


Figure 3.12 The proposed tRNA structure of the *ileX* tRNA gene at *att*B. The solid line indicates the *att*B common sequence, the dashed line the sequence identity 3' to the 1 bp mismatch. The positions of strand cleavage relative to tRNA structure is denoted by two arrows. The sequence differences between *att*B, the *ileX* gene and *att*P are shown.

*ile*X is at 67 min (Komine *et al.*, 1990), suggesting that any use of *ile*X as an attachment site is infrequent, or more probably, that a prophage integrated at the *ile*X site renders the *ile*X gene inactive and the resultant *E. coli* (*ile*X$^-$) is no longer viable.

3.3.2.ii Position of attP

The left breakpoint of a phage deletion isolated from a prophage (186Δ1; Kalionis *et al.*, 1986) was predicted to have a common end-point in the core of the *att* site. Results from work presented in this Chapter confirm this to be correct, the left end of the *del*1 deletion does locate the rightside of the common sequence in *att*P.

3.3.3 Analysis of the mutagenesis data

3.3.3.i Overlap mutants

In section 3.2.6, the *trans in vitro* recombination assay showed the basepair at position 6 in the overlap region (TTATGAG) was important for recombination (Figure 3.8(a)). For the –1 site of the λ overlap region (TTTATAC), it too has also been shown that base substitution results in reduced recombination even if the same change is made in *att*P (Kitts and Nash, 1987). It is interesting to note that for both 186 and λ that the site of first strand exchange lies to the side closest to this basepair in the overlap region (see below).

i.e.

3.3.3.ii Asymmetric importance of bases in the B and B' dyad

I investigated, by limited mutagenesis, the important base pairs in B and B' for maintaining recombinational competency of the *att*B site. The data generated indicated asymmetry with respect to the important bases within the almost perfect dyad for B and B'. When the same basepair changes were made in each equivalent

position of the dyad, and the substrates reacted separately with the wild-type POP' site, results indicated that only at two equivalent positions in the dyad (-1,-6, +1 and +6) were all highly important, of which one is the basepair in each dyad adjacent to positions of strand cleavage. Mutagenesis of other base pairs within the B half-dyad produced a different profile of importance compared to the profile produced by the same mutations in the B' half-dyad. The result raises two questions: What is the role of the dyad? Further, if it is unimportant to have a conserved dyad, provided only the positions of important base pairs are maintained, then how does 186 Int interact with two highly related sites so differently?

The apparent non-equivalent recognition of B-186 and B'-186 by Int questions the notion that the dyad permits recognition by the core-type binding domain of the integrase at the junctions of the overlap sequence. Mutagenesis has also been performed on the FRT site of the FLP recombination system which showed a direct correlation between decreases in recombination activity and in FLP binding (Senecoff *et al.*, 1988). Further it was shown that, at least at one position within a FRT site, an identical change in the FLP binding sites (double mutant) produced a significantly different effect than the single mutant. The double mutant exhibited greatly decreased reactivity when present in both inverted repeats (FLP binding sites) than when present in only one. From the data presented it is not known whether this is an additive effect of the double mutant, or whether the mutation of one site had a more deleterious effect than the same mutation in the other site indicating asymmetry in their importance to FLP binding. Other double mutants were analysed however, when the single mutant produced a large deleterious effect this masked the effect of the other site, experiments using single (half-site) mutants of left and right sites were not undertaken. Hence the extent of the asymmetry of important bases for FLP binding in the two inverted repeats is not known.

Kim and Landy (1992) found that the same, single base substitution in either of the *att*L core sites resulted in a lower level of excisive recombination than the

equivalent substitution in *att*R core sites. Experiments have been conducted on the highly related λ and HK022 core-type sites to locate the base pairs involved in the specificity difference for the two integrases (Nagaraja and Weisberg, 1990; Dorgai *et al.*, 1998), and again both studies indicated non-equivalent recognition of the four core-type sites. Nagaraja and Weisberg (1990) found the replacement of B'-λ by HK022 sequences repressed recombination more severely than the corresponding replacement of B-λ and similarly, the substitution of B'-HK022 by λ sequences repressed recombination more severely than the corresponding replacement of B-HK022. Dorgai (1998) also found that the same nucleotide substitutions within a core-site had different effects depending on which core-site was changed, indicating that the integrase does not recognise the core-sites equivalently, and further, Dorgai suggested that the differences in sequence recognition among core binding sites are likely to be confined to only a few positions. In my results, the profiles for Int recognition of B-186 and B'-186 are different at most positions within the consensus dyad for B and B'. It would be interesting to perform similar mutagenesis analysis on the λ and HK022 core sites to determine the degree of non-equivalent recognition by integrase.

3.4 Conclusions

In this Chapter, I precisely located the 186 *att*B site, identified the positions of strand cleavage and the order of strand exchange. The mutagenesis conducted on the BOB' site determined the position and numbers of bases important for integrase action and further, some sequence degeneracy in the system was identified. Taken together, the data indicated that we would not find a unique, naturally occurring mammalian target site (mtBOB') and that more extensive mutagenesis analysis would need to be conducted. Further, I investigated whether insertion of an extra helical turn of DNA sequence could provide the extra sequence necessary for selecting a potential unique genomic mtBOB' and found, on a first analysis, that it was unsuccessful. At this stage no further examination of

the BOB' site would be pursued until the 186 integrative recombination system had been shown to operate in a eukaryotic environment.

CHAPTER 4

Studies of 186 site-specific integration using transgenic mice containing genomic copies of the 186 *att*B site

Chapter 4

Studies of 186 site-specific integration using transgenic mice containing genomic copies of the 186 *att*B site

4.1 Introduction

I have investigated an alternative means of efficient transgenic animal production, which may also have applications in human gene therapy. To this end I have been characterising a directional site-specific recombinase system which may permit precise and stable single-copy genetic manipulation of a mammalian genome. To date, no directional systems have been shown to operate in a eukaryotic environment. Related systems, such as the non-directional Cre/lox and FLP/FRT systems (one protein mediates both excision and integration), have been shown to operate in a number of eukaryotic cells (see *1.7.2.i* and *1.7.2.ii*), including fertilised eggs (via microinjection), indicating that such integrases can operate successfully in the presence of chromatin.

The aim of this section of the project was to investigate whether a directional integrase system, namely the 186 site-specific recombinase system, could function in a eukaryotic environment. Two possible approaches for accomplishing this were:

1. Transfect a mammalian cell line with *att*B DNA, select a stable cell line containing at least one genomic copy of *att*B (randomly placed), transfect with pP'OPNeo and the purified proteins Int and IHF, select Neo^R cells and, by PCR and Southern analysis, determine if any resulted from site-specific insertion into *att*B rather than a random insertion.

2. Create transgenic *att*B mice by microinjection of *att*B DNA into a pronucleus of fertilised mouse egg, identify *att*B F0 progeny by PCR, mate the transgenic (TG) male pups to wild-type (WT) females, isolate the

fertilised eggs, microinject pP'OPNeo and the purified proteins Int and IHF into a pronucleus, transfer to a pseudopregnant female and test resultant pups by PCR and Southern analysis for 186 site-specific or random insertion of pP'OPNeo.

While both approaches have advantages and disadvantages, I chose the second method, the primary reason being that the transfer of molecules into a fertilised egg by microinjection is an efficient procedure for studying the activity of molecules in a non-native environment as it permits the transfer of molecules directly into the nuclei of a recipient cell. Hence a pre-formed intasome could be delivered straight to the nucleus where the target site (*att*B) is located. It is likely that an intasome, consisting of several IHF and Int proteins bound to supercoiled *att*P DNA, would be too large to permit passive diffusion into the nucleus, as would be required for the first approach, if introduced by standard transfection methods such as lipofection and electroporation. It is likely that such large protein/DNA complexes would require a nuclear localisation signal for active transportation to the nucleus, with the one caveat that during cell division the intasome may be permitted access to the nucleus. Microinjection into the nucleus can be accomplished in somatic cell lines using specialised equipment (computer-automated microinjection systems), but such equipment was not available to me at the commencement of this section of work. Due to existing limitations in technology, for most livestock animals only a genetically manipulated fertilised egg can produce a transgenic animal. In the future, if nuclear transfer and 'reprogramming' of somatic cells becomes routine, then genetic manipulation of somatic cells to produce a transgenic animal will be possible. The disadvantage of the second approach is that, by comparison to the first, only relatively low numbers of cells (pups) can be assessed, however given current equipment limitations, the advantage of delivering intasomes directly to a nucleus of a fertilised egg by microinjection, in my opinion, outweighed this disadvantage.

The question of how the intasome might access the *att*B sites in nucleosomally organised DNA needs to be addressed. All eukaryotic chromosomes consist of a

regularly repeating protein-DNA complex called the nucleosome (Kornberg, 1977). Each nucleosome consists of a protein octamer, made of two copies each of histones H2A, H2B, H3 and H4, which, together with the fifth histone, H1, organises about 200 bp of DNA. Further organisation involves the assembly of nucleosomes into higher-order chromatin structures. During transcription, the chromatin must unfold to permit the transcriptional machinery access to the DNA. Increased acetylation of histone tails is associated with transcriptional activity, and under-acetylation has been linked with repression (Turner, 1993). Furthermore, it has been discovered that both acetyltransferases and deacetylases form complexes with certain transcription factors, providing a mechanism by which the appropriate region of chromatin can be targeted for 'remodelling' (Gregory *et al.*, 1998). The prevailing view is that acetylation promotes the disruption of the higher-order structure of chromatin (Walia *et al.*, 1998; Luger *et al.*, 1997) and constitutes an important point at which transcription, and presumably other nuclear processes, can be regulated.

If *att*B is located in an area of non-silenced chromatin (euchromatin) it is likely the intasome will be able to gain access for recombination to take place. In the likelihood *att*B is located in an area of silenced chromatin structure (heterochromatin), one can imagine that during cell cycle progression some areas of heterochromatin are remodelled to an open configuration (euchromatin) concomitant with gene activation, and thereby possibly permitting access of the intasome to *att*B. Successful recombination would then depend upon the stability of the intasome through various cell cycles.

In Chapter Four I report the creation of lines of transgenic mice containing randomly located 186-*att*B sites, and examination of whether microinjection of active intasomes (pre-tested *in vitro*) into transgenic fertilised eggs can mediate integrative recombination. However, initially I tested for integrative recombination using isolated genomic DNA (naked) containing *att*B, sites and later performed similar experiments on chromatin isolated from the livers of transgenic mice. The results of these analyses are presented below.

4.2 Results

4.2.1 Creating lines of transgenic mice for the 186 *att*B site

4.2.1.i Source of attB DNA

At the time of creating lines of transgenic mice the minimal *att*B site had not yet been fully defined, therefore the clone containing the RL fragment, pRL (Figure 3.2), shown to permit *in vitro* integrative recombination with pP'OPNeo (Figure 3.3(a), Control), was used as the source of DNA. pRL contained a 177 bp fragment of *att*B, cloned into pBS SKII+ (2.96 kb). The pRL clone was linearised with *Sca*I (located in the ampicillin gene), purified (6.6.13) and the resultant 3.14 kb fragment, containing the 186-*att*B site, used for microinjection into fertilised mouse eggs (*6.6.13.i*). The presence of extra 5' and 3' vector sequence would buffer the effect of any exonuclease activity on the fragment after microinjection and prior to random integration into the mouse genome.

4.2.1.ii PCR detection of transgenic mice

Pups (approximately 6 weeks old) were analysed by removal of a small section of tail from which genomic DNA was prepared (6.6.15) and used for PCR analysis (6.6.11). The primers 2RR1 and LHIII (6.4.7; Table 1(b)), used for the initial construction of the pRL clone, were used to detect mice transgenic for the RL fragment (BOB'). Initially 38 pups were analysed and three were found to be transgenic by PCR analysis (data not shown), giving a 7.9% frequency of transgenesis. The mice #23 (female), #29 and #31 (both male) all generated a fragment of 203 bp representative of RL, as did the positive control reaction containing pRL DNA (100 pg). The negative control (no DNA) and all other genomic DNA reactions failed to generate a band.

4.2.1.iii Breeding of transgenic mice

The three founder mice #23, #29 and #31 were bred with wild-type mice and the

F1 pups analysed as in *4.2.1.i*, the results of which are presented below:

F_0	*F1 progeny*	*Transgenics*
#23f	23.1 – 23.16	23.2, 4 and 16
#29m	29.1 – 29.10	0
#31m	31.1 – 31.26	31.3, 5, 6, 7, 10, 11, 13, 14, 17, 21, 25 and 26

No transgenic pups were identified for the founder #29; it appeared to be a mosaic giving non-germline transmission. Founder #23 gave rise to three transgenic (TG) pups giving a frequency of transmission of 19%. Founder #31 gave rise to twelve TG pups giving a frequency of transmission of 46%. We would expect 50% of the F1 progeny to be TG. Why the rate of transmission for #23 is lower than this is unknown but perhaps it too is a mosaic with only half the germcells carrying the construct. Further breeding and PCR analysis gave rise to the following:

F1 x WT	*Progeny*	*Transgenics*	*% TG*
23.2 f	23.2.1 - 23.2.6	23.2.2 and 23.2.5	33%
31.3 m	31.3.1 – 31.3.20	31.3.5, 6, 8, 9, 10, 11, 13, 14 and 17	45%
31.5 f	31.5.1 – 31.5.11	31.5.1, 3, 5, 6, 8 and 9	54%
31.7 f	31.7.1 – 31.7.7	31.7 .1, 5 and 7	43%

At this point enough TG mice had been generated for the required experiments in this section of work to commence, and no further breeding of mice took place.

4.2.2 Investigating site-specific recombination in genomic DNA *in vitro*

4.2.2.i Generating a positive control template and a PCR assay for detecting site-specific recombination in genomic DNA

In order to detect a possible 186 site-specific integration event of pP'OPNeo into a randomly placed genomic BOB' site by PCR, I first needed to generate a positive control template. This was accomplished by performing the *trans in vitro* recombination assay (*6.6.12.ii*) using pP'OPNeo, linearised (*Sca*1) pRL and

purified Int and IHF proteins (Figure 4.1). The reaction was diluted 1/10 with sterile H_2O and 1 µl used in a standard PCR reaction, using the primers 66 (Appendix, Figure 1) and 2RR1 (6.4.7, Table 1(b) and Figure 3.2) to detect the P'OB composite site generated by the reciprocal exchange of the second strands (3.2.6) which complete integration of pP'OPNeo into BOB' (pRL). Results of the PCR reactions are presented in Figure 4.2. Integration of pP'OPNeo into BOB' had occurred as evidenced by the presence of a band of 260 bp representative of P'OB in track one. Another band of slower mobility was also present. It was non-specific as it also appeared in the control reaction which had no proteins added, and hence was likely to be a result of cross-reactivity with vector sequence in either pP'OPNeo or pRL. A *trans in vitro* control reaction containing no DNA indicated, by PCR, that no contamination of equipment or solutions had occurred. The primer control reactions, using either 63 and 66 or LHIII and 2RR1, both detected the correct POP' (307 bp) and BOB' (203 bp) bands respectively.

Figure 4.1

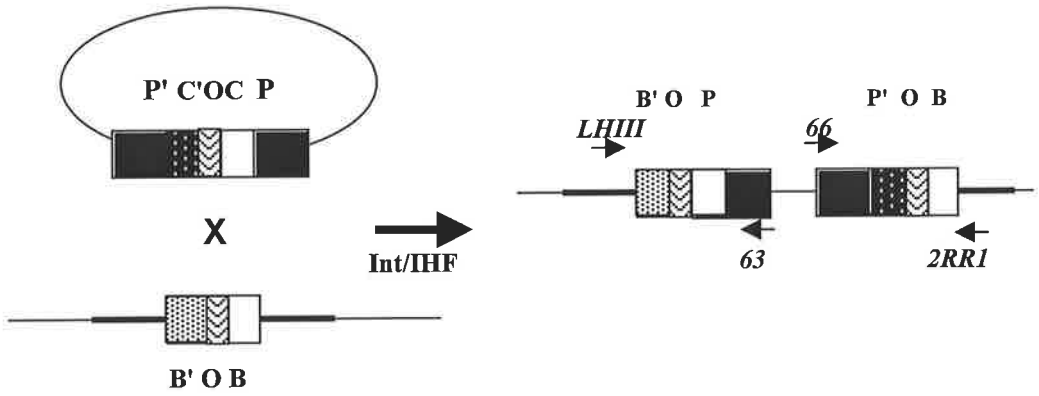


Figure 4.1 A schematic representation of the integration of pP'OPNeo into pRL (BOB'), in the presence of Int and IHF, to generate the composite sites B'OP and P'OB. The PCR primers 66 and 2RR1 detect P'OB which results from the exchange of the second strands, and the controls LHIII and 2RR1 detect B'OB, and 66 and 63 detect P'OP.

Figure 4.2

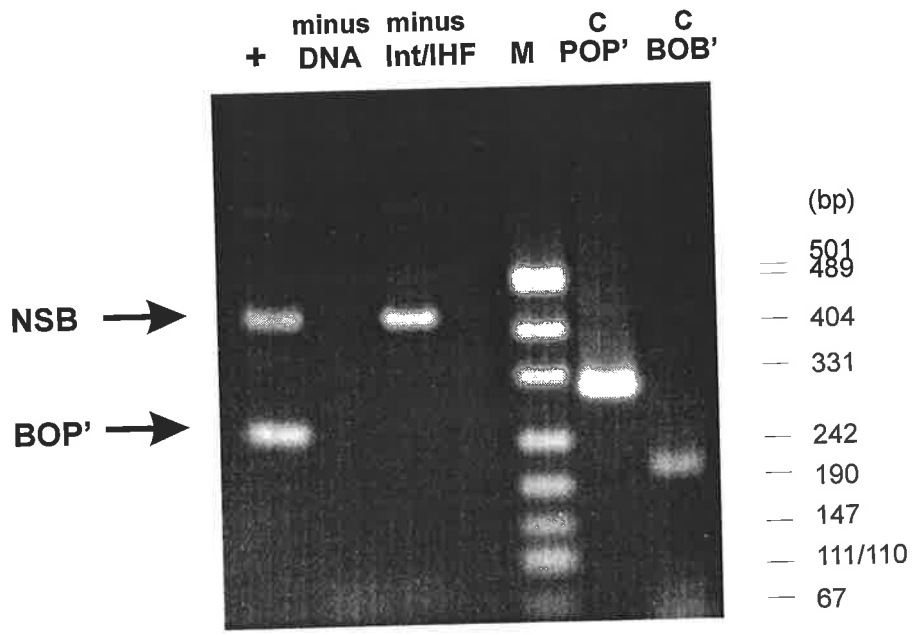


Figure 4.2 A 3% agarose-TAE gel of the PCR assay to detect the composite site BOP' (260 bp) resulting from the integration of pP'OPNeo into 186-genomic BOB' sites using the PCR primers 66 and 2RR1. A non-specific band (NSB) is also evident; the control reaction, minus Int/IHF, also gave rise to this band. The DNA markers (M) are DMW-P1 (6.4.4). Primers 63 and 66 gave rise to the POP' band (307 bp) control and LHIII and 2RR1 the BOB' band (203 bp) control.

4.2.2.ii Detection of site-specific integration into attB-containing genomic DNA in vitro

Isolated and purified genomic DNA from transgenic mouse #31.25 (*4.2.1.iii*) was used in the *trans in vitro* recombination assay (5 µl and 2 µl) using pP'OPNeo and the purified proteins Int and IHF. The genomic DNA from #31.25 had been shown (PCR) to contain at least one randomly located *att*B site. PCR using the primers 66 and 2RR1 was performed on 1 µl of the *trans* recombination assays. The data (Figure 4.3) indicated that site-specific recombination had taken place *in vitro* resulting in the integration of pP'OPNeo into *att*B site(s) located in genomic DNA for both the 2 µl and 5 µl samples. A *trans in vitro* control recombination assay using the plasmid pRL was performed and PCR showed the presence of a band specific for BOP' and a non-specific (cross-reacting) band. A *trans in vitro* negative control reaction (minus DNA) indicated that no contamination of equipment or solutions had occurred prior to PCR analysis. The non-specific band is likely to arise from cross-reacting sequences present in pP'OPNeo or the mouse genomic DNA. Later I show the non-specific band is due to sequences in pP'OPNeo (*4.2.3.ii*).

4.2.3 Investigating site-specific recombination into chromatin *in vitro*

Chromatin was prepared from nuclei isolated from livers of transgenic mice (6.6.16). The purified chromatin (186-*att*B) was assessed for its ability to undergo site-specific recombination with intasomes (pP'OPNeo + Int + IHF) in the *trans in vitro* recombination assay by PCR designed to detect the composite sites B'OP and P'OB.

4.2.3.i Microccocal nuclease assessment of chromatin quality

In order to assess the quality of the isolated chromatin, *Microccocal nuclease* digestion was performed (6.6.17). The nucleosomal ladder generated (Figure 4.4) showed that much of the chromatin was intact, as indicated by a strong band at 200-210 bp representative of the presence of whole nucleosomes, and at the higher

Figure 4.3

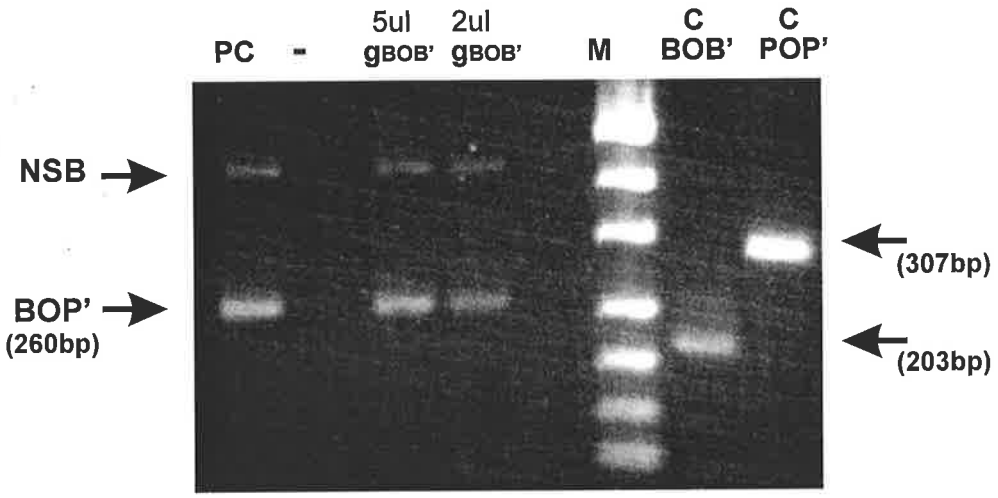


Figure 4.3 A 3% agarose-TAE gel showing the PCR results for the detection of integration of pP'OPNeo into genomically placed *att*B sites *in vitro* (primers: 66/2RR1). PCR using the positive control template (PC) showed the BOP' band and the non-specific band (NSB). PCR performed on the recombination assay using either 5ul or 2ul of genomic DNA from the TG mouse #31.25 both showed the presence of the BOP' band indicating integrative recombination into genomically located *att*B sites of pP'OPNeo. The DNA markers (M) are DMW-P1. The control reactions for BOB' (primers: 2RR1/LHIII) and POP' (primers: 63/66) are also shown.

Figure 4.4

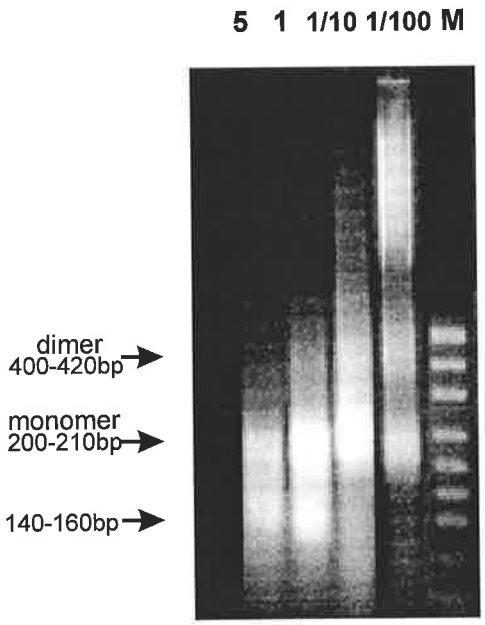


Figure 4.4 The quality of liver chromatin preparation was assessed using micrococcal nuclease digestion (6.6.17) at amounts of 5,1,1/10 and 1/100, where 1 is equal to 0.5U/ul. Samples were analysed on a 3% agarose-TAE gel. The markers are DMW-P1. The locations of presumed nucleosome dimer, monomer and core nucleosome (140-160bp) are indicated.

levels of *Microccocal nuclease* a faster migrating band of approximately 140-160 bp indicating areas of chromatin that have perhaps lost linker histones and consist only of the core nucleosome. The prepared chromatin was considered suitable for *in vitro* recombination assays.

4.2.3.ii In vitro 186 site-specific recombination occurs in chromatin

The *trans in vitro* recombination assay was conducted using 30 µl of prepared intasome mix (6.7.4) and 10 µl of chromatin diluted 1/3, 1/10, 1/100 and 1/200 and incubated at 37°C for 60 minutes. Without purification of the recombination reaction, I initially tested whether PCR could detect the 186-*att*B sites known to be present in the chromatin and if the presence of histone proteins was inhibitory to the PCR reaction. Using the primers LHIII and 2RR1 PCR was conducted, and the results (Figure 4.5(a)) indicated that, at a 1/3 dilution of the recombination reaction, the chromatin was inhibitory to PCR. The further dilutions of 1/10, 1/100 and 1/200 adequately overcame any inhibitory effect, permitting the detection of *att*B. The PCR controls consisted of a positive using the template generated in *4.2.2.i*, a spiked positive control consisting of the 1/100 chromatin dilution and the template generated in *4.2.2.i*, and a negative control using no chromatin. All gave the expected results.

PCR was performed on the same samples using primers 66 and 2RR1 to detect P'OB which results from reciprocal exchange of the second strands. The results (Figure 4.5(b)) indicated that 186 site-specific integrative recombination had occurred *in vitro* using chromatin as a source of *att*B DNA. The PCR controls consisted of a positive using the template generated in *4.2.2.i*, a spiked positive control consisting of the 1/100 chromatin dilution and the template generated in *4.2.2.i*, and a negative control reaction using no chromatin. Using primers 66 and 2RR1 the negative control showed that in the absence of chromatin (*att*B) no site-specific recombination could be detected as expected. Again a non-specific band (*4.2.2.i*) was evident and therefore it could be concluded it must arise from cross-reacting sequence present in the pP'OPNeo plasmid. I could detect site-specific

Figure 4.5

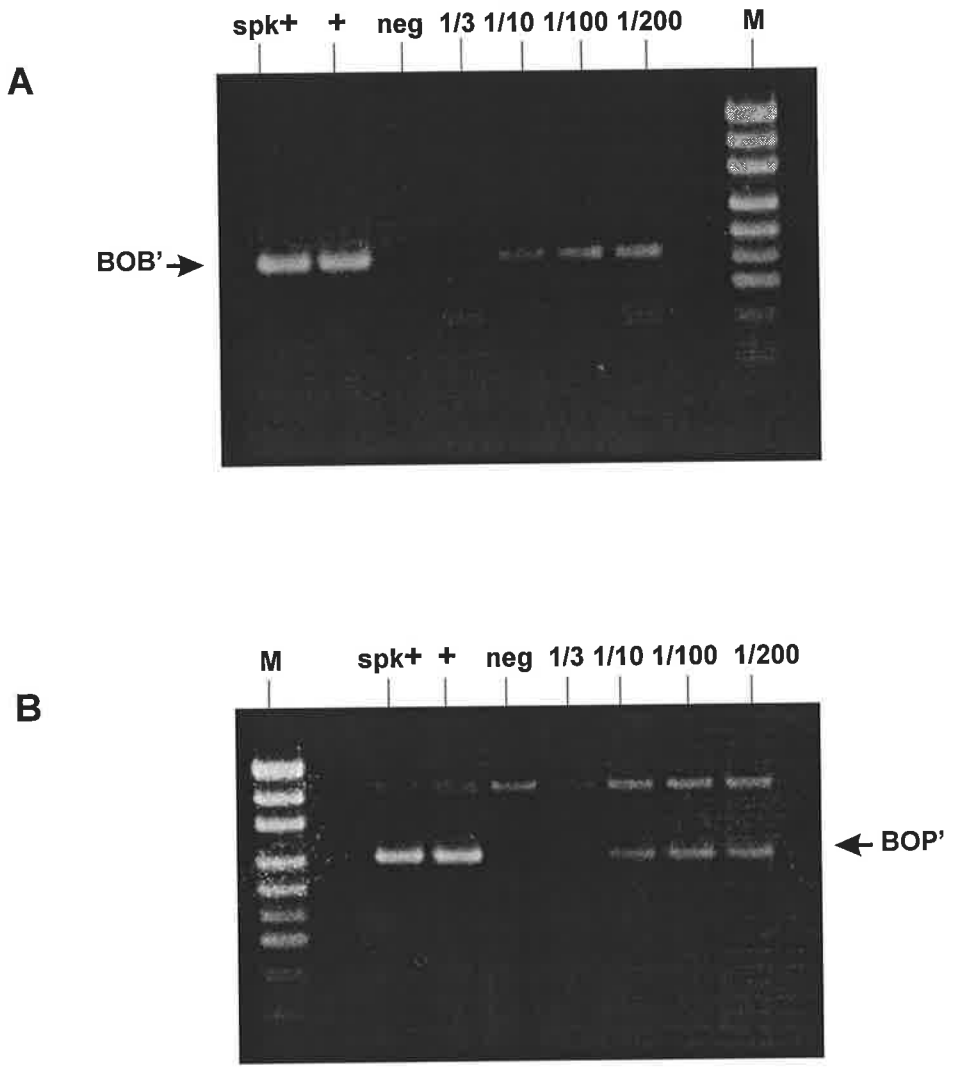


Figure 4.5 A 3% agarose-TAE gel showing PCR analysis of chromatin diluted 1/3, 1/10, 1/100 and 1/200 and used in the *trans in vitro* recombination assay using A) primers LHIII/2RR1 to detect 186-*att*B sites and B) using primers 66/2RR1 to detect BOP'. The controls consisted of a positive control using the template generated in *4.2.2.i* (+), a spiked positive control containing the 1/100 chromatin dilution (spk+), and a negative control containing no chromatin. The markers (M) are DMW-P1.

recombination in both the positive and spiked positive control reactions. In the samples using chromatin as the only source of *att*B DNA, site-specific recombination was detectable in the diluted chromatin samples 1/10, 1/100 and 1/200. The reaction containing a 1/3 dilution of chromatin was inhibitory to both the PCR reaction, with a very low level of the non-specific band detectable, and the initial recombination reaction, with no P'OB band detectable.

4.2.4 Investigating 186 site-specific recombination in transgenic fertilised mouse eggs

The final part of my project involved initial investigations into whether 186 site-specific recombination could be demonstrated to occur in fertilised mouse eggs containing randomly located BOB' sites. The approach was to microinject (performed by S. MacIlfattrick, BresaGen) pre-formed 186 intasomes into the fertilised mouse eggs, place them back into pseudopregnant female mice (performed by Dr. Zhong Tao Du, BresaGen) and analyse the resultant progeny, via PCR and Southern analysis, for precise integration of pP'OPNeo. Ideally for the microinjection process, which took place for 2-3 hours each day over a 3-4 week period, aliquots of the pre-formed intasomes would be required to maintain their recombinational ability after being frozen and then kept on ice. Hence the stability of pre-formed and frozen intasomes was investigated.

4.2.4.i Intasome stability

For microinjection into fertilised mouse eggs, intasomes were prepared in 1 x PBS (phosphate-buffered saline). These contained approximately a 1.5-fold excess of purified Int and IHF proteins based on *trans in vitro* assay data, whereby 300 ng of pP'OPNeo in 50 µl required 57.6 ng of Int and 2 ng of IHF proteins (minimally) for maximal integrative recombination. Therefore, for stability analysis, 250 µl of pre-formed 186 intasomes (10 ng/µl pP'OPNeo, 768 ng Int, 26.8 ng IHF) were prepared (6.7.4) in both PBS and recombination buffer (10 mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 100 mM KCl). After 15 minutes incubation at 4°C, a 50 µl aliquot was

removed from each, placed in liquid nitrogen and stored at –80°C for 2 hours, while a second 10 µl aliquot was removed from each and added to 40 µl of a *trans in vitro* recombination assay, containing kinased 30mer BOB', and incubated at 37°C for 60 min prior to analysis (*6.6.12.ii*). The remainder of the two pre-formed intasome mixes were further incubated at 4°C for 60 min and 120 min and 10 µl aliquots were removed for analysis. After 2 hours at –80°C the samples were thawed on ice for 30 min and 10 µl aliquots were removed and assayed as above.

Results (Figure 4.6) indicated that pre-formed 186 intasomes were stable when frozen and thawed, and maintained their recombinational competency after 1 and 2 hours at 4°C. Further, at 15 min the level of recombination achieved was lower than all later time points indicating that maximal intasome formation requires greater than 15 minute incubation at 4°C to allow protein binding. Furthermore, recombinational activity was identical for intasomes prepared in either PBS or recombination buffer, indicating that PBS does not have a detrimental affect on intasome activity.

4.2.4.ii Intasome toxicity test

Three sets of intasome mixes were prepared (6.7.4) containing either 2.5, 5 or 10 ng/µl pP'OPNeo DNA. A number of fertilised mouse eggs were microinjected with each mix for toxicity analysis (performed by S. MacIlfattrick, BresaGen). The controls consisted of PBS-injected and non injected fertilised eggs. The toxicity test results are presented below:

	1 cell injected	*Development to* *2 cell*	*blastocyst*	*%*
10 ng/µl	25	23	18	72%
5 ng/µl	25	25	14	56%
2.5 ng/µl	29	28	23	79%
PBS injected	26	24	17	65%
non injected	22	22	22	100%

Figure 4.6

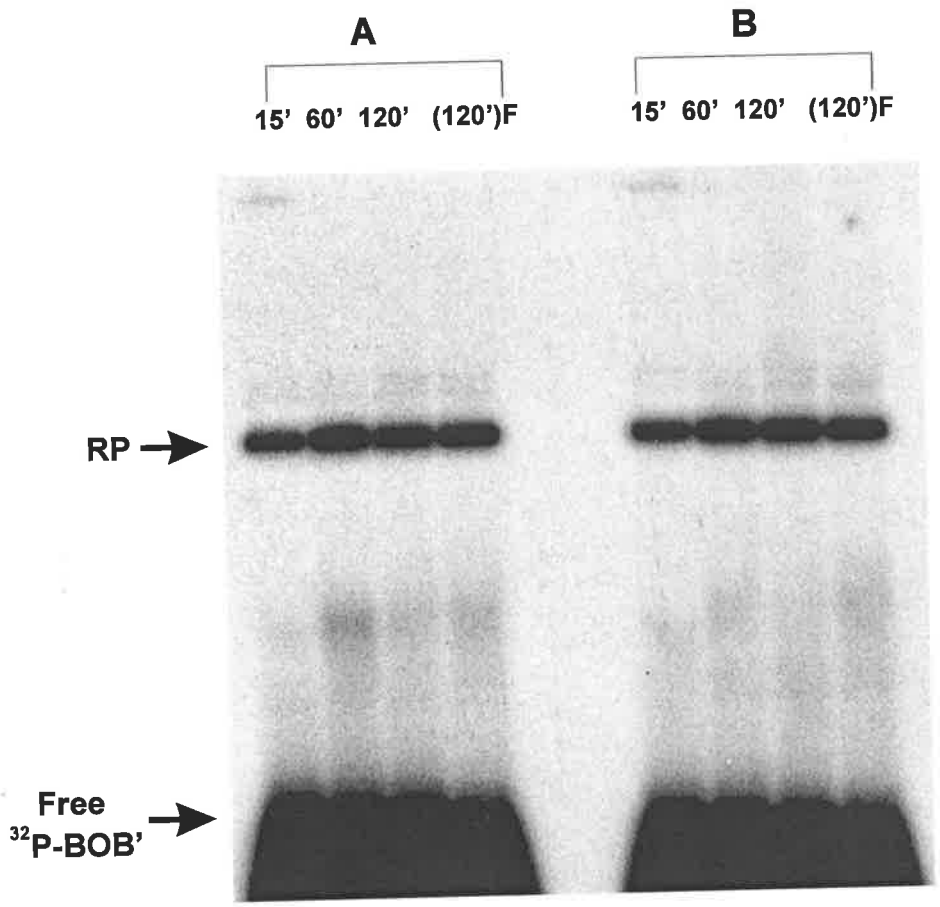


Figure 4.6 Pre-made intasomes were analysed for stability in A) PBS and B) recombination buffer. Activity was analysed after 15, 60 and 120 min at 4°C using the *trans in vitro* recombination assay. Furthermore, an aliquot removed at 15' (4°C), frozen (liquid N_2) and stored at -80°C for 2 hours [(120')F] maintained maximal recombination ability when thawed and analysed.

The data indicated that no component, in all three intasome mixes, was toxic to the fertilised eggs (>50% of injected eggs developed to blastocyst stage). To maximise the possibility of insertion, intasomes with the highest DNA content (10 ng/μl) were used for the following microinjection studies.

4.2.4.iii Analysis of pups generated from microinjection of intasomes into transgenic fertilised mouse eggs

Prior to injection of the intasomes into fertilised mouse eggs, an aliquot of the frozen 10 ng/μl intasomes was analysed for recombinational activity as described previously (*4.2.4.i*). Results indicated that the intasomes maintained their ability to generate the correct recombination product, with no observed decrease in the level of product generated (data not shown).

Eggs were generated from the crossing of TG males (heterozygotes) to WT females, and after injection, placed back into pseudopregnant WT females. Hence, 50% of the eggs injected would contain 186-*att*B sites. A total of 119 pups were generated from the microinjection of 186 intasomes into transgenic fertilised mouse eggs.

PCR to detect BOB' transgenic mice

Initially PCR was used to identify mice transgenic for BOB', as described previously in section *4.2.1.ii.* A total of 62 positive BOB' transgenic mice (transmission frequency = 52%) were identified. Since the 62 mice, transgenic for BOB', all arose from the breeding of three founder mice it was reasoned it was highly unlikely that any would contain a single-copy, randomly inserted BOB' such that a 186 site-specific insertion of pP'OPNeo would eliminate detection of the BOB' site (using microinjection such single-copy insertion events are rare).

An example of the PCR results obtained is given in Figure 4.7. The controls consisted of a plasmid control, a genomic DNA (BOB'-minus) spiked positive control and a no DNA negative control. All controls gave the expected results.

Figure 4.7

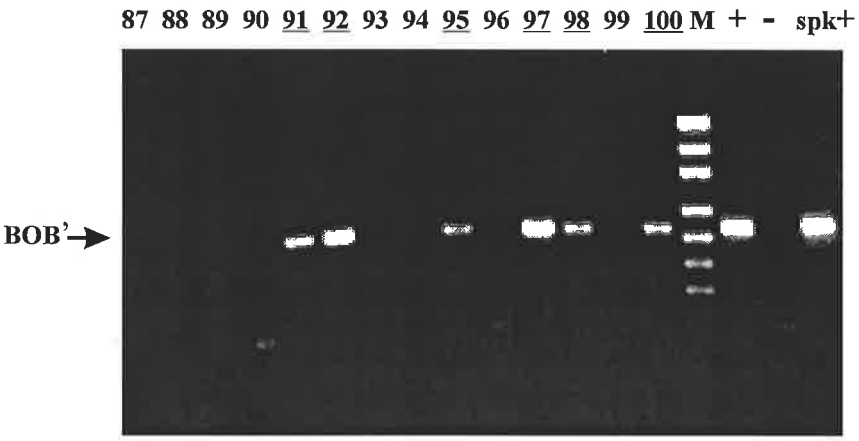


Figure 4.7 An example of the PCR results obtained for the identification of BOB' transgenic mice (underlined) using the primers LHIII and 2RR1. The controls were the plasmid control, pRL (+), a no DNA negative control (-) and a genomic DNA spiked positive plasmid control (spk+).

Listed below are the identified BOB' transgenic mice:

Mice: #1, 3, 4, 6, 7, 8, 9, 11, 17, 18, 19, 21, 23, 25, 27, 29, 31, 33, 35, 36, 37, 43, 45, 46, 47, 48, 50, 53, 54, 55, 59, 60, 63, 64, 65, 67, 70, 71, 72, 75, 76, 77, 78, 80, 81, 85, 91, 92, 95, 97, 98, 100, 101, 103, 105, 107, 108, 109, 112, 114, 116 and 117.

PCR to detect 186-mediated integration events

PCR analysis could have been conducted on just the 62 transgenic BOB' mice however, I performed PCR on all 119 genomic DNA tail samples to firstly, provide internal BOB'-minus genomic DNA negative controls and secondly, because of the small possibility of there being multiple site-specific insertions which also would have eliminated BOB' detection by PCR.

If 186 recombination had occurred there would be, at minimum, one copy of correctly integrated pP'OPNeo in the mouse genome. Initially it was important to ascertain that the PCR system (using primers 66/2RR1) could detect a single-copy integrant. In *4.2.2.ii* I showed detection of site-specific integration into *att*B-containing genomic DNA *in vitro*. The actual number of copies of BOP' that were detected in this system was significantly higher due to firstly, the amount of genomic DNA used and secondly, it was likely that integration occurred multiple times within one copy of the mouse genome.

The mouse genomic DNA tail preparations routinely contained approximately 100 ng/µl DNA. Given that one copy of the mouse genome approximates 3.3 pg of DNA, 100 ng would be equivalent to 3.3×10^4 copies. The positive control template (8.4 kb=pRL+pP'OPNeo) generated in *4.2.2.i* was calculated to be 5 ng/µl (UV spectroscopy). Given that the Mw of one basepair is 660, then 1 µg of 8.4 kb positive control template would equal 1.8×10^{-13} moles/µg or 1.11×10^{11} copies/µg (where Avogadro's number=6.023×10^{23}). Hence, 5 ng of positive control template equals 5.5×10^8 copies. The positive control template was serially

diluted down to $2.8x10^4$ copies per µl, and tested in the PCR reaction, which was also spiked with 1 µl of genomic DNA (BOB' negative). The results are presented in Figure 4.8.

I surmised I would be able to detect a single-copy integration event by PCR analysis, since the correct PCR product was detectable down to $2.8x10^4$ copies (Lane 8). Furthermore, in these results the non-specific band, evident in earlier experiments, was absent. This may be attributed to either a new batch of primers or Taq DNA polymerase, both of which had changed since performing the same assays earlier, under identical PCR conditions.

Results showed no positives for BOP' were detected (data not shown). The controls, a spiked positive control, positive control and a no DNA negative control all gave the correct results. The BOP' band was evident in the two positive controls, spiked and unspiked, while the negative control showed no band, indicating no contamination of samples had occurred. Again the presence of the non-specific band was eliminated.

Southern analysis

Although the PCR data strongly indicated no positives for 186-mediated recombination in fertilised mouse eggs (in this experiment), further confirmation was made by Southern analysis. Only the 62 mice identified as transgenic for BOB' were analysed. Analysis was conducted using 3 µg of genomic DNA which was digested with *Hind*III or *Eco*RI restriction enzymes. 186 site-specific integrants of pP'OPNeo would give a defined fragment of 5.1 kb *Eco*RI or 2.0 kb *Hind*III (Figure 4.9) when analysed with a neomycin gene specific probe (*Xma*III-*Nco*I fragment of *neo* [526 bp]; courtesy of R. Faast, BresaGen, Adelaide). In contrast, single random integrants of pP'OPNeo would give rise to a randomly sized fragment, depending on where the next restriction site is in the mouse genome for the end repeat, whereas multiple insertions of pP'OPNeo would give one randomly sized fragment and a band of 5.25 kb representative of pP'OPNeo head to tail repeats.

Figure 4.8

Figure 4.8 PCR using the positive control template serially diluted to give the copies/reaction:**1.** $1.1x10^7$, **2.** $5.5x10^6$, **3.** $1.1x10^6$, **4.** $5.5x10^5$, **5.** $2.2x10^5$ **6.** $1.1x10^5$, **7.** $5.5x10^4$ **8.** $2.8x10^4$. The markers (M) are DMW-P1. The negative control (no DNA) showed no band.

Figure 4.9

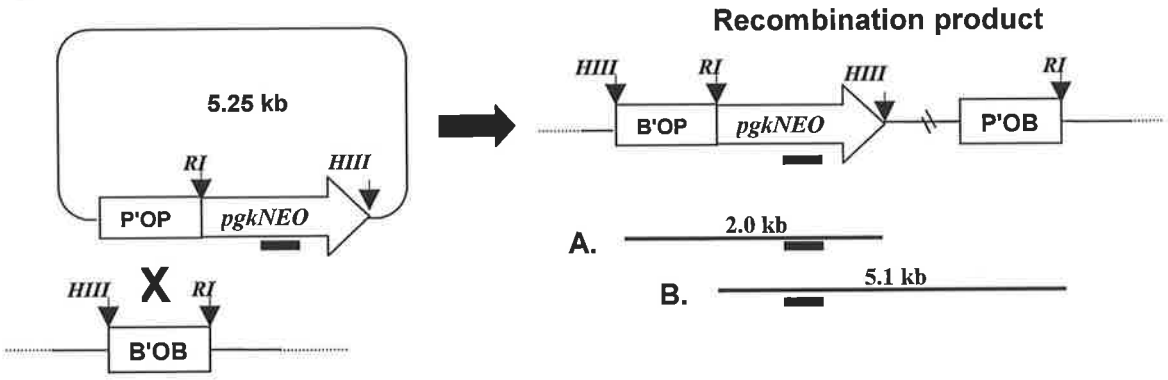


Figure 4.9 Recombination between pP'OPNeo and a genomically placed BOB' results in the formation of the illustrated recombination product. For 186 mediated site-specific insertion Southern analysis of either *Hind*III and *Eco*RI digested DNA, using a *neo* specific probe (▬) would detect a 2.0 kb (A.) or 5.1 kb (B.) fragment respectively. Random integration of pP'OPNeo would detect a 5.25 kb fragment for head to tail multiple copies, or a random sized fragment for a single copy.

Southern results, using *Hind*III-digested DNA (Figure 4.10) showed no 186 site-specific integration events. The control lane contained approximately 0.5 pg of *Hind*III digested positive recombination control template, pRL+pP'OPNeo (generated in *4.2.2.i)*. Two bands are present in the positive control track: the slower migrating band represents pP'OPNeo/*Hind*III (5.25 kb), equivalent to head to tail arrays of randomly integrated pP'OPNeo, and a faster migrating band, representative of the correct 186 Recombination product/ *Hind*III (2.0 kb). In mouse #85, random integration of pP'OPNeo was observed, as evidenced by the presence of a strong band at 5.25 kb (head to tail repeats) and a larger band of approximately 6.5 kb, characteristic of a randomly sized *Hind*III fragment generated by the end repeat of the randomly integrated pP'OPNeo, the size of which depends on the position of the next genomically located *Hind*III site.

Southern analysis of *Eco*RI-digested #85 mouse genomic DNA further confirmed that this mouse had a random integration of pP'OPNeo (data not shown), as again a strong band at 5.25 kb was present, and a larger band of approximately 8 kb, characteristic of a randomly sized *Eco*RI fragment generated by the end repeat of

Figure 4.10

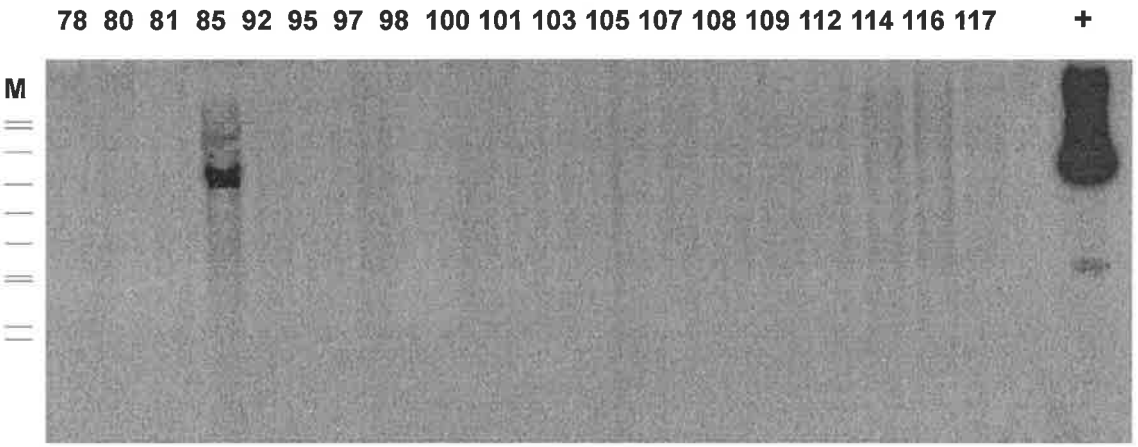


Figure 4.10 Southern analysis of genomic DNA digested with *Hind*III from a subset of the identified BOB' transgenic mice, where the oocyte was microinjected with 186-intasomes, and when analysed with a neomycin gene specific probe. The control template (*4.2.2.i*) has two bands, the slower of which represents pP'OPNeo/*Hind*III (5.25 kb) equivalent to the random, head to tail insertion of pP'OPNeo, and a faster migrating band representing the correct *Hind*III/recombination product band (2.0 kb). Mouse #85 has a random insertion of pP'OPNeo (5.25 kb) and a slower migrating band at approximately 6.5 kb, representing a randomly sized *Hind*III fragment generated by the end repeat of pP'OPNeo. The markers (M) are DMW-S1.

the randomly integrated pP'OPNeo, the size of which depends on the position of the next genomically located *EcoR*I site.

4.3 Discussion

Promisingly, the *trans in vitro* recombination data showed integration of pP'OPNeo firstly into naked genomic DNA and secondly into chromatin. The integration into chromatin was a significant step forward for my project as it strongly indicated that 186 intasomes could access nucleosomal structure. However, it is possible that the only integration that occurred was because there was, unavoidably, low levels of DNA that had lost enough histones, or even whole nucleosomes, to allow integration to take place. Nevertheless, the results were encouraging.

Microinjection studies of intasomes into fertilised mouse eggs were pursued. To detect site-specific integrants, the two methods of PCR and Southern analysis were undertaken. No site-specific integrants were detected, although random integration of pP'OPNeo was found (1/62=1.6%). The frequency of random integrants, via microinjection, was significantly lower than that routinely obtained (5-25%) by this laboratory, using other constructs at the same concentration (10 ng/μl). This result may indicate that the proteins, Int and IHF, when bound to pP'OPNeo forming an intasome, may help to prevent the random integration of DNA by sequestering the DNA, sterically precluding it from integrating. If the 186 recombination system can be shown to work in mammals, this finding, if proven correct, would be highly advantageous since random integrants would be minimised.

4.4 Conclusions

The present study has been a preliminary examination of 186-mediated integration into mammalian genomic DNA, *in vitro* and *in vivo*. Encouraged by positive *in vitro* results, microinjection studies were undertaken, combining the accumulated

knowledge gained from the work presented in each chapter of this thesis. Although the final outcome has been a negative result on this occasion, it does not necessarily mean there exists a fundamental problem with the 186 system in a eukaryotic environment *per se*. The system may require further fine tuning of the conditions used, such as increasing the amount of intasomes delivered. The problem could simply be one of frequency, investigations using greater numbers of injected eggs, or microinjection into cells could be pursued in the future. Thereafter, a positive result would provide a starting point to enable further fine tuning of the conditions used.

CHAPTER 5

Final Discussion

Chapter 5

Final Discussion

5.1 Transgenic technology: Where are we today?

The development of transgenic technology, whereby genes, or modifications, can be stably introduced into the germline of experimental mammals, now permits scientists to assess the consequences of these genes (modifications) in the developing mammal. The development of techniques that allow modification of the mammalian genome have made a considerable contribution to many areas of both basic and applied biological science. Despite this, for efficient livestock production, challenges remain in two key areas of transgenic technology, namely gene regulation and precise gene transfer or modification.

The technique of microinjection is so far the only routine method of gene transfer in farm animals. In contrast to the traditional "gain of function" mutations typically created by microinjection of the gene of interest into the one-celled zygote, gene targeting via homologous recombination in pluripotent ES cells allows one to modify precisely the gene of interest. To date the lack of pluripotent embryonic stem cell lines in farm animal species hampers the development of this area of work. Possible alternative methods for generating transgenic livestock involve the use of homologous recombination in fertilised eggs, although the frequency of success is currently too low for consideration as a routine method. The Cre/*lox* and FLP/FRT recombinase systems have been shown to operate in eukaryotic systems and importantly, via microinjection into fertilised mouse eggs (excision). The drawback with these systems is their non-directionality, one protein directing both excision and integration.

The development of successful nuclear transfer techniques provides a means of manipulating a farm animal's genome. This can be accomplished by first performing homologous recombination in a somatic cell line, followed by nuclear

transfer and reprogramming of the nucleus. A drawback with using homologous recombination to insert genes into livestock species is the arduous task of constructing a targeting vector containing expansive regions of isogenic homology.

As already discussed (1.9.1), the target sites for Cre, FLP, λ, P2 and 186 recombinase systems could be positioned into a somatic cell line via homologous recombination and thereafter, the genomically placed target site could be used to insert various transgenes using the appropriate site-specific recombination system. Further, the recombinase systems may present an alternative path to homologous recombination for the initial introduction of the target site, and that is by using closely matching naturally occurring target sites in combination with an altered integrase, one could bypass the homologous recombination step altogether. If any of the λ, P2 or 186 recombinase systems can be shown to operate in a mammalian environment it is likely the integration frequencies achieved would be significantly higher than either Cre or FLP due to their extreme directionality.

5.2 Targeted transgenesis and the 186 site-specific recombinase system

The initial aims of this project were two-fold: to characterise the 186 integration reaction and to conduct preliminary investigations into the possible use of such a system for targeted transgenesis. The development of a directional recombinase system for use in higher eukaryotes for precise, single copy placement of transgenes, or inactivation of a gene, would be beneficial to both pure and applied research. Prior to this study, to the best of my knowledge, there have been no reports of using a directional recombination system, such as λ, P2 or 186 for the *in vivo* site-specific integration of DNA into a mammalian genome.

In this study I have examined *in vitro* the requirements for 186 integrative site-specific recombination, firstly using cell-free extracts containing a source of 186 Int and IHF proteins, and later using purified components. Furthermore, I have

characterised the bacterial attachment site, *att*B, located its exact position, sites of cross-over, order of strand exchange, and with preliminary mutational analysis, identified bases important for recombination. Knowing the exact protein requirements and some important characteristics of the recombination reaction, the data from this work permitted me to pursue genomic DNA recombination experiments *in vitro,* using naked genomic DNA and chromatin, and *in vivo* using fertilised mouse eggs. Encouragingly, I was able to demonstrate 186-mediated integration into both naked genomic DNA and chromatin. In contrast, the first attempt at microinjecting active intasomes into fertilised mouse eggs, harbouring an *att*B site, did not show a specific 186 integration event. However, it is early days, further experiments are poised to begin and there remains much cause for optimism.

5.2.1 A 186-mediated "knockout" of the pig $\alpha_{1\text{-}3}$-galactosyltransferase gene using closely matching, naturally occurring 186-*att*B sites.

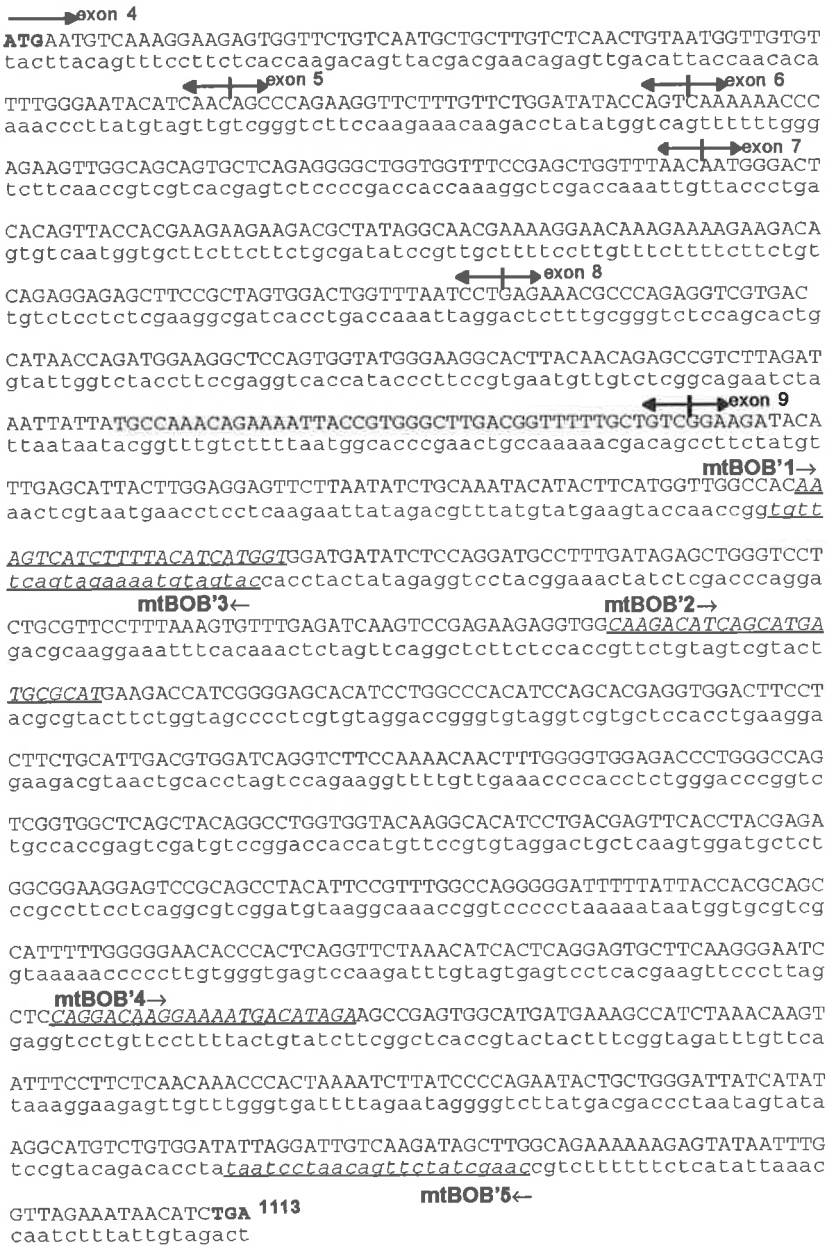
To demonstrate the feasibility of using a directional recombination system to insert DNA into a gene to modify, disrupt, or insert an entirely new function, using closely matching, naturally occurring phage target sites, I have performed a preliminary homology search of the porcine $\alpha_{1\text{-}3}$GalT coding sequence for mammalian genomic target sites akin to the bonafide 186-*att*B site which could enable insertion of DNA, using 186-intasomes, to disrupt the gene's function.

The coding sequence and structure of the porcine $\alpha_{1\text{-}3}$GalT gene is known (Dabkowski *et al.*, 1993). There are nine exons, the first three are non-coding, while exons four through nine encode discrete structural regions of the protein. Importantly, exon 9 encodes the catalytic domain which will be the major target for disruption (Figure 5.1 A and B). In mice, which share 76% amino acid identity to the pig gene (Strahan *et al.*, 1995), it has already been demonstrated that the targeted disruption of the catalytic domain significantly reduces the hyperacute rejection response (Tearle *et al.*, 1996; van Denderen *et al.*, 1997).

Figure 5.1

Porcine α-GalT coding sequence

A.



B.

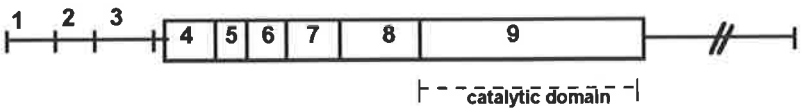


Figure 5.1 A. The porcine α-GalT coding sequence. The locations of exons 4 to 9 are indicated. Mammalian target site sequences mtBOB'1-5 are shown in the boxes. **B.** A schematic diagram showing the structure of the porcine α-GalT gene. Exons 1 to 3 are non-coding, while exons 4 to 9 encode discrete structual regions of the protein. The catalytic domain is encoded by exon 9.

A homology search of the pig $\alpha_{1\text{-}3}$GalT coding sequence and the eleven identified (3.2.4) 186 Int determinants (**CAAG**nn**A** nnnnn**a**n **T**nn**C**n**T**G**T**) revealed three sequences closely matching the native *att*B site, both contained in the 5' region of exon 9, one of which is on the complementary strand. Furthermore, I identified two sequences with a high number of 186 Int determinants, both contained in the 3' region, of which one is again on the complementary strand (below and Figure 5.1);

		N° Determinants	+6(overlap)	=specificity (>4^{16})	
186-BOB'	5' **CAAG**CG**A** ttatg**a**g **T**CG**C**C**T**G**C**T	**11**	**+6**		
mtBOB'1	5' **AAAG**TC**A** tctttta CATCA**T**GG**T**	8	+7	=15	(4^{15})
mtBOB'2	5' **CAAG**AC**A** tcagc**a**t GATGCGCA**T**	7	+6	=13	(4^{13})
mtBOB'3 (complementary strand)	5' **CATG**ATG taaaaga **T**GACTTTG**T**	7	+7	=14	(4^{14})
mtBOB'4	5' **CAGG**AC**A** aggaa**a**a **T**GACA**T**AGA	9	+6	=15	(4^{15})
mtBOB'5 (complementary strand)	5' **CAAG**CT**A** tcttg**a**c AATCC**T**AA**T**	9	+6	=15	(4^{15})

The sequences mtBOB'1 and mtBOB'2 would prove most useful as integration target sites as both are contained close to the start of exon 9 encoding the catalytic region of the enzyme α-$_{1\text{-}3}$GalT. An insertion into either of these two sites would disrupt enzyme functionality more favourably than insertion into mtBOB'3. The mammalian target sites mtBOB'4 and mtBOB'5 illustrate two sites that closely match the native phage site. Although neither would be useful targets to knock-out the $\alpha_{1\text{-}3}$GalT gene activity, they would be useful insertion sites for new genes as only two rounds of mutagenesis of 186 Int would be required to evolve Int recognition toward the new site. We know that any disruption of the $\alpha_{1\text{-}3}$GalT gene would not be detrimental to the animal.

Although much emphasis has been placed on selecting sites which closely match the Int determinants of the native phage site some caution may need to be exercised

if altering bases in the other positions. The report by Dorgai *et al.*, (1998) showed that the specificity difference between λ and HK022 can be attributed to the absence of specific bases in contrast to just the presence of integrase specific determinants. Therefore, alteration of bases in positions other than the identified Int determinants may restrict recombination. One can imagine docking of the Int protein onto the core sites to be a two-step meshing mechanism. Firstly the Int determinants are sterically aligned and secondly, as the protein nestles into position, further meshing to the other bases occurs. In this scenario a base change in a position other than a determinant could prevent correct second-step meshing and disrupt Int function. Hence the importance of conducting more extensive mutagenesis of the 186-*att*B site at all positions with all three base pair changes.

5.3 Future work

The results presented in this thesis provide a strong base from which further work can be conducted into the development of an alternative transgenic technology, namely targeted insertion using directional recombination systems. The primary question which needs to be addressed is one of frequency, that is, was the lack of a positive integration event in fertilised mouse eggs a frequency problem or a failure of the system to operate in mammals *per se*? Therefore, initially we would wish to examine a greater number of cells by using a computer-automated microinjection system (now available during the course of this work) with cells, either COS or mouse ES cells given their large and accessible nuclei, in combination with selection for neomycin resistant colonies. This would isolate any insertion events of pP'OPNeo, both random and 186-mediated events, then PCR and Southern analysis would be employed to identify any specific 186 integration events. Furthermore, modification of the proteins via fluorescent tagging could be to used to examine specific problem areas such as intasome stability in the nucleus. Given the range of fluorescent tags available today it will be possible to tag all three proteins with different identifiable markers so that the fate of each protein can be individually assessed.

5.4 Concluding Remarks

The development of a proficient system for genetic modification of farm animals (and human somatic cells) is of importance for agriculture, biotechnology and human medicine. In this study I have characterised the 186 site-specific integration reaction *in vitro* with the expectation that we could exploit its use in a mammalian system, to insert genes, or modify existing ones, to provide a much needed alternative method for the production of transgenic livestock species. On a first attempt we have not shown the 186 system to work *in vivo*, however further investigations are on going, and in the near future we hope to show a directional recombinase-mediated integration event in a mammalian genome *in vivo,* prior to pursuing evolution of the integrase to recognise natural target sites, closely matching the native phage site.

CHAPTER 6

Materials and Methods

Chapter 6

Materials and Methods

6.1 Abbreviations

A_{xxx}	absorbance at xxx nm
APS	ammonium persulphate
BCIG	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
bisacrylamide	N,N'-methylene-bisacrylamide
bp	base pairs
BSA	bovine serum albumin
CIP	calf intestinal phosphatase
ddNTP	dideoxynucleotide triphosphate
DMF	dimethylformamide
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
dNTPαs	α-phoshorothioate nucleotide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
IPTG	isopropyl-β-D-thio-galactopyranoside
Kb	kilobases
kD	kilodaltons
min	minutes
PEG	polyethylene glycol
PBS	phosphate buffered saline
RNase	ribonuclease
r.p.m.	revolutions per minute
SDS	sodium dodecyl sulphate
TEMED	N,N,N',N'-tetramethylethylenediamine

UV	ultraviolet light
XXmer	oligonucleotide of length xx bases

6.2 Bacterial Strains

BL21: F^- *ompT* r_B^- m_B^- *lon*$^-$ (Studier *et al.*, 1990).

BL21(λDE3): λDE3 lysogen of BL21. For expression of pET clones λDE3 carries a copy of the T7 RNA polymerase gene expressed from the plac UV5 promoter (Studier et al., 1990).

C600 Δ*himA*: F- *thr*$^-$1 *leu*B6 *thi*$^{-1}$ *lac*Y1 Δ*him*A::*cat*. IHF^- strain for use in recombination assays (M, Reed., Ph.D. thesis)

DH5α: F^- endA1 hsdR17($r_k^-m_k^+$) *sup*E44 *thi*$^{-1}$ *rec*A1 *gyr*A($Na1^R$) *rel*A1 Δ(*lac*ZYA-*arg*F) U169. Used for routine cloning (Hanahan, 1983).

MM294: F^- *sup*E44 hsdR17 endA1 *thi*$^{-1}$ λ- (ATCC 33625)

6.3 Plasmids

pBluescript: General cloning and double-stranded sequencing vector (Stratagene).

pEC611: A derivative of pKT52. PEC611 is a high copy-number expression vector containing the powerful p*trc* promoter (Brosius *et al.*, 1985) with a modified Shine-Dalgarno sequence (AGGAGG), and a spacer region with a unique *Nde*I site adjacent to the ATG start codon, TAACATATG (T. Brumby, unpublished data). Strong transcriptional terminators (T_1T_2) from the *rrnB* ribosomal RNA operon (Brosius *et al.*, 1981) are necessary to prevent transcription from p*trc* extending around the plasmid.

pACYC($lacI^q$): Kanamycin resistant pACYC177 derivative containing a 1.23kb *Sal*I fragment carrying *lac*I (*lac* repressor) expressed from the I^q promoter (Lanzer & Bunjard, 1988).

pFLIP(X): Used in the recombination assays (Chapter 1). Contains 186 *att*P and *att*B oriented such that site-specific recombination between the two sites inverts the DNA fragment between them.

pFLIP(I): The "integrated" form of pFLIP(X). It is the result of exposing pFLIP(X) to 186 Int and *E. coli* IHF and contains *att*L and *att*R in an inverted orientation (M. Reed, Ph.D. thesis).

pIHF: *E. coli* IHF expression plasmid created using pEC611 (this work, Chapter 2).

pMRR18: 186 Int expression plasmid created using pET3a (Studier *et al.*, 1990) which carries a T7 promoter and ribosome binding site used for expression of proteins from inserted DNA fragments. (M. Reed, Ph.D. thesis).

pPgkP'OPNeo: pBS KSII+ vector containing the 307 bp *att*P site and the Neomycin resistance gene under the constitutive control of the murine phosphoglycerate kinase promoter and polyadenylation cassette.

pUHE24-2: Contains a tightly controlled IPTG-inducible promoter (Lanzer & Bujard, 1988).

6.4 Reagents

6.4.1 Enzymes

Enzymes used were obtained from the following suppliers:

Calf intestinal phosphatase: Sigma Chemical Co.

DNA polymerase I (Klenow fragment): Bresatec.

Lysozyme: Sigma Chemical Co.

Micrococcal nuclease: Sigma Chemical Co.

Pfu (cloned): Stratagene.

Polynucleotide kinase: Bresatec.

Proteinase K: Sigma Chemical Co.

Restriction endonucleases: Pharmacia, Bresatec, Boehringer Mannheim.

RNase: Sigma Chemical Co. 10mg/ml stock solutions were heated at 95°C for 20 min to inactivate DNases.

T4 DNA ligase: Bresatec.

Taq DNA polymerase: Bresatec.

6.4.2 Radiochemicals

(α-^{32}P)-dATP and (α-^{32}P)-dCTP of specific activity 3000 Ci/mmol (radioactive concentrations of 10mCi/ml) and (γ-^{32}P)-ATP of specific activity 4000 Ci/mmol (radioactive concentration of 10mCi/ml) were purchased from Bresatec.

6.4.3 Chemicals

All chemicals were of analytical grade or of the highest purity available.

5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal): Sigma Chemical Co. Stock solutions at 20 mg/ml in dimethyl formamide were kept at -20°C.

Acetic acid: B.D.H. Labs,. Australia.

Acrylamide: Sigma Chemical Co.

Agarose: Sigma Chemical Co.

Ammonium persulphate (APS): May and Baker Ltd. Stock solutions at 25% (w/v) in H_2O were stored at -20°C.

Ammonium sulphate: May and Baker Ltd.

Ampicillin (sodium salt): Sigma Chemical Co. Stock solutions at 100 mg/ml in H_2O were millipore filter sterilised and stored at -20°C.

β-mercaptoethanol: Sigma Chemical Co.

Bacto-tryptone, yeast extract and Bacto-agar: Difco Labs., U.S.A.

Bovine serum albumin (BSA): Boehringer Mannheim.

Bromophenol blue: B.D.H. Labs., Australia.

Calcium chloride: Sigma Chemical Co.

Chloramphenicol: Sigma Chemical Co. Stock solutions kept as 30 mg/ml in ethanol and stored at -20°C.

Chloroform: B.D.H. Labs., Australia.

Coomassie blue R-250: Sigma Chemical Co.

Deoxyribnucleoside triphosphates: Sigma Chemical Co. Stock solutions at 20 mg/ml prepared in TE and stored at -20°C.

Di-potassium hydrogen orthophosphate: B.D.H. Labs., Australia.

Di-sodium hydrogen orthophosphate: B.D.H. Labs., Australia.

Dialysis membrane (18/32): Union Carbide. Prepared by boiling 10 cm strips in 2% sodium bicarbonate. 1 mM EDTA for 10 min and storing in 20% ethanol.

Dithiothreitol (DTT): Sigma Chemical Co. Prepared as a 100 mM solution in H_2O and stored at -20°C.

Ethanol (95%): Redistilled before use.

Ethidium bromide: Sigma Chemical Co.

Ethylenediaminetetraacetic acid (EDTA): Disodium salt. Sigma Chemical Co.

Formamide: B.D.H. Labs., Australia. Deionised and stored in the dark at -20°C.

Glucose: Ajax.

Glycerol: B.D.H. Labs., Australia.

Glycogen: Boehringer Mannheim

Heparin-sepharose CL-6B: Pharmacia Biotech

Hydrochloric acid (HCl): B.D.H. Labs., Australia.

Isopropanol: May and Baker Ltd.

Isopropyl-β-D-thiogalactopyranoside (IPTG): Sigma Chemical Co. Stock solutions of 1 M in H_2O were millipore filtered and stored at -20°C.

Kanamycin (sulphate): Sigma Chemical Co. Stock solutions of 50 mg/ml in H_2O were millipore filtered and stored at -20°C.

Magnesium acetate: Ajax.

Magnesium chloride: Ajax.

Magnesium sulphate: Ajax.

Methanol: B.D.H. Labs ., Australia.

N,N'-methylene-bis-acrylamide(bis): Sigma Chemical Co.

N, N,N',N',-tetramethylethylenediamine (TEMED): Eastern Kodak Co.

Polyethylene glycol (PEG) 8000: Sigma Chemical Co.

Phenol: B.D.H. Labs., Australia.

Phenylmethylsulfonyl fluoride (PMSF): Sigma Chemical Co.

Potassium acetate: B.D.H. Labs., Australia.

Potassium chloride: B.D.H. Labs., Australia.

Potassium dihydrogen phosphate: B.D.H. Labs., Australia.

rATP: Boehringer Mannheim.

Salmon sperm DNA: Sigma Chemical Co. Sonicated and stored as a 20 mg/ml solution in TE at –20°C.

Sodium acetate: B.D.H. Labs., Australia.

Sodium chloride: B.D.H. Labs., Australia.

Sodium dihydrogen phosphate: B.D.H. Labs., Australia.

Sodium dodecyl sulphate (SDS): Ajax

Sodium hydroxide: Ajax.

Spermidine: Sigma Chemical Co.

Spermine: Sigma Chemical Co.

Sucrose: Ajax

Thiamine: Sigma Chemical Co.

Tris acetate: B.D.H. Labs., Australia.

Trizma base and Tris 7-9: Sigma Chemical Co.

Urea: Sigma Chemical Co.

Xylene cyanol: Sigma Chemical Co.

Zeta-Probe membrane: Biorad

Zinc chloride: May and Baker Ltd.

6.4.4 DNA molecular weight markers

DNA markers were obtained from Bresatec and 500 ng were routinely loaded on an agarose gel.

DMW-S1: **SPP1/*Eco*RI.**

Fragment sizes in bp: 8510, 7350, 6110, 4840, 3590, 2810, 1950, 1860, 1510, 1390, 1160, 980, 720, 480, 360.

DMW-L1: **λ/*Hind* III**

Fragment sizes in bp: 23130, 9416, 6557, 4361, 2322, 2027, 564, 125.

DMW-P1: **pUC19/*Hpa* II**

Fragment sizes in bp: 501, 489, 404, 331, 242, 190, 147, 111, 110, 67, 34, 34, 26.

6.4.5 Protein molecular weight markers

Pre-Stained Low Range Molecular Weight Markers were obtained from BRL. ovalbumin 43 kD, carbonic anhydrase 29 kD, β-lactoglobulin 18.4 kD, lysozyme 14.3 kD, bovine trypsin inhibitor 6.2 kD, insulin α+β 3.0 kD.

6.4.6 Kits

QIAamp Tissue/Blood kit: QIAGEN

QIAquick PCR purification kit: QIAGEN

Maxi Plasmid kit: QIAGEN

QIAquick gel extraction kit: QIAGEN

T7 Super-base sequencing kit: Bresatec Ltd.

6.4.7 Oligonucleotides

The oligonucleotides used during the course of this work were constructed by Bresatec (Australia) using an Applied Biosystems 380B DNA synthesiser.

Table 1(a) Oligonucleotides used to clone IHF

Oligo-nucleotide	Sequence (5'-3')
HimD-1	TCACCTGCAGAGGAGGTAACATATGACCAAGTCAGAATTGATAGAA
HimD-2	CTAACTCAAGCTTTTAACCGTAAATATTGCGATCGCGCAGTTCTTT
HimA-1	TCACCATATGGCGCTTACAAAAGCTGAAATGTCA
HimA-2	ACTCCTGCAGTTACTCGTCTTTGGGCGAAGCGTT

Table 1(b) Deletion analysis and cleavage site oligonucleotides

Oligo-nucleotide	Sequence (5'-3')	Use
L-5RI	TAGACTGGAATTCTAATGATTTGG	BOB' deletions
L6RI	TAGACTGGAATTCGTGGCCCTTG	"
L16RI	TAGACTGGAATTCCTGGACTTGA	"
L26RI	TAGACTGGAATTCACCAGCGACC	"
L35RI	TAGACTGGAATTCAAGCGATTAT	"
L46RI	TAGACTGGAATTCGAGTCGCCTG	"
LHIII	ACGTAGCAAGCTTTGTTTATTGCG	"
R+5HIII	ACGTAGCAAGCTTAGGCGACTCA	"
R44HIII	ACGTAGCAAGCTTTAATCGCTTG	"
R34HIII	ACGTAGCAAGCTTGTCGCTGGTT	"
R24HIII	ACGTAGCAAGCTTCAAGTCCAGC	"
R14HIII	ACGTAGCAAGCTTAAGGGCCACC	"
R4HIII	ACGTAGCAAGCTTAAATCATTAT	"
2RRI	TAGACTGGAATTCTAGACCGCCAT	"
R+10HIII	ACGTAGCAAGCTTAGAGCAGGCG	"
R+20HIII	ACGTAGCAAGCTTCTCAGTGGTT	"
R+30HIII	ACGTAGCAAGCTTGGCCCTTTAG	"
40PT-T	GCGACCAAGCGATTATGAGTCGCCTGCTCT AACCACTGAGCTAAA	cleavage site/ top strand
41PT-T	GCGACCAAGCGATTATGAGTCGCCTGCTCT AACCACTGAGCTAAA	"
42PT-T	GCGACCAAGCGATTATGAGTCGCCTGCTCT AACCACTGAGCTAAA	"
43PT-T	GCGACCAAGCGATTATGAGTCGCCTGCTCT AACCACTGAGCTAAA	"
46PT-B	TTTAGCTCAGTGGTTAGAGCAGGCGACTCA TAATCGCTTGGTCGC	cleavage site/ bottom strand
47PT-B	TTTAGCTCAGTGGTTAGAGCAGGCGACTCA TAATCGCTTGGTCGC	"
48PT-B	TTTAGCTCAGTGGTTAGAGCAGGCGACTCA TAATCGCTTGGTCGC	"
49PT-B	TTTAGCTCAGTGGTTAGAGCAGGCGACTCA TAATCGCTTGGTCGC	"
WT-T	GCGACCAAGCGATTATGAGTCGCCTGCTCT AACCACTGAGCTAAA	wild-type seq. top
WT-B	TTTAGCTCAGTGGTTAGAGCAGGCGACTCA TAATCGCTTGGTCGC	wild-type seq. bottom

Table 1(c) Synthetic oligonucleotides used for B and B' mutagenesis

Oligo-nucleotide	Sequence (5'-3')
-7BOB'-T	CGAC**A**AAGCGATTATGAGTCGCCTGCTCTA
-7BOB'-B	TAGAGCAGGCGACTCATAATCGCTT**T**GTCG
-6BOB'-T	CGACC**C**AGCGATTATGAGTCGCCTGCTCTA
-6BOB'-B	TAGAGCAGGCGACTCATAATCGCT**G**GGTCG
-5BOB'-T	CGACCA**C**GCGATTATGAGTCGCCTGCTCTA
-5BOB'-B	TAGAGCAGGCGACTCATAATCGC**G**TGGTCG
-4BOB'-T	CGACCAA**T**CGATTATGAGTCGCCTGCTCTA
-4BOB'-B	TAGAGCAGGCGACTCATAATCG**A**TTGGTCG
-3BOB'-T	CGACCAAG**A**GATTATGAGTCGCCTGCTCTA
-3BOB'-B	TAGAGCAGGCGACTCATAATC**T**CTTGGTCG
-2BOB'-T	CGACCAAGC**T**ATTATGAGTCGCCTGCTCTA
-2BOB'-B	TAGAGCAGGCGACTCATAAT**A**GCTTGGTCG
-1BOB'-T	CGACCAAGCG**C**TTATGAGTCGCCTGCTCTA
-1BOB'-B	TAGAGCAGGCGACTCATAA**G**CGCTTGGTCG
+1BOB'-T	CGACCAAGCGATTATGAG**G**CGCCTGCTCTA
+1BOB'-B	TAGAGCAGGCG**C**CTCATAATCGCTTGGTCG
+2BOB'-T	CGACCAAGCGATTATGAGT**A**GCCTGCTCTA
+2BOB'-B	TAGAGCAGGC**T**ACTCATAATCGCTTGGTCG
+3BOB'-T	CGACCAAGCGATTATGAGTC**T**CCTGCTCTA
+3BOB'-B	TAGAGCAGG**A**GACTCATAATCGCTTGGTCG
+4BOB'-T	CGACCAAGCGATTATGAGTCG**A**CTGCTCTA
+4BOB'-B	TAGAGCAG**T**CGACTCATAATCGCTTGGTCG
+5BOB'-T	CGACCAAGCGATTATGAGTCGC**A**TGCTCTA
+5BOB'-B	TAGAGCA**T**GCGACTCATAATCGCTTGGTCG
+6BOB'-T	CGACCAAGCGATTATGAGTCGCC**G**GCTCTA
+6BOB'-B	TAGAGC**C**GGCGACTCATAATCGCTTGGTCG
+7BOB'-T	CGACCAAGCGATTATGAGTCGCCT**T**CTCTA
+7BOB'-B	TAGAG**A**AGGCGACTCATAATCGCTTGGTCG
+8BOB'-T	CGACCAAGCGATTATGAGTCGCCTG**A**TCTA
+8BOB'-B	TAGA**T**CAGGCGACTCATAATCGCTTGGTCG
+9BOB'-T	CGACCAAGCGATTATGAGTCGCCTGC**G**CTA
+9BOB'-B	TAG**C**GCAGCCGACTCATAATCGCTTGGTCG
+10BOB'-T	CGACCAAGCGATTATGAGTCGCCTGCT**A**TA
+10BOB'-B	TA**T**AGCAGGCGACTCATAATCGCTTGGTCG
+11BOB'-T	CGACCAAGCGATTATGAGTCGCCTGCTC**G**A
+11BOB'-B	T**C**GAGCAGGCGACTCATAATCGCTTGGTCG
+12BOB'-T	CGACCAAGCGATTATGAGTCGCCTGCTCT**C**
+12BOB'-B	**G**AGAGCAGGCGACTCATAATCGCTTGGTCG

Table 2(a) Synthetic oligonucleotides used for BOB' overlap mutagenesis

Oligonucleotide	Sequence (5'-3')
1BOB'T	CGACCAAGCGA**G**TATGAGTCGCCTGCTCTA
1BOB'B	TAGAGCAGGCG**C**TCATA**C**TCGCTTGGTCG
2BOB'T	CGACCAAGCGAT**G**ATGAGTCGCCTGCTCTA
2BOB'B	TAGAGCAGGCG**C**TCAT**C**ATCGCTTGGTCG
3BOB'T	CGACCAAGCGATT**C**TGAGTCGCCTGCTCTA
3BOB'B	TAGAGCAGGCGACTCA**G**AATCGCTTGGTCG
4BOB'T	CGACCAAGCGATTA**G**GAGTCGCCTGCTCTA
4BOB'B	TAGAGCAGGCGACTC**C**TAATCGCTTGGTCG
5BOB'T	CGACCAAGCGATTAT**T**AGTCGCCTGCTCTA
5BOB'B	TAGAGCAGGCGACT**A**ATAATCGCTTGGTCG
6BOB'T	CGACCAAGCGATTATG**C**GTCGCCTGCTCTA
6BOB'B	TAGAGCAGGCGAC**G**CATAATCGCTTGGTCG
7BOB'T	CGACCAAGCGATTATGA**T**TCGCCTGCTCTA
7BOB'B	TAGAGCAGGCGA**A**TCATAATCGCTTGGTCG

Table 2(b) Synthetic oligonucleotides used for POP' overlap mutagenesis

Oligonucleotide	Sequence (5'-3')
1POP'T	TTCTTGTAGGAACTCATA**C**TCGCTTGGTCG
1POP'B	CGACCAAGCGA**G**TATGAGTTCCTACAAGAA
2POP'T	TTCTTGTAGGAACTCAT**C**ATCGCTTGGTCG
2POP'B	CGACCAAGCGAT**G**ATGAGTTCCTACAAGAA
3POP'T	TTCTTGTAGGAACTCA**G**AATCGCTTGGTCG
3POP'B	CGACCAAGCGATT**C**TGAGTTCCTACAAGAA
4POP'T	TTCTTGTAGGAACTC**C**TAATCGCTTGGTCG
4POP'B	CGACCAAGCGATTA**G**GAGTTCCTACAAGAA
5POP'T	TTCTTGTAGGAACT**A**ATAATCGCTTGGTCG
5POP'B	CGACCAAGCGATTAT**T**AGTTCCTACAAGAA
6POP'T	TTCTTGTAGGAAC**G**CATAATCGCTTGGTCG
6POP'B	CGACCAAGCGATTATG**C**GTTCCTACAAGAA
7POP'T	TTCTTGTAGGAA**A**TCATAATCGCTTGGTCG
7POP'B	CGACCAAGCGATTATGA**T**TTCCTACAAGAA

Table 3 Synthetic oligonucleotides used to create 10 bp insertion in POP' and BOB'

Oligonucleotide	Sequence (5'-3')
POP'∇10T	TTCTTGTAGGAACTCAATGATGTATCTAATCGCTTGGTCGCT
POP'∇10B	AGCGACCAAGCGATTAGATACATCATTGAGTTCCTACAAGAA
BOB'∇10T	CGACCAAGCGATTAGATACATCATTGAGTCGCCTGCTCTA
BOB'∇10B	TAGAGCAGGCGACTCAATGATGTATCTAATCGCTTGGTCG

6.4.8 Growth media

6.4.8.i Liquid media

All media were prepared using glass distilled H_2O and sterilised by autoclaving for 25 min at 120°C and 120 kPa.

LB medium (LB): 1% bacto-tryptone, 1% yeast extract, 1% NaCl, pH 7.0

2YT medium (2YT): 1.6% bacto-tryptone, 1% yeast extract, 0.5% NaCl, pH 7.0

SOC medium (SOC): 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl and H_2O to 1 litre. The solution was autoclaved and allowed to cool to room temperature before the addition of 10 ml of 2 M $MgSO_4$ and 20 ml of 1 M glucose.

6.4.8.ii Solid media

LB plates: LB-agar plates contained LB-medium with 1.5% (w/v) bacto-agar.

Antibiotic supplements: When required, filter sterilised antibiotics were added to the media after it was autoclaved and had cooled. Ampicillin was added to 50-100 μg/ml, kanamycin to 50 μg/ml and chloramphenicol to 30 μg/ml.

6.4.9 Buffers and solutions

2x agarose gel load buffer: 5% glycerol, 1 mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue

10 x PNK: 500 mM Tris pH 7.9, 100 mM MgCl2, 100 mM DTT, 100 mM EDTA. Stored at –20°C.

10 x SD restriction buffer: 33 mM Tris acetate pH 7.85, 0.65 M potassium acetate, 0.1 M magnesium acetate, 4 mM spermidine, 20μm DTT.

10 x T4 ligase buffer: 500 mM Tris pH 7.5, 25 mM DTT, 100 mM MgCl2, 10 mM ATP

10 x TAE: 0.4 M Tris-acetate, 0.2 M Na acetate, 10 mM EDTA, pH 8.2

10 x TBE: 0.89 M Tris-HCl, 0.89 M boric acid, 2.7 mM EDTA, pH 8.3

GET buffer: 50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0

NaOH/SDS buffer: 200 mM NaOH, 1% SDS

Formamide loading buffer: 95 % Formamide, 0.1% bromophenol blue, 0.1% xylene cyanol, 10 mM EDTA

Neutralisation buffer: 3 M Na Acetate pH 5.2

10 x Taq PCR buffer: 670 mM Tris-HCl pH 8.8, 166 mM $(NH_4)_2SO_4$, 2 mg/ml gelatin, 4.5 % Triton X-100

TE: 10 mmTris pH8.0, 1 mM EDTA

Zeta probe transfer buffer: 400 mM NaOH

Depurination buffer: 0.25 M HCl

Phenol/Chloroform (1:1): To melted phenol was added 8-hydroxyquinoline to 0.1 % final concentration. Buffer equilibration was performed by the addition of an equal volume of 1 M Tris-HCl pH 8.0 and the mixture heated until the phenol and aqueous layers mixed. The phases were allowed to separate and the aqueous phase removed. Two volumes of TE was then added to the phenol phase, mixed and allowed to stand until the two phases separated. This was repeated two more times. After the removal of the last aqueous phase an equal volume of chloroform was added to the buffered phenol, mixed and allowed to stand. Equilibrated phenol/chloroform was stored under TE at 4C in the dark. 10 ml aliquots were removed as required and allowed to equilibrate to room temperature.

PBS: Dissolve 8 g NaCl, 0.2g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4.in 800 ml glass distilled H_2O. Adjust pH to 7.4 with HCl and add H_2O to 1000 ml and filter.

6% PAGE/8 M urea: Stock gel solution was prepared by dissolving 57 g acrylamide, 3 g bis-acrylamide and 480.5 g urea in 400 ml glass-distilled H_2O at room temperature. The solution was made to 900 ml with glass-distilled H_2O. The solution was deionised by adding 10 % (w/v) of mixed bed resin and gently stirring the solution at room temperature for 30 min. Mixed bed resin was removed by filtration. The solution was de-gassed.

6.5 Bacterial Procedures

6.5.1 Storage of bacterial stocks

For short term storage bacteria were maintained on the appropriate plates at 4°C. Long term storage of bacterial cultures was at –80°C after addition of glycerol to a final concentration of 20%.

6.5.2 Growth of bacterial strains

Stationary phase bacterial cultures were prepared by inoculating broth with a single colony of bacteria from a plate stock or a loopful of bacteria directly from a glycerol stock, and incubating for 16 hours at the appropriate temperature (usually 30°C or 37°C).

6.5.3 Preparation of high efficiency electrocompetent cells

500 ml of LB-broth was inoculated with 5 ml of an overnight culture of DH5α. Cells were grown at 37°C in a shaking incubator to an O.D. A_{600} of 0.6. The flask was cooled on ice for 15-30 minutes and the cells harvested by centrifugation at 3000 r.p.m. for 15 minutes in a Sorvall GSA rotor. The cells were washed in 500ml (1x), 250ml (1x) and 10ml (1x) ice-cold 10% glycerol. The cells were finally resuspended in 1.5 ml of cold 10% glycerol and 45 µl aliquots frozen and stored at -80°C.

6.5.4 Bacterial electroporation procedure

The BIORAD electroporater was set at 2.0 kV, 200 Ohms and 25 uF. To 45 µl of electroporation competent cells (DH5α) 1 µl of purified ligated DNA was added. The mix was placed into a cooled 0.1 cm cuvette and pulsed. Routinely time constants of 4.7 msec were obtained. The cells were rescued with 1 ml of SOC buffer and incubated at 37°C for 30 minutes. They were then spun in a microfuge

for one minute at 6000 r.p.m., resuspended in 0.2 ml of SOC buffer and plated onto LB + Amp (50 μg/ml) plates and incubated at 37°C overnight. For colour selection, once the cells had been resuspended in 0.2 ml of SOC buffer 20 μl of IPTG (20 mg/ml) and 20 μl of BCIG (20 mg/ml in DMF) was added and the cells immediately plated as above.

6.6 DNA Techniques

6.6.1 Agarose gel electrophoresis of DNA

1% (w/v) agarose in TAE buffer (40 mM Tris-acetate, 20 mM Na-acetate, 1 mM EDTA pH 8.2) was poured onto an appropriately sized glass slide. An equal volume of 2 x loading buffer was added to the DNA sample, mixed and an appropriate volume loaded into each well. Electrophoresis was carried out in 1 x TAE buffer at 100 mA until the bromophenol blue marker dye had run the required distance. Gels were routinely stained in 0.2 μg/ml Ethidium Bromide and visualised under shortwave UV light (254nm).

6.6.2 Restriction endonuclease digestion of DNA

Restriction enzyme digests were performed using the conditions recommended by the supplier. Analytical digests were performed in 20 μl reactions containing 500 ng of DNA and a two-fold excess of enzyme. Preparative digests were performed in 100 μl reactions containing 2-5 μg of DNA and a five-fold excess of enzyme. Reactions were digested for one hour at the appropriate temperature for the restriction enzyme, although for preparative digests this time was extended to overnight to ensure complete digestion.

6.6.3 Isolation of DNA fragments from agarose gels

DNA to be isolated was visualised under longwave UV light (304nm) and the appropriate band excised. The DNA was purified using the spin column method.

Siliconized glass wool was packed into a 0.5 ml eppendorf and a hole was made in the bottom with a fine syringe needle. The excised DNA band was placed in this eppendorf which was itself placed inside a 1.5 ml eppendorf and spun in a microfuge 8000 r.p.m. for ten minutes. The eluate was collected and made up to 100 µl with sterile H_2O and phenol/chloroform (1:1) extracted and ethanol precipitated by the addition of 10 µl of 7.5 M NH_4 acetate and 200 µl of ethanol. The tube was spun at 14000 r.p.m. for 10 minutes and the pellet was washed once with cold 70% ethanol and briefly vacuumed dried. Finally the DNA was resuspended in 10-20 µl of TE buffer.

6.6.4 Subcloning of DNA fragments

6.6.4.i Dephosphorylation of vector ends

Vector DNA was linearised with the appropriate restriction enzyme(s) and dephosphorylated with one unit of calf intestinal phosphatase for 1 hr at 37°C in a 50 µl reaction volume containing 50 mM Tris-HCl pH 9.0, 1 mM $MgCl_2$, 0.1 mM $ZnSO_4$. The vector was then purified as described in section 6.6.3.

6.6.4.ii End-filling of Vectors and Inserts For Blunt End Ligations

To 50 µl of the restriction digest 10 µl of 10 x Med salt buffer (100 mM Tris-HCl pH 7.4, 100 mM $MgCl_2$, 500 mM NaCl, 10 mM DTT), 10 µl of 250 µM dNTPs, H_2O to 99 µl and 1 µl of DNA polymerase I (Klenow fragment) at 2U/µl was added. The reaction was incubated at 37°C for 30 minutes. End-filled vector was then dephosphorylated as described in 6.6.4.1. End-filled DNA fragments were then purified by the spin column method (6.6.3).

6.6.5 Ligations

Both sticky and blunt ended ligations were performed in a 20 µl reaction volume containing 2 µl of 10 x Ligation buffer (supplied by Boehringer Mannheim) and 1U of T4 DNA ligase. Reactions contained 10-20 ng of vector and sufficient

insert to give a 1:1 molar ratio. Reaction were incubated at 16°C overnight. A vector-only control was always included to determine background ligation levels.

6.6.6 Purification of ligation reactions

Prior to electroporation ligation reactions were purified to remove salts which if present results in a lower electroporation efficiency. To a 20 µl ligation reaction 80 µl of sterile H_2O and 1 µl of 5mg/ml glycogen was added. The mix was phenol/chloroform extracted (1:1) and ethanol precipitated. The pellet was washed once in cold 70% ethanol and the pellet briefly dried under vacuum and resuspended in 10 µl of TE.

6.6.7 Small scale isolation of DNA

A single colony was picked into 1.5 ml of LB + Ampicillin (50 µg/ml) and grown overnight at 37°C in a roller drum. The cells were pelleted for 1 minute at 12000 rpm in a microfuge and the supernatant discarded. The pellet was resuspended in 100 µl of GET buffer (50 mM Glucose, 25 mM Tris-HCl pH 7.6, 10 mM EDTA) and incubated at room temperature for 5 minutes. Following which 200 µl of freshly made 200 mM NaOH/1% SDS was added and incubated on ice for 5 minutes. 150 µl of 3M Na Acetate pH 5.2 was then added, mixed by inversion and incubated on ice for 10 minutes. The mix was then centrifuged 12000 r.p.m/10 minutes and the supernatant removed and phenol/chloroform extracted. After spinning to separate the phases the top aqueous phase was carefully removed and ethanol precipitated by the addition of 40 µl of 3M Na Acetate pH 5.2 and 900 µl of ethanol. This was incubated on ice for 10 minutes and spun in a microfuge for 10 minutes at 12000 r.p.m. The pellet was rinsed in cold 70% ethanol, briefly vacuumed dried and resuspended in 30 µl of TE and 1 µl of RNase A (10 mg/ml).

6.6.8 Large scale isolation of plasmid DNA

Cells from a 500 ml overnight culture were harvested by centrifugation at 6000 rpm for 5 minutes (Sorvall SS34 rotor). High purity DNA was obtained by using Qiagen plasmid purification columns (QIAGEN Maxiplasmid kit) and following the manufacturer's recommended protocol.

6.6.9 Sequencing of DNA

6.6.9.i Dideoxy sequencing method

Double-stranded sequencing was performed using the T7 Super-Base sequencing kit from BRESATEC following the supplied protocol. Products of the sequencing reactions were separated by electrophoresis on 6% (w/v) polyacrylamide/8 M urea gels containing 1 x TBE buffer. For polymerisation, to 80 ml of gel stock solution was added 400 µl of freshly prepared 25% (w/v) ammonium persulphate (APS) and 100 µl of TEMED. The gels used were 40cm x 20cm x 0.25mm and were pre-electrophoresed for 30 minutes at 1200 V/25 mA. The gels were run in 1 x TBE buffer at 35 mA/1700 V until the bromophenol blue migrated off the bottom of the gel. Gels were fixed with 200 ml of 10% (v/v) acetic acid/20% (v/v) ethanol and washed with 2 litres of 20% (v/v) ethanol. After drying the gels were autoradiographed overnight at room temperature.

6.6.9.ii Automated fluorescent sequencing

Toward the end of this study, sequencing was performed using an Perkin Elmer 373 Automated Fluorescence Sequencer located at the Institute of Medical and Veterinary Science (IMVS), Adelaide. DNA samples were purified using a Qiaquick Spin kit following the supplied protocol. To a 0.5 ml PCR tube was added 0.5 µg DNA template, 100 ng primer, 8 ul Dye Terminator mix (IMVS) and H_2O to 20 µl. Samples were placed into a PCR machine (PTC-100, MJ Research Inc.) and the following program executed: Step1, 96°C 30 seconds., Step2, 50°C 15 seconds., Step 3, 60°C 4 minutes., for 25 cycles (Steps 1-3). The samples were

purified by the addition of 2 µl 3 M Na acetate pH 4.6, 50 µl 95 % ethanol, precipitated at –80°C for 15 min, spun in a microfuge for 20 minutes. The pellet was washed in 70 % ethanol and vacuumed dried before delivery to the IMVS, Adelaide.

6.6.10 Autoradiography, phosphorimaging and autoradiograph scanning

^{32}P-labelled DNA, which had been electrophoresed on polyacrylamide or agarose gels (transferred to Zeta–Probe membrane) was visualised by autoradiography. A sheet of Fuji RX medical X-ray film or Kodak XAR5 was placed over the gel in an Ilford autoradiography cassette and exposed at room temperature for the required amount of time. For detection of low levels of radioactivity, autoradiography was carried out in a cassette with a tungsten intensifying screen at -80°C. A Molecular Dynamics phosphorimager with ImageQuant software was used for some analyses. For figure preparation, autoradiographs were scanned using an Apple ColorOne Scanner.

6.6.11 PCR amplifications

Each reaction contained PCR buffer (6.4.9), 50 ng of each primer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.1-0.3 U Taq polymerase in a total reaction volume of 50 µl. The amount of target template used was in the range of 50-200 pg DNA or a fraction of a bacterial colony carrying the plasmid was used. The following cycles were used for standard PCR amplifications: Step1; 94°C/2 min., Step2; 94°C/1 min., Step3; 55°C/30 sec., Step4; 74°C/30 sec., Repeat steps 2-4 for 30 cycles.

6.6.12 *In vitro* recombination assays

6.6.12.i cis

Approximately 300 ng of pFLIP(X) plasmid was incubated with 2 µl of cell-free extract (or 1µl of purified Int or IHF, and 1 µl of CE of IHF or Int), in a total volume of 50 µl (10 mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 100 mM KCl and 2 mM EDTA), at 37°C for 60 min (or as indicated in the text). The mix was phenol/chloroform extracted (1:1) and ethanol precipitated. The pellet was washed once in cold 70% ethanol, the pellet briefly dried under vacuum, resuspended in 20 µl of 1 x SD and 2 U of *Xho*I and incubated for 4 hours at 37°C. To the digests was added 2 µl of 10 x loading buffer and then 8 µl was run on a 1% agarose-TAE gel for analysis.

6.6.12.ii trans

Approximately 300 ng of pPgk(P'OP)Neo plasmid and a linear end-labelled *att*B fragment (~2-5 ng) was incubated with Int and IHF proteins (as indicated in the text or figure legends) in a total volume of 50 µl (10 mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 100 mM KCl), at 37°C for 60 min (or as indicated in the text). To the mix was added 5 µl of 10 x loading buffer/1%SDS, the sample heated at 95°C for 5 min and 8 µl run on a 1% agarose-TAE gel prior to transfer to Zeta-probe membrane (6.6.21) for analysis.

6.6.13 Preparation of DNA for microinjection into mouse fertilised eggs

Plasmid DNA was prepared using a QIAGEN Maxi Plasmid kit following the manufacturer's protocol. The plasmid DNA was cut with the appropriate restriction enzyme(s). The digested DNA was run on an agarose gel with appropriate markers, the gel stained in ethidium bromide, de-stained and visualised on a medium wave UV light box. The appropriate DNA band was cut out with a scalpel blade. The DNA was isolated from agarose using a QIAquick gel extraction kit following the manufacturer's protocol. 10 x PBS was added to

the purified DNA to make it 1 x PBS. The purity and concentration of the sample was determined by using a spectrophotometer to scan a diluted sample from 210-310 nm, and by running aliquots on an agarose gel against known standards. The DNA was diluted in 1 x PBS to give a final concentration of 10 ng/μl.

6.6.13.i Microinjection of DNA

Microinjections were performed by Steve MacIlfatrick, Obstetrics and Gynaecology Dept., University of Adelaide, South Australia. The prepared DNA (6.6.13) was injected (1-3 pl) into zygotes using standard procedures (Hogan *et al.*, 1986).

6.6.15 Preparation of genomic DNA from mouse tails

A 0.5 cm piece of tail was cut using surgical scissors, from 4-6 week old mice, snap frozen in liquid N_2 and stored at -80°C until required. Genomic DNA was prepared using the QIAamp Tissue/Blood kit following the supplied protocol.

6.6.16 Isolation of chromatin from mouse livers

6.6.16.i Removal of livers from transgenic mice

Mice were killed by dislocation of the neck. The abdomen was opened just below the rib cage and the skin torn open up to the neck and down to the anus. Using scissors, the abdomen lining was cut open up through the ribs and sternum, exposing the liver (1-2 g/mouse). Livers were excised with scissors by cutting away connective tissue. Livers were placed into 50 ml falcon tubes (3 livers/tube) and snap frozen in liquid N_2, and stored at -80°C until required.

6.6.16.ii Isolation of nuclei/chromatin

Using a mortar and pestle, frozen liver tissue (3 livers) was crushed under liquid N_2 to a powder. The powder was tipped into a 50 ml falcon tube and thawed in a 42°C water bath for 1 min. Following which 10 ml of ice-cold homogenisation

buffer (10 mM HEPES pH 7.6, 25 mM KCl, 1 mM EDTA, 2 M sucrose, 10 % glycerol, 0.15 mM spermine, 0.5 mM spermidine, 1 mM DTT and 0.1 mM PMSF) was added and mixed into a slurry. The tissue was homogenised using a motor driven teflon-glass homogeniser, 5 strokes on low speed. The homogenate was diluted to 32 ml with homogenisation buffer and layered (4 x 8 ml homogenate) onto 4 x 4 ml cushions of homogenisation buffer in SW41 centrifuge tubes. Samples were spun at 24 K for 30 min. at 4°C in an SW41 swing-out rotor in a Beckman ultracentrifuge. The supernatants were carefully aspirated with a 10 ml pipette followed by a pasteur pipette for the remaining 4 ml of supernatant. The nuclei pellets obtained from this procedure were washed in 3 x 3 ml cold nuclei lysis buffer (10 mM HEPES pH 7.6, 100 mM KCl, 0.1 mM EDTA, 3 mM $MgCl_2$, 10 % glycerol, 1 mM DTT and 0.1 mM PMSF) and finally resuspended in a total volume of 800 µl of nuclei lysis buffer, aliquotted into eppendorf tubes (16 x 50 µl), snap frozen in liquid N_2 and stored at –80°C.

6.6.17 Micrococcal nuclease digests

In a total volume of 50 µl, prepared chromatin (25 µl) was resuspended in 5 mM $CaCl_2$, 10 mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 100 mM KCl and 1 µl of prewarmed Micrococcal nuclease serially diluted (1/10, 1/100, 1/1000) down from 0.5 U/µl. The reactions were incubated for 5 min at 37°C and terminated by the addition of EDTA to 20 mM. Samples were RNAse treated for 1 hour at 37°C, phenol/chloroform extracted (1:1) and ethanol precipitated. The pellet was washed once in cold 70% ethanol, the pellet briefly dried under vacuum, resuspended in 9 µl of water, 1 µl of 10 x LB and then 5 µl was run on a 3% agarose-TAE gel for analysis.

6.6.18 Annealing complementary oligonucleotides

Duplex oligonucleotides (20 µl), each at a concentration of 200 ng/µl, were added together (1:1), mixed and heat denatured for 5 min at 95°C in a waterbath. The

tubes were removed from the waterbath and allowed to cool to room temperature to facilitate the annealing process.

6.6.19 Kinasing of DNA restriction fragments and oligonucleotides

Oligonucleotides were 5'-end-labelled with T4 polynucleotide kinase and γ-^{32}P-ATP. Approximately 200 ng of oligonucleotide or annealed oligonucleotides were kinased in a 50 µl reaction volume containing 50 uCi of γ-^{32}P-ATP and 2U of PNK in 1 x PNK buffer. The reaction mix was incubated at 37°C for 30 min. The reactions were purified by adding 1 µl of 5mg/ml glycogen, phenol/chloroform extracting (1:1) and ethanol precipitating with 5 µl of 3 M Na acetate pH 5.2 and 110 µl of 95 % ethanol. The pellet was washed once in cold 70% ethanol, briefly dried under vacuum, and resuspended in 20 µl of TE.

6.6.20 End-labelling and end-filling using the Klenow fragment of DNA polymerase I

DNA restriction fragments to be used as radioactive size markers were end-labelled by an endfilling reaction containing 10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 25 uM dGTP/dTTP and dATP or dCTP (added from stocks of 0.25 mM dNTPs in 5 mM Tris-HCl pH 8.0, 1 mM EDTA and depending on which ^{32}P-labelled dNTP was used), 1 unit of Klenow fragment and 50 µCi α-^{32}P-dATP and/or α-^{32}P-dCTP. The reaction was incubated at 37°C for 30 min. and terminated by heating at 70°C for 15 min or by the addition of EDTA to a final concentration of 50 mM. When a 3'overhang needed to be blunted or radiolabelled, or a blunt end needed to be radiolabelled, a chew-back fill-in reaction was performed by allowing 2 min of Klenow fragment 3'-5'exonuclease activity before adding the nucleotides. If the DNA was to be used for another downstream reaction the reactions were purified by adding 1 µl of 5mg/ml glycogen, phenol/chloroform extracting (1:1) and ethanol precipitating with 5 µl of 3 M Na acetate pH 5.2 and 110 µl of 95 % ethanol. The pellet was washed

once in cold 70% ethanol, briefly dried under vacuum, and resuspended in 20 µl of TE.

6.6.21 Transfer to Zeta-probe blotting membrane

Agarose gels were soaked in 0.25 M HCl for 10-15 min and rinsed with distilled water. Using a Hoeffer vacuum transfer apparatus, and following the manufacturer's supplied protocol, the DNA was transferred from the gel to the Zeta-probe blotting membrane in the presence of 0.4 M NaOH for 30 min.

6.6.22 Southern analysis

6.6.22.i Oligo-labelling of DNA

Oligo-labelling of DNA fragments was performed using the reagents and protocol provided by Bresatec based on the method described by Feinberg and Vogelstein (1983). Approximately 10-25 ng of a DNA restriction fragment was taken up in 12 µl of water and denatured by heating at 100°C for 3 minutes and snap chilling on ice. The DNA was then added to 50 µCi of lyophilised α-^{32}P-dATP resuspended in 12 µl of a solution containing 40 µM dATP, dCTP, dGTP, dTTP in a buffer of 100 mM Tris-HCl pH7.6, 100mM NaCl, 20 mM $MgCl_2$, 200 µg/ml BSA. Also contained in this solution was 6.5 µg of chemically synthesised decamer primer. Then 5 units of DNA polymerase I, Klenow fragment was added.

After incubation at 37°C for 30 minutes the volume was increased to 100 µl with water and the reaction terminated by phenol/chloroform extraction. Glycogen was added to 100 ng/ml and 10 µl 7 M ammonium acetate and 300 µl ethanol also added to precipitate the labelled DNA fragments. The probe was resuspended in water and stored frozen until required. Before use, the probe was denatured by boiling for 5 minutes.

6.6.22.ii Hybridisation of DNA immobilised on a membrane

Filters were placed in bottles and prehybridised at 42°C for at least 4 hours in a solution containing 10% (w/v) Dextran sulphate, 1 M NaCl, 1% (w/v) SDS, 40% (v/v) formamide, 50 mM Tris-HCl pH 7.5, 5 x Denhardts and 500 µg/ml denatured sonicated salmon sperm DNA (Maniatis *et al.*, 1982). Heat denatured probe was added and allowed to hybridise for 16-24 hours at 42°C. Filters were washed in 2 x SSC, 0.1% (w/v) SDS at 65°C and in 0.5 x SSC,0.1% (w/v) SDS at 65°C for 20 minutes each, prior to autoradiography or phosphorimaging.

6.6.23 Oligonucleotide-directed mutagenesis

Mutagenesis was performed using the Stratagene QuikChange Site-Directed Mutagenesis Kit and following the manufacturer's protocol.

6.7 Protein techniques

6.7.2 Polyacrylamide gel electrophoresis

6.7.2.i 15 % SDS polyacrylamide gels

The separating gels were 0.5 mm thick, and prepared from a 30 % 38:1 acryl:bis stock, diluted to 15 % with 0.75 M Tris HCl pH 8.8, 0.2 % SDS containing 10 µl/ml APS and 1 µl/ml TEMED. The stacking gel was prepared from the same acrylamide stock but was diluted to 0.4 % with 0.44 M Tris HCl pH 8.8, 0.12 % SDS containing 10 µl/ml APS and 1 µl/ml TEMED. Gels were run at 25 mAmp, 250 V in 125 mM Tris pH 8.5, 0.96 M glycine, 0.5 % SDS.

Gels were stained in Coomassie blue R-250, 50 % methanol, 10 % acetic acid, destained for 2 hours in 50 % methanol, 10 % acetic acid and dried onto Whatman 3MM paper under vacuum at 65°C for one hour.

6.7.2.ii Phastsystem gels

1 ml of cells to be analysed were spun for 5 min in a microfuge and resuspended in 100 µl of loading buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA, 3.3 % SDS, 10 mM DTT, 0.01 % Bromophenol blue). Samples were heated to 100°C for 5 min and microfuged for 15 min, discarding any pellet. Then 1 ul samples were run on Phastgel (8-25) Gradient gels with SDS buffer strips (Pharmacia Biotech). The separation procedure used was described in the Pharmacia Phastsystems Users Manual, Separation Technique File No. 110. Gels were stained with Coomassie Blue and destained according to the development procedure outlined in Development Technique File No. 200.

6.7.2.iii Protein gel scanning and analyses

A Molecular Dynamics Personal Densitometer was used to scan gels and analyses were performed using ImageQuant software.

6.7.3 Preparation of cell-free extracts/storage

Overnight cultures of *E. coli* C600 (IHF+), C600∇*himA* (IHF-), C600/pFLIP(X) (IHF+/Int+) and C600∇*himA*/pFLIP(X) (IHF-/Int+) grown in LB containing ampicillin (100 µg/ml) at 37°C, were diluted 200-fold into 50 ml of the same broth and incubated with aeration at 37°C. At an A_{600}=0.5, IPTG was added to 0.4 mM. After 2 hours the cultures were chilled on ice and the cells harvested by centrifugation (8000 rpm, 10 min, 4°C, JA20 rotor). The pellets were resuspended in 3 ml of ice-cold cell sonication buffer (100 mM Tris-HCL pH7.8, 200 mM KCl, 0.1 mM DTT, 1 mM EDTA, 1 mM PMSF) and subjected to disruption by sonication (3 x 1 min bursts, 5 min rest on ice between bursts) with a Branson Sonic Power Co. B-30 Sonifer Cell Disrupter. The lysates were cleared by centrifugation (18000 rpm, 30 min, 4°C, SS34 rotor) and aliquots (30 µl) frozen in dry ice and stored at –80°C.

6.7.4 186 Intasome preparation and storage

Three sets of Intasome mixes were prepared containing either 2.5, 5 or 10 ng/µl pPgk(P'OP)Neo DNA. For microinjection into mouse oocytes intasomes were prepared in 1 x PBS and contained approximately an 1.5-fold excess of purified Int and IHF proteins based on *in vitro trans* assay data whereby 300 ng of pPgk(P'OP)Neo in 50 µl required 57.6 ng of Int and 2 ng of IHF proteins (minimally) for maximal integrative recombination. Therefore 250 µl of 2.5 ng/µl pPgk(P'OP)Neo contained 192 ng Int and 6.7 ng IHF, 5 ng/µl pPgk(P'OP)Neo contained 384 ng Int and 13.4 ng IHF and 10 ng/µl pPgk(P'OP)Neo contained 768 ng Int and 26.8 ng IHF. Each intasome mix was aliquotted out, 10 µl per eppendorf tube, snap frozen in liquid nitrogen and stored at –80°C until required.

6.8 Containment facilities and animal ethics

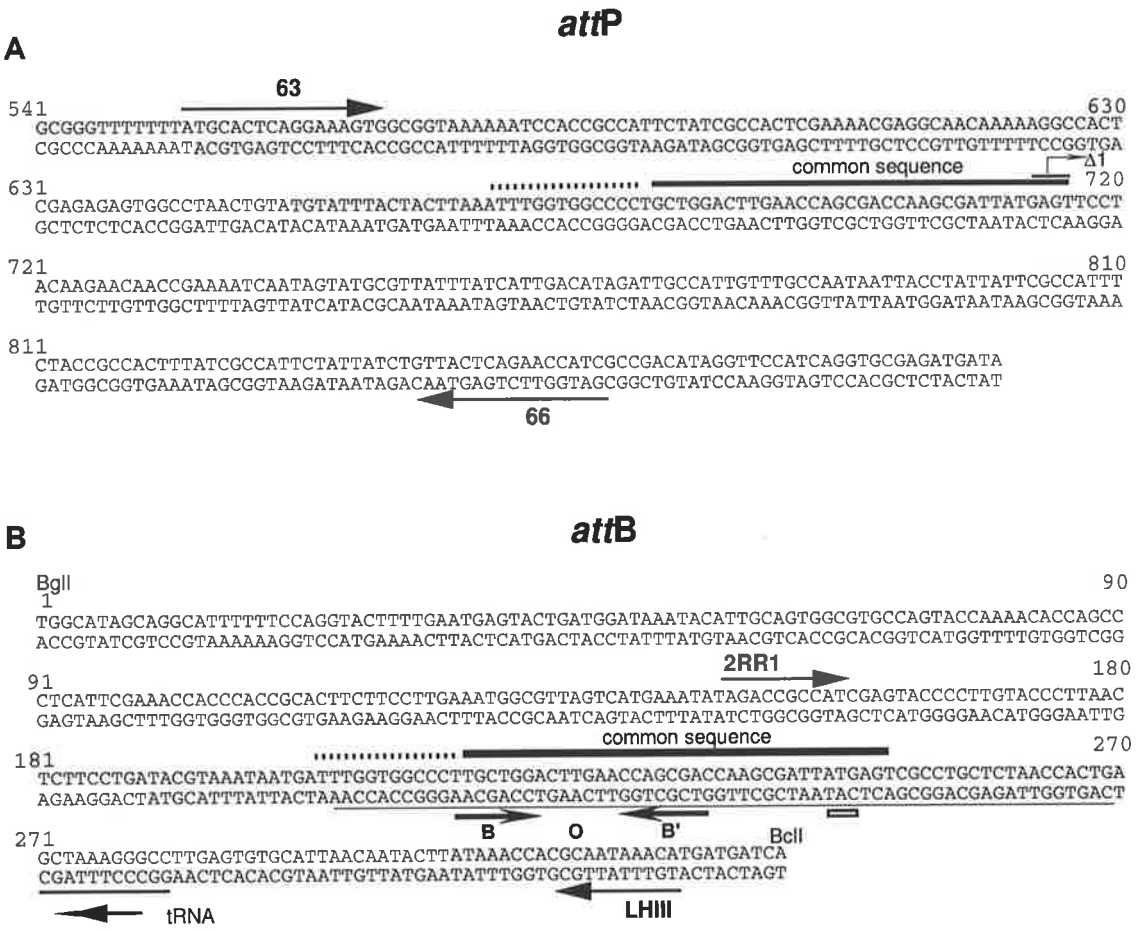
All manipulations involving recombinant DNA were carried out in accordance with the regulations and approval of the Australian Academy of Science Committee on Recombinant DNA and the University Council of the University of Adelaide. All procedures involving animals were carried out with the approval of the University of Adelaide Animal Ethics Committee.

APPENDIX

APPENDIX

Figure 1

Phage and bacterial attachment sites of 186



A). The *att*P common sequence (thick black line) located 681-716 bp from the *Pst*I site at 65.5% on the 186 map (Kalionas *et al.*, 1986). Identity between *att*P and *att*B extends 12 bp to the left of the common sequence following a single basepair mismatch (dashed line). The locations of oligonucleotides 63 and 66 is shown. The leftbreak point of a deletion, 186 Δ1 (Kalionis *et al.*, 1986) is shown.

B). The *att*B common sequence (thick black line) is located in the 3' half of a putative *Ile* tRNA gene (underlined). The CAU codon is shown (open rectangle). Numbering is from the *Bgl*I site. The locations of oligonucleotides LHIII and 2RR1 is shown. The predicted locations of B and B' (Reed, 1994) are shown.

BIBLIOGRAPHY

Bibliography

Albert. H., Dale E.C., Lee E., Ow D. W., (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J.* **7:** 649-659.

Ambreski K., Gottesman S., (1981). Site-specific recombination: Xis-independent excisive recombination of bacteriophage λ. *J. Mol. Biol.* **153(1):** 67-78.

Ambreski K., Hoess R., Sternberg N., (1983). Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell* **32(4):** 1301-1311.

Ambreski, K., Hoess, R., (1984). Bacteriophage P1 site-specific recombination, purification and properties of the cre recombinase protein. *J. Biol. Chem.* **259:** 1509-1514.

Ambreski, K., Hoess, R., (1992). Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng.* **5:** 87-91.

Anderson G. B., Choi S. J., BonDurant R. H., (1994). Survival of porcine inner cell masses in culture and after injection into blastocysts. *Theriogenology* **42:** 204-212.

Araki K., Masatake A., Miyazaki J., Vassali P,. (1995). Site specific recombination of a transgene in fertilised eggs by transient expression of Cre recombinase. *Proc. Natl. Acad. Sci.* **92:** 160-164.

Araki K., Araki M., Yamamura K., (1997). Targeted integration of DNA using mutant lox sites in embryonic stem cells. *Nuc. Acids Res.* **25(4):** 868-872.

Argos P. A., Landy A., Abreski K., Egan J. B., Haggard-Ljungquist E., Hoess R. H., Kahn M., Kalionis B., Leong J. M., (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5:** 433-440.

Astumian J. H., Waldman A. S., Scocca J. J., (1989). Site-specific recombination between cloned attP and attB sites from the Haemophilus influenzae bacteriophage HP1 propagated in recombination-deficient E. coli. *J. Bacteriol.* **171(3):** 1747-1750.

Auvray F., Coddeville M., Ritzenthaler P., Dupont L., (1997). Plasmid integration in a wide range of bacteria mediated by the integrase of Lactobacillus delbrueckii bacteriophage mv4. *J. Bacter.* **179(6):** 1837-1845.

Azaro M. A., Landy A., (1997). The isomeric preference of Holliday junctions influences resolution bias by lambda integrase. *EMBO J.* **16(12):** 3744-3755.

Baubonis W., Sauer B., (1993). Genomic targeting with purified Cre recombinase. *Nuc. Acids res.* **21(9):** 2025-2029.

Bertani L. E., Bertani G., (1971). Genetics of P2 and related phages. *Adv. Genetics* **16:** 199-237.

Bertani L. E., Six E. W., (1988). The P2-like phages and their parasite, P4. In R. Calendar (ed.), The bacteriophages, vol 2, Plenum Publishing Corp. NY. Pages:73-143.

Bradley C., Ling O. P., Egan J. B., (1975). Isolation of phage P2.186 intervarietal hybrids and 186 insertion mutations. *Mol. Gen. Genet.* **140:** 123-125.

Bradley I. (1984). Formation of germline chimaeras from embryo derived teratocarcinoma cell lines. *Nature*. **309:** 255-256.

Brinster R. L. (1981). Somatic expression of Herpes Thymidine Kinase in mice following injection of a fusion gene into eggs. *Cell*. **27:** 223-231.

Brinster R. L., Chen, H.Y. (1985) Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. *P.N.A.S.* **82:** 4438-4442.

Brinster R. L. (1988). Introns increase transcriptional efficiency in transgenic mice. *P.N.A.S.* **85:** 836-840.

Brinster R. L., Avarbock M. R., (1994). Germline transmission of donor haplotype following spermatogonial transplantation. *Proc. Natl. Acad. Sci.* **91:** 11303-11307.

Brinster R. L., Zimmermann J. W., (1994). Spermatogenesis following male germ-cell transplantation. *Proc. Natl. Acad. Sci.* **91:** 11298-11302.

Brosius J., Ullrich A., Raker M. A., Gray A., Dull T. J., Gutell R. R., Noller H. F., (1981). Construction and fine mapping of recombinant plasmids containing the rrnB ribosomal RNA operon of E. coli. *Plasmid* **6(1):** 112-118.

Brosius J., Erfle M., Storella J., (1985). Spacing of the –10 and-35 regions in the tac promoter. Effect on its in vivo activity. *J. Biol. Chem.* **260(6):** 3539-3541.

Buchholtz F., Ringrose, L., Angrand, P.O., Rossi, F., Stewart, A.F. (1996). Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. *Nuc. Acids Res.* **24:** 4256-4262.

Buchholtz F., Angrand, P.O., Stewart, A.F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotech.* **16(7):** 657-662.

Burgin Jr. A. B., Nash H. A., (1992). Symmetry in the mechanism of bacteriophage λ integrative recombination. *P.N.A.S.* **89:** 9642-9646.

Bushman F., (1995). Targeting Retroviral Integration. *Science* **267:** 1443-1444.

Bushman W., Landy A., (1984). Determinants of directionality in λ site-specific recombination *Cell* **39:** 699-706.

Bushman W., Thompson J. F., Vargas L., Landy A., (1985). Control of directionality in lambda site specific recombination. *Science* **230(4728)**: 906-911.

Campbell A. M., (1992). Chromosomal insertion sites for phages and plasmids. *J. Bact.* **174(23):** 7495-7499.

Campbell K. H., McWhir J., Ritchie W. A., Wilmut I., (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **380(6569):** 64-66.

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

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

Choo A., (1997). Human artificial chromosomes and gene therapy. *Today's Life Science* **Nov:** 14-18.

Connolly A. A., Browning G. G., (1997). Randomised trial of neomycin dexamethasone spray vs drop preparation for the treatment of active chronic mucosal otitis media. *Clin Otolaryngol.* **22(6):** 529-531.

Constantini F., Lacy, E. (1981). Introduction of a rabbit β-globin gene into the mouse germ line. *Nature*: **294:** 92-94.

Coutelle C., Douar A., Colledge W., Froster U., (1995). The challenge of foetal gene therapy. *Nature Medicine* **1(9):** 864-866.

Cox M.M. (1988). FLP Site-Specific Recombination System of *Saccharomyces cerevisiae*. In Genetic Recombination, R. Kucherlapati and G.R. Smith, eds. (Washington, D.C. : American Society for Microbiology), pp. 429-443.

Cozzi E., White D. J. G., (1995). The generation of transgenic pigs as potential organ donors for humans. *Nature Medicine:* **1(9):** 964-966.

Craig N. L., Nash H. A., (1983). The mechanism of phage λ site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell* **35:** 795-803.

Craig N. L., Nash H. A., (1984). *E. coli* integration host factor binds to specific sites in DNA. *Cell* **39:** 707-716.

Craig N. L., (1988). The mechanism of conservative site-specific recombination. *Ann. Rev. Genet.* **22:** 77-105.

Dabkowski P. L., Vaughan H. A., McKenzie I. F., Sandrin M. S., (1993). Characterisation of a cDNA clone encoding the pig α1-3 galactosyltransferase: implications for xenotransplantation. *Transplant. Proc.* **25(5):** 2921.

Damak S., (1996). Targeting gene expression to the wool follicle in transgenic sheep *Biotechnology* **14(2):** 181-184.

Danks D. M., (1994). Potential and ethics of somatic and germ-line gene therapy. *Today's Life Science* **Jan:** 10-13.

Ditto M. D., Roberts D., Weisberg R. A., (1994). Growth phase variation of integration host actor level in *Escherichia coli. J. Bact.* **176(12):** 3738-3748.

Doetschmann T., Williams, P. (1988). Establishment of hamster blastocyst-derived embryonic stem (ES) cells. *Dev. Biol.* **127:** 224-227.

Dorer D. R., Henikoff S., (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* **77:** 993-1002.

Dorgai L., Yagil E., Weisberg R. A.,. (1995). Identifying determinants of recombination specificity: construction and characterisation of mutant bacteriophage integrases. *J. Mol. Biol.* **252:** 178-188.

Dorgai L., Sloan S., Weisberg R. A., (1998). Recognition of core binding sites by bacteriophage integrases. *J. Mol. Biol.* **277(5):** 1059-1070.

Drlica K., Rouviere-Yaniv J., (1987). Histone-like proteins of bacteria. *Microbiol Rev.* **51(3):** 301-319.

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

Eissenberg J. C., Elgrin S. C., (1991). Boundary functions in the control of gene expression. *Trends Genet.* **7(10):** 335-340.

Eposito D., Fitzmaurice W. P., Benjamin R. C., Goodman S. D., Waldman A. S., Scocca J. J., (1996). The complete nucleotide sequence of bacteriophage HP1 DNA. *N. A. R.* **24(12):** 2360-2368.

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

Evans B. R., Chen J. W., Parsons R. L., Jayaram M., (1990). Identification of the active site tyrosine of FLP recombinase. Possible relevance of its location to the mechanism of recombination. *J. Biol. Chem.* **266(11):** 7312.

Feil R., Brocard J., Mascrez B., LeMeur M., Metzger D., Chambon P., (1996). Ligand-activated site-specific recombination in mice. *P.N.A.S.* **93(20):** 10887-10890.

Feinberg A. P., Vogelstein B., (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132(1):** 6-13.

Filutowicz M., Grimek H., Appelt K,. (1994). Purification of the *Escherichia coli* integration host factor (IHF) in one chromatographic step. *Gene* **147:** 149-150.

Flamm E. L., Weisberg R. A., (1985). Primary structure of the *hip* gene of *Escherichia coli* and of its product, the β subunit of integration host factor. *J. Mol. Biol.* **183:** 117-128.

Freundlich M., Ramani N., Mathew E., Sirko A., Tsui P., (1992). The role of integration host factor in gene expression in E. coli. *Mol. Micro.* **6(18):** 2557-2563.

Friedman D. I., (1988). Integration host factor: a protein for all reasons. *Cell* **55:** 545-554.

Galili U., Swanson, K., (1991). Gene sequences suggest inactivation of α-1-3-galactosyltransferase in catarrhine after the divergence of apes from monkeys. *P.N.A.S.* **88:** 7401-7404.

Gamas P., Burger A. C., Churchward G., Caro L., Galas D., Chandler M., (1986). Replication of pSC101: effects of mutations in the E. coli DNA binding protein IHF. *Mol. Gen. Genet.* **204(1):** 85-89.

Gardner J. F., Nash H. A., (1986). Role of *Escherichia coli* IHF protein in lambda site-specific recombination.: A mutational analysis of binding sites. *J. Mol. Biol.* **191:** 181-189.

Garrick D., Fiering S., Martin D. I., Whitelaw E., (1998) Repeat-induced gene silencing in mammals. *Nat. Genet.* **18(1):** 56-59.

Gellert M., Nash. H., (1987). Communication between segments of DNA during site-specific recombination. *Nature* **325:** 401-404.

Golic K.G., Lindquist, S. (1989). The FLP Recombinase of Yeast Catalyses Site-Specific Recombination in the Drosophila Genome. *Cell.* **59:** 499-509.

Goodman S. D., Nash H. A., (1989). Functional replacement of a protein-induced bend in a DNA recombination site. *Nature* **341:** 251-254.

Goodman S. D., and Scocca J. J., (1989). Nucleotide sequence and expression of the gene for the site-specific integration protein from bacteriophage HP1 of Haemophilus influenzae. *J. Bacteriol.* **17(8):** 4232-4240.

Goodman S. D., Nicholson S .C., Nash H,. (1992). Deformation of DNA during site-specific recombination of bacteriophage λ: Replacement of IHF protein by HU protein or sequence-directed bends. *Proc. Natl. Acad. Sci.* **89:** 11910-11914.

Goodrich J. A., Schwartz M. L., McClure W. R.,. (1990). Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). *Nuc. Acids Res.***18(17):** 4993-5000.

Gordon J.W., Scangos, G.A., (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. *P.N.A.S.* **77:** 7380-7384.

Gordon J.W., Ruddle, F.H. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science.* **214:** 1244-1246.

Görlich D., Mattaj I. W., (1996). Nucleoplasmic transport. *Science* **271:** 1513-1518.

Gottschling D.E., Aparicio, O.M, Zakian, B.A. (1990). Position effect at *S cerevisiae* telomeres: reversible repression of Pol II transcription *Cell.* **63:** 751-762.

Gregory P. D., Horz W., (1998). Life with nucleosomes: chromatin remodelling in gene regulation *Curr. Opin. Cell. Biol.* **10(3):** 339-345.

Grindley N. D. F., (1997). Site-specific recombination: Synapsis and strand exchange revealed. *Current Biology* **7(10):** 608-612.

Gronostajski R.M., Sadowski, P.D. (1985). The FLP Protein of the 2-micron Plasmid of Yeast. *J. Biol. Chem.* **260**: 12328-12335.

Grosveld F., van Assendelft G. B., Greaves D. R., Kollias G. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51(6):** 975-985.

Grosveld F., Kollias, G. (1992). Transgenic Animals. Acad. Press Ltd. London. Chpt 11. 247-265.

Gu H., Zou Y. R., Rajewsky K., (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73(6):** 1155-1164.

Gu H., Marth J. D., Orban P.C., Mossman H., Rajewsky K., (1994). Deletion of a DNA polymerase β gene segment in T cells using cell type-specific gene targeting. *Science* **265:** 103-106.

Guo F., Gopaul D. N., Van Duyne G. D., (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389(4):** 40-46.

Hales L.M., Gumport R. I., Gardner J. F., (1994). Determining the DNA sequence elements required for binding integration host factor to two different target sites. *J. Bact.* **176(10):** 2999-3006.

Hales L.M., Gumport R. I., Gardner J. F., (1996). Examining the contribution of a dA+dT element to the conformation of E. coli integration host factor-DNA complexes. *NAR.* **24(9):** 1780-1786.

Halligen B.D., Davis, J.L., Edwards, K.A., Liu, L.F. (1982). Intramolecular and intermolecular strands transfer by HeLa DNA topoisomerase 1. *J. Biol. Chem.* **257:** 3995-4000.

Hamada T., Hiroyuki, S., Seki, R., Sakaki, Y. (1993). Mechanism of chromosomal integration of transgenes in microinjected mouse eggs; sequence analysis of genome-transgene and transgene-transgene junctions at two loci. *Gene* **128:** 197-202.

Han Y. W., Gumport R. I., Gardner J. F,. (1993). Complementation of bacteriophage lambda integrase mutants: evidence for an intersubunit active site. *EMBO J.* **12(12)**: 4577-4584.

Han Y. W., Gumport R. I., Gardner J. F., (1994). Mapping the functional domains of bacteriophage lambda integrase protein. *J. Mol. Biol.* **235:** 908-925.

Harrington R. E., (1992). DNA curving and bending in protein - DNA recognition. *Mol. Micro.* **6(18):** 2549-2555.

Hartung M., Kisters-Woike B., (1998). Cre mutants with altered DNA binding properties. *J. Biol. Chem.* **273(36):** 22884-22891.

Hauser M. A., Scocca J. J., (1992) Site-specific integration of Haemophilus influenzae bacteriophage HP1. *J. Biol. Chem.* **267:** 6859-6864.

Hazelrigg T., Levis, R., Rubin, G.M. (1984). Transformation of white locus DNA in drosophila: dosage compensation, zeste interaction, and position effects. *Cell* **36:** 469-472.

Hickman A. B., Waninger S., Scocca J. J., Dyda F., (1997). Molecular organisation in site-specific recombination: The catalytic domain of bacteriophage HP1 integrase at 2.7Å resolution. *Cell* **89:** 227-237.

Hoess R. H., Ambreski K., (1985). Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *J. Mol. Biol.* **181:** 351-362.

Hogan B., Constantini E., Lacy E., (1986). Manipulating the mouse embryo: A laboratory manual. *Cold Spring Harbour Laboratory Manual*, Cold Spring Harbour.

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

Hoyt M. A., Knight D. M., Das A., Miller H. I., Echols H., (1982). Control of phage λ development by stability and synthesis of *c*II protein: role of the viral *c*II and host *hfl*A, *him*A and *him*D genes. *Cell* **31:** 565-573.

Huang L.C. (1991). A bacterial model system for chromosomal targeting. *Nuc. Acids. Res.* **19:** 443-448.

Hwang E. S., Scocca J. J., (1990). Interaction of integration host factor from *Escherichia coli* with the integration region of the *Haemophilis influenzae* bacteriophage HP1. *J. Bact.* **172(9):** 4852-4860.

Hwang D. S., Kornberg A., (1992). Opening of the replication origin of E. coli by DnaA protein with HU or IHF. *J. Biol. Chem.* **267(32):** 23083-23086.

Jackson I. J., (1989). Genes and Embryos. Oxford IRL press. 165-221.

Jaenisch R., (1976). Germline integration and mendelian transmission of the exogenous Moloney leukemia virus. *P.N.A.S.* **73:** 1260-1264.

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

Jamieson A. C., Kim S. H., Wells J. A., (1994). In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* **33(19):** 5689-5695.

Jayaram M., (1985). Two-micrometer circle site-specific recombination: The minimal substrate and the possible role of flanking sequences. *Proc. Natl. Acad. Sci.* **82:** 5875-5879.

Jayaram M., (1994). Phosphoryl transfer in FLP recombination: a template for strand transfer mechanisms. *TIBS* **19:** 78-82.

Jayaram M., (1997). The cis-trans paradox of integrase. *Science* **276:** 49-51.

Jiang R., Gridley T., (1997). Gene targeting: Things go better with Cre. *Current Biol.* **7:** R321-R323.

Joyner A. L., (1991). Gene targeting and gene trap screens using embryonic stem cells: new approaches to mammalian development. *Bioessays* **13(12):** 649-656.

Joyner A. L., Auerbach A., Skarnes W. C., (1992). The gene trap approach in embryonic stem cells: the potential for genetic screens in mice. *Ciba Found. Symp.* **165:** 277-288.

Jung, S., Rajewsky, K. (1993). Shutdown of class switch recombination by deletion of a switch region control element. *Science*. **259:** 984-987.

Kalionis B., Dodd. I. B., Egan J. B., (1986). Control of gene expression in the P2-related template coliphages III. DNA sequence of the major control region of phage 186. *J. Mol. Biol.* **191:** 199-209.

Kalos M., Fournier E. K., (1995). Position-Independent transgene expression mediated by boundary elements from the Apolipoprotein B chromatin DNA. *Mol. Cell. Biol.* **15(1):** 198-207.

Kato Y., Tetsuya T. Sotomaru Y., Kurokawa K., Kato J., Dogushi H., Yasue H., Tsunodo Y., (1998). Eight calves cloned from somatic cells of an adult. *Nature* **282(5396):** 2095-2098.

Kho S. H., Landy A., (1994). Dissecting the resolution reaction of λ integrase using suicide Holliday junction substrates. *EMBO J.* **13(11):** 2714-2724.

Kikuchi Y., Nash H. A., (1978). The bacteriophage λ *int* gene product. *J Biol. Chem.* **253(20):** 7149-7157.

Kilby N. J., Snaith M. R., Murray J. A. H., (1993). Site-specific recombinases: tools for genome engineering. *TIG.***9(12):** 413-421.

Kim S., Landy A., (1992). Lambda Int protein bridges between higher order complexes at two distant chromosomal loci *att*L and *att*R. *Science* **256:** 198-203.

Kim J.H., Jung-Ha, H.S., Chung, K.S. (1997). Development of a positive method for male stem cell-mediated gene transfer in mouse and pig. *Mol. Reprod. Dev.* **46(4):** 515-526.

Kitts P. A. Nash H. A,. (1987). Homology dependent interactions in phage λ site-specific recombination. *Nature* **329:** 346-348.

Kitts P. A. Nash H. A,. (1988). Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. *Mol. Biol.* **204:** 95-107.

Kitts P. A. Nash H. A,. (1988). An intermediate in the phage λ site-specific recombination reaction is revealed by phosphorothioate substitution in DNA. *Nucleic Acids Res.* **16:** 6839-6856.

Kolb A. F., Ansell R., McWhir J., Siddell S. G., (1999). Insertion of a foreign gene into the β-casein locus by Cre-mediated site-specific recombination. *Gene* **227:** 21-31.

Kolot M., Yagil E., (1994). Position and direction of strand exchange in bacteriophage HKO22 integration. *Mol Gen Gen.* **245:** 623-627.

Kono T., (1997). Nuclear transfer and reprogramming. *Rev. Reprod.* **2(2):** 74-80.

Kornberg A., (1977). Multiple stages in the enzymic replication of DNA. *Biochem. Soc. Trans.* **5(2):** 359-374.

Kosturko L., Murialdo H., (1989). The interaction of *E. coli* integration host factor and λ cos DNA: multiple complex formation and protein-induced bending. *Nucl. Acids Res.*, **17:** 317-334.

Kühn R., Schwenk F., Aguet M., Rajewsky K., (1995). Inducible gene targeting in mice *Science* **269:** 1427-1429.

Kühn R., Schwenk F., (1997). Advances in gene targeting methods. *Curr. Opin. Immunol.* **9(2):** 183-188.

Kwon H. J., Tirumalai R., Landy A., Ellenberger T., (1997). Flexibility in DNA recombination: Structure of the lambda integrase catalytic core. *Science* **276:** 126-131.

Landy A., Ross W., (1977). Viral integration and excision: structure of the λ att sites. *Science* **197:** 1147-1160.

Landy A., Hoess R. H., Bidwell K., Ross W., (1978). Site specific recombination in bacteriophage λ: Structural features of recombining sites. *Cold Spring Harbour Symp. Quant. Biol.* **43:** 1089-1097.

Landy A., (1989). Dynamic, structural and regulatory aspects of λ site-specific recombination. *Ann. Rev. Biochem.* **58:** 913-949.

Landy A., (1993). Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. *Current opinion in genetics and development* **3:** 699-707.

Lanzer M., Bujard H., (1988). Promoters largely determine the efficiency of repressor action. *P.N.A.S.* **85:** 8973-8977.

Lee J., Whang I., Lee J., Jayaram M., (1994). Directed protein replacement in recombination full sites reveals trans-horizontal DNA cleavage by Flp recombinase. *EMBO J.* **13(22):** 5346-5354.

Lee J., Jayaram M., Grainge I., (1999). Wild-type Flp cleaves DNA in trans. *EMBO J.* **18(3):** 784-791.

Lee Y., Park J., (1998). A novel mutant loxP containing part of long terminal repeat of HIV-1 spacer region: presentation of possible target site for antiviral strategy using site-specific recombinase. *Biochem Biophys. Res. Commun.* **253(3):** 588-593.

Leong J. M., Nunes-Duby S. E., Lesser C. F., Youderian P., Susskind M. M., Landy A., (1985).The Ø80 and P22 attachment sites. *J. Biol. Chem.* **260(7):** 4468-4477.

Leong J. M., Nunes-Duby S. E., Oser A. B., Lesser C. F., Youderian P., Susskind M. M., Landy A., (1986). Structural and regulatory divergence among genes of lambdoid phage. *J. Mol. Biol.* **189:** 603-616.

Luger K., Rechsteiner T. J., Flaus A. J., Waye M. M., Richmond T. J., (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272(3):** 301-311.

MacWilliams M. P., Gumport R. I., Gardner J. F., (1996). Genetic analysis of the bacteriophage λ *att*L nucleoprotein complex. *Genetics* **143:** 1069-1079.

MacWilliams M. P., Gumport R. I., Gardner J. F., (1997).Mutational analysis of protein binding sites involved in formation of the bacteriophage λ *att*L complex. *J. Bact.* **179(4):** 1059-1067.

Maniatis T., Fritsch E. F., Sambrook J., (1982). Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory*, Cold Spring Harbor.

Mark W.H., Signorelli, K., Blum, M., Kwee, L., Lacy, E. (1992). Genomic structure of the locus associated with an insertional mutation in line 4 transgenic mice. *Genomics* **13:** 159-166.

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

McFarlane M., Wilson J. B., (1996). A model for the mechanism of precise integration of a microinjected transgene. *Transgenic Res.* **5:** 171-177.

Mehtali M., LeMeur M., Lathe R., (1990). The methylation-free status of a housekeeping transgene is lost at high copy number. *Gene* **91:** 179-184.

Meng L., Ely J. J., Stouffer R. L., Wolf D. P., (1997). Rhesus monkeys produced by nuclear transfer. *Biol. Reprod. 57(2):* 454-459.

Miller H. I., Kikuchi A., Nash H. A., Weisberg R. A., Friedman D. I., (1979). Site-specific recombination of bacteriophage λ: the role of host gene products. *Cold Spring Harbour Symp. Quant. Biol.* **43:** 1121-1126.

Miller H.I., Friedman D. L., (1980). An *E. coli* gene product required for λ site-specific recombination. *Cell* **20:** 711-719.

Miller H. I., Mozola M.A., Friedman D. I., (1980). Int-h: An int mutation of phage λ that enhances site-specific recombination. *Cell* **20(3):** 721-729.

Miller H. I., (1981). Direct role of the himA gene product in phage λ integration *Cell* **25(1):** 269-276.

Miller H. I., Abraham J., Benedik M., Campbell A., Court D., Echols H., Fischer R., Sosa L., (1981). Regulation of the integration-excision reaction by bacteriophage λ. *Cold Spring Harbour Symp. Quant. Biol.* **45:** 439-445.

Miller H. I., (1984). Primary structure of the *himA* gene of *Escherichia coli*: Homology with DNA binding protein HU and association with the phenylalanyl-tRNA synthetase operon. *Cold Spring Harbor Symp. Quant. Biol.* **49:** 691-698.

Mizuuchi K., Nash H.A., (1976). Restriction assay for integrative recombination of bacteriophage lambda DNA in vitro: requirement for closed circular DNA substrate. *P. N. A. S.* **73(10):** 3524-3528.

Mizuuchi K., Mizuuchi M., (1978). Integrative recombination in bacteriophage λ: In vitro study of the intermolecular reaction. *Cold Spring Harbour Symp. Quant. Biol.* **43:** 1111-1115.

Mizuuchi M., Mizuuchi K., (1980). Integrative recombination of bacteriophage λ: extent of the DNA sequence involved in attachment site function. *Proc. Natl. Acad. Sci.* **77(6):** 3220-3224.

Mizuuchi K., Weisberg R., Enquist L., Mizuuchi M., Buraczynska M., Foeller C., Hsu P. L., Ross W., Landy A., (1981). Structure and function of the phage λ att site: Size, Int-binding sites, and location of the crossover point. *Cold Spring Harbour Symp. Quant. Biol.* **45:** 429-437.

Mizuuchi M., Mizuuchi K., (1985). The extent of DNA sequence required for a functional bacterial attachment site of phage lambda. *Nuc. Acids Res.* **13(4):** 1193-1207.

Moitoso de Vargas L., Pargellis C. A., Hasan N. M., Bushman E. W., Landy A., (1988). Autonomous DNA binding domains of λ integrase recognise two different sequence families. *Cell* **54:** 923-929.

Moitoso de Vargas L, Kim S, Landy A., (1989). DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. *Science* **244:** 1457-1461.

Mondragón A., (1997). Solving the cis/trans paradox in the Int family of recombinases. *Nature Structural Biol.* **4(6):** 427-429.

Morris A.C. (1991). FLP-mediated recombination in the vector mosquito, *Aedes aegypti.*. *Nuc. Acids Res.* **19:** 5895-5900.

Mu Chen C., Bund Choo, K., Cheng, W.T.K. (1995). Frequent deletions and sequence aberrations at the transgene junctions of transgenic mice carrying the papillomavirus regulatory and the SV40 Tag gene sequences. *Transgenic Res.* **4:** 52-59.

Nagaraja R., Weisberg R. A., (1990). Specificity determinants in the attachment sites of bacteriophage HK022 and λ. *J. Bact.* **172(11):** 6540-6550.

Nandi A. K. (1988). Regulated expression of genes inserted at the human chromosomal β-globin locus by homologous recombination. *P.N.A.S.* **85:** 3845-3849.

Nash H. A., (1975). Integrative recombination of bacteriophage lambda DNA *in vitro*. *Proc Natl Acad Sci.* **72(3):** 1072-1076.

Nash H. A., (1981). Integration and excision of bacteriophage λ the mechanism of conservative site specific recombination. *Annu. Rev. Genet.* **15:** 143-167.

Nash H. A., Robertson C. A., (1981). Purification and properties of the E. coli protein factor required for λ integrative recombination. *J. Biol. Chem.* **256(17):** 9246-9253.

Nash H. A., Bauer C. E., Gardner J. F., (1987). Role of homology in site-specific recombination of bacteriophage λ: Evidence against joining of cohesive ends. *Proc. Natl. Acad. Sci.* **84:** 4049-4053.

Nash H. A., (1990). Bending and supercoiling of DNA at the attachment site of bacteriophage λ. *TIBS* **15:** 222-227.

Neumann E., Schaeffer-Ridder M., Wang Y., Hofschneider P H,. (1982). Gene transfer into mouse myeloma cells by electroporation in high electric fields. *EMBO J.* **1:** 841-845.

Nunes-Düby., Matsumoto L., Landy A., (1987). Site-specific recombination intermediates trapped with suicide substrates *Cell* **50:** 779-788.

Nunes-Düby., Matsumoto L., Landy A., (1989). Half-*att* site substrates reveal the homology independence and minimal protein requirements for productive synapsis in λ excisive recombination. *Cell* **59:** 197-206.

Nunes-Düby S. E., Tirumalai R. S., Dorgai L., Yagil E., Weisberg R. A., Landy A., (1994). λ integrase cleaves DNA in cis. *EMBO J.* **13(18):** 4421-4430.

Nunes-Düby S. E., Azaro M. A., Landy A., (1995). Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr. Biol.* **5(2):** 139-148.

Nunes-Düby S. E., Yu D., Landy A., (1997). Sensing homology at the strand-swapping step in λ excisive recombination. *J. Mol. Biol.* **272:** 493-508.

O'Gorman S., Fox D. T., Wahl G. M., (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251:** 1351-1355.

Paleyander R K., Velander W. H., Lee T. K., Lubon H., (1997). Transgenic pigs produce functional human factor VIII in milk. *Nat Biotech.* **15(10):** 971-975.

Palmiter, R.D., Brinster, R.L. (1982). Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature* **300:** 611-615.

Palmiter, R.D., Norstedt, G. (1983). Metallothionein-human GH fusion genes stimulates growth of mice. *Science* **222:** 809-814.

Palmiter, R.D., Brinster, R.L. (1986). Germline transformation of mice. *Ann. Rev. Genet.* **20:** 465-499.

Parekh B. S., Hatfield G. W., (1996). Transcriptional activation by protein-induced DNA bending: evidence for a structural transmission model. *P.N.A.S.* **93(3):** 1173-1177.

Pargellis C. A., Nunes-Duby S. E., Moisoto de Vargas L., Landy A., (1988). Suicide recombination substrates yield covalent λ integrase-DNA complexes and lead to identification of the active site tyrosine. *J. Biol. Chem.* **263:** 7678-7685.

Pascoe W. S., Kemler R., Wood S. A., (1992). Genes and functions: trapping and targeting in embryonic stem cells. *Biochemica et Biophysica Acta.* **1114:** 209-221.

Patsey R. L., Bruist M. F.,. (1995). Characterisation of the interaction between the lambda intasome and *att*B. *J. Mol. Biol* **252:** 47-58.

Peña C. E. A., Stoner J.E., Hatfull G.F., (1996). Positions of strand exchange in Mycobacteriophage L5 integration and characterization of the *att*B site. *J. Bacteriol.* **178:** 5533-5536.

Peña C. E. A., Lee M. H., Pedulla M. L., Hatfull G. F., (1997). Characterization of the Mycobacteriophage L5 attachment site, attP. *J. Mol. Biol.* **266(1):** 76-92.

Perez-Martin J., Timmis k. N., de Lorenzo V., (1994). C0regulation by bent DNA. Functional substitutions of the IHF site at sigma 54-dependent promoter Pu of the upper-TOL operon by intrinsically curved sequences. *J. Biol. Chem.* **269(36):** 22657-22663.

Perreten, V., Scharwz, F., Cresta, L., Teuber, M., (1997). Antibiotic resistance spread in food. *Nature* **389:** 801-802.

Pierson L. S., Kahn M. L., (1984). Cloning of the integration and attachment regions of bacteriophage P4. *Mol. Gen. Genet.* **195:** 44-51.

Plasterk R. H. A., (1992). Genetic switches: mechanism and function. *TIG* **8(12):** 403-406.

Platt, J.L., Bach, F.H. (1991). The barrier to xenotransplantation. *Transplantation* **52(6):** 937-947.

Prentki P., Chandler M., Galas D. J., (1987). E. coli integration host factor bends the DNA at the ends of IS1 and in an insertion hotspot with multiple IHF binding sites. *EMBO J.* **6:** 2479-2487.

Pursel, V.G. (1989). Genetic engineering of livestock. *Science.* **244:** 1281-1287.

Qian X., Cox M. M., (1995). Asymmetry in active complexes of FLP recombinase. *Genes and Dev.***9:** 2053-2064.

Reed M. R., Shearwin K., Pell L., Egan J. B., (1997). The dual role of Apl in prophage induction of coliphage 186. *Mol. Micro.* **23(4):** 669-681.

Reiter W., Palm P., Yeats S., (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nuc. Acids Res.* **17(5):** 1907-1911.

Rice P. A., Yang S., Mizuuchi K., Nash H. A., (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87(7):** 1295-1306.

Richet E., Abcarian P., Nash H. A., (1986). The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in λ integrative recombination. *Cell* **47:** 1011-1021.

Richet E., Abcarian P., Nash H. A., (1988). Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a Protein-DNA complex, *Cell* **52:** 9-17.

Rijkers, T., Peetz, A., Ruther, U. (1994). Insertional mutagenesis in transgenic mice. *Transgenic Res.* **3:** 203-215.

Robertson C. A., Nash H. A., (1988) Bending of the bacteriophage λ attachment site by *E. coli* integration host factor. *J. Biol. Chem.* **263:** 3554-3557.

Robertson, E.J. (1986) Embryo derived stem cell lines. Pluripotential stem cell lines as a route into the mouse germline. *TIGS.* **2:** 9-13.

Rohlmann A., Gotthardt M., Willnow T. E., Hammer R. E., Herz J., (1996). Sustained somatic gene inactivation by viral transfer of Cre recombinase. *Nat. Biotech.* **14:** 1562-1564.

Rosenberg A. H., Lade B. N., Chui D. S., Lin S. W., Studier F. W., (1987). Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56(1):** 125-135.

Ross W., Landy A., (1982). Bacteriophage λ Int protein recognises two classes of sequence in the phage att site: characterisation of arm-type sites. *P.N.A.S.* **79:** 7724-7728.

Ross W., Landy A., (1983). Pathways of λ Int recognition in the region of strand exchange. *Cell* **33:** 261-272.

Ross W., Shulman M., Landy A. (1982). Biochemical analysis of *att*-defective mutants of the phage lambda site-specific recombination system. *J Mol. Biol.* **156:** 505-529.

Rossant J., Bernelot-Moens C., Nagy A,. (1993). Genome manipulation in ES cells. *Philos. Trans R. Soc. Lond. B. Biol. Sci.* **339(1283):** 207-215.

Rossant J., Nagy A,. (1995). Genome engineering: the new mouse genetics. *Nature Medicine* **1(6):** 592-594.

Sadowski P., (1986). Site-specific recombinases: changing partners and doing the twist. *J. Bacteriol.* **165:** 341-347.

Sauer, B. (1989). Cre-stimulated recombination at *loxP*-containing DNA sequences placed into the mammalian genome. *Nuc. Acids Res.* **17:** 147-161.

Sauer, B. (1990). Targeted Insertion of Exogenous DNA into the Eukaryotic Genome by the Cre Recombinase. *The New Biologist.* **2(5):** 441-449.

Schnieke A. E., Kind A. J. Ritchie W., Mycock K., Scott A. R., Ritchie M., Wilmut I., Colman A., Campbell H. S., (1997). Human factor IX transgenic sheep produced by transfer of nuclei from transfected foetal fibroblasts. *Science* **278:** 2130-2133.

Segal D. J., Carrol D., (1995). Endonuclease-induced, targeted homologous extrachromasomal recombination in Xenopus oocytes. *Proc. Natl. Acad. Sci.* **92:** 806-810.

Segall A. M., Goodman S. D., Nash H. A., (1994). Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. *EMBO J.* **13(19):** 4536-4538.

Sells M. A., Li J., Chernoff J,. (1995). Delivery of protein into cells using polycationic liposomes. *Biotechniques* **19 (1):** 73-78.

Senecoff, J.F. (1985) Cox M. M., The FLP recombinase of the yeast 2-μm plasmid: characterisation of its recombination site. *P.N.A.S.* **82:** 7270-7274.

Senecoff, J.F., Cox, M.M. (1986). Directionality in FLP Protein-promoted Site-specific Recombination Is Mediated by DNA-DNA Pairing. *J. Mol. Biol.* **261:** 7380-7386.

Senecoff, J.F. (1988). DNA Recognition by the FLP Recombinase of the Yeast 2-micron Plasmid. A Mutational Analysis of the FLP Binding Site. *J. Mol. Biol.* **201:** 405-421.

Shaikh A. C., Sadowski P. D., (1997). The Cre recombinase cleaves the lox site in trans. *J. Biol. Chem.* **272(9):** 5695-5702.

Shamblott M. J., Gearhart J. D., (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells. *P.N.A.S.* **95:** 13726-13731.

Shim, H., Gutierrez-Adan, A., Chen, L.R., BonDurant, R.H., Behboodi, E., Anderson, G.B. (1997). Isolation of pluripotent stem cells from cultured porcine primordial germ cells. *Biol. Reprod* **57(5):** 1089-1095.

Smith A. J., De Sousa M. A., Kwabi-Addo B., Heppell-Parton A., Impey H., Rabbitts P., (1995). A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. *Nat Genet.* **9:** 376-385.

Smith G. R., (1988). Homologous recombination in prokaryotes. *Microbiol. Rev.* **52:** 1-28.

Snyder U. K, Thompson J. F, Landy A., (1989). Phasing of protein-induced DNA bends in a recombination complex. *Nature* **341:** 255-257.

Studier F. W., Moffat B. A., (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189(1):** 113-1130.

Susulic V. S., Frederich R. C., Lawitts J., Lowell B. B., (1995). Targeted disruption of the β3-adrenergic receptor gene. *J. Biol. Chem.* **49:** 29483-29492.

Tearle, R.G., Tange, M.J., Zannetino, Z.L., Katerelos, M., Shinkel, T.A., van Denderen, B.J., Lonie, A.J., Lyons, I., Nottle, M.B., Becker, C., Peura, A.M., Wigley, P.L., Crawford, R.J., Robins, A.J., Pearse, M.J., d'Apice, A.J. (1996). The alpha-1,-3-galactosyltransferase knockout mouse. Implications for xenotransplantation. *Transplantation* **61(1):** 13-19.

te Riele H., Maandag E. R., Berns A., (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *P.N.A.S.* **89:** 5128.

Thaler D. S., Stahl F., (1988). DNA-double chain breaks in recombination of phage λ and of yeast. *Ann. Rev. Genet.* **22:** 169-197.

Thomas K. R., Folger K. R., Capecchi M. R., (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44(3):** 419-428.

Thomas K.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell.* **51:** 503-512.

Thompson R.F., Moisoto de Vargas L., Koch C., Kahmann R. Landy A., (1987). Cellular factors couple recombination with growth phase: characterisation of a new component in the λ site-specific recombination pathway. *Cell* **50:** 901-908.

Thompson and Landy (1988). Empirical estimation of protein-induced DNA bending angles: applications to λ site-specific recombination complexes. *Nucl. Acids Res.* **16:** 9687-9705.

Tirumalai R. S., Healey E., Landy A., (1997). The catalytic domain of lambda site-specific recombinase. *Proc. Natl. Acad. Sci.* **94(12):** 6104-6109.

Tsurushita N., Fu H., Warren C., (1996). Phage display vectors for in vivo recombination of immunoglobulin heavy and light chain genes to make large combinatorial libraries. *Gene* **172(1):** 59-63.

Turner B. M., (1993). Decoding the nucleosome. *Cell* **75(1):** 5-8.

van Denderen, B.J., Salvaris, E., Romanella, M., Aminian, A., Tange, M.J., Pearse, M.J., d'Apice, A.J. (1997). *Transplantation* **64(6):** 882-888.

Verma I. M., Somia N., (1997). Gene therapy-promises, problems and prospects *Nature* **389:** 239-242.

Vita R. C., Campos M., Belfort R. Paiva E. R., (1998). Alterations in blood-aqueous barrier after corneal refractive surgery. *Cornea* **17(2):** 158-162.

Vize, P.D. (1988). Introduction of a porcine growth hormone fusion gene into transgenic pigs promotes growth. *J. Cell Sci.* **90:** 295-300.

Volkert F.C. (1986)., Site-specific recombination promotes plasmid amplification in yeast. *Cell.* **46:** 541-550.

Voytas D., Boeke J. D., (1993). Yeast retrotransposons and tRNAs. *TIG* **9(12):** 421-427.

Wakayama T., Perry A. C., Zuccotti M., Johnson K. R., Yanagimachi R., (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394(6691):** 369-374.

Walia H., Chen H. Y., Sun J. M., Holth L. T., Davie J. R., (1998). Histone acetylation is required to maintain the unfolded nucleosomal structure associated with transcribing DNA. *J. Biol. Chem.* **273(23):** 14516-14522.

Wang S., Cosstick R., Gardner JF., Gumport R. I., (1995). The specific binding of E coli integration host factor involves both major and minor grooves of DNA. *Biochemistry* **34(40):** 13082-13090.

Weisberg R. A., Landy A., (1983). Site-specific recombination in phage λ. Lambda II. New York: *Cold Spring Harbour Lab.* p 211-250.

Whitelaw, C.B.A. (1991). Targeting expression to the mammary gland: Intronic sequences can enhance the efficiency of gene expression in transgenic mice. *Transgenic Research.* **1:** 3-13.

Wilkie, T.J., Brinster, R.L., Palmiter, R.D. (1986). Germline and somatic mosaicism in transgenic mice. *Dev. Biol.* **118:** 9-18.

Wilkie. T.J., Palmiter, R.D. (1987). Analysis of the integrant in MyK-1103 transgenic mice in which males fail to transmit the integrant. *Mol Cell. Biol.* **7:** 1646-1655.

Willadsen S. M., (1986). Nuclear transplantation in sheep embryos. *Nature.* **320(6057):** 63-65.

Willadsen S. M., (1989). Cloning of sheep and cow embryos. *Genome* **31(2):** 956-962.

Wilmut, I., Whitelaw, C.B. (1994). Strategies for production of pharmaceutical proteins in milk. *Reprod. Fertil, Dev.* **6(5):** 625-630.

Wilmut I., Schnieke A. E., McWhir J., Kind A. J., Campbell K. H., (1997). Viable offspring derived from foetal and adult mammalian cells. *Nature* **385(6619):** 810-813.

Woods W. H., Egan J. B., (1972). Integration site of noninducible coliphage 186. *J. Bacteriol.* **111(2):** 303-307.

Wu H. M., Crothers D. M., (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature* **308(5959):** 509-513.

Yang C. C. Nash H. A., (1989). The interaction of E. coli IHF protein with its specific binding sites. *Cell* **57:** 869-880.

Yagil E., Dorgai L., Weisberg R. A., (1995). Identifying determinants of recombination specificity: construction and characterisation of chimeric bacteriophage integrases. *J. Mol. Biol.* **252:**163-177.

Yin S., Bushman W., Nash H. A., (1985). Interaction of the λ site-specific recombination protein Xis with attachment site DNA. *P.N.A.S.* **82:** 1040-1044.

Yong Z., Yuiang L., (1998). Nuclear-cytoplasmic interaction and development of goat embryos reconstructed by nuclear transplantation: production of goats by serially cloning embryos. *Biol. Reprod.* **58(1):** 266-269.

Yu A., Bertani L. E., Haggård-Ljungquist E., (1989). Control of prophage integration and excision in bacteriophage P2: nucleotide sequences of the *int* gene and *att* sites. *Gene* **80:** 1-12.

Yu A., Haggård-Ljungquist E., (1993a). The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination. *J. Bacteriol.* **175:** 7848-7855.

Yu A., Haggård-Ljungquist E., (1993b). Characterisation of the binding sites of two proteins involved in the bacteriophage P2 site-specific recombination system. *J. Bacteriol.* **175:** 1239-1249.

Zulianello L., de la Gorgue de Rosny E., van Ulsen P., van de Putte P., Goosen N., (1994). The HimA and HimD subunits of IHF can specifically bind to DNA as homodimers. *EMBO J. 13(7):* 1534-1540.

Zou Y-R., Müller W., Gu H., Rajewsky K., (1994). Cre-*lox*P-mediated gene replacement: a mouse strain producing humanised antibodies. *Current Biol.* **4:** 1099-1103.