Cloning and characterisation of DNA flanking highly expressed integrated plasmids in two mouse myeloma cell lines

Paul Walker B.Sc (Leeds 1993)

Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester

June 1999
August 1998

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

Acknowledgements

All of the work presented in this thesis was done with the supervision of Dr Tim Harrison.

Special thanks are due to Sarahs Munson and Uff for their help and advice throughout the project. Nick Suttcliffe also provided help and advice at the start of the work.

Thanks also to all members, past and present, of David Critchley's lab who provided friendship and support throughout project. Particular thanks for this to Mark Holt, Jim Norman and Bipin Patel.

The completion of this work is due, in no small way, to my wife Rebecca without whose constant love and support I would almost certainly have given up and done something more sensible.

ABSTRACT

The work presented in this thesis is a continuation of a project attempted by several previous workers. Attempts to optimise plasmid vectors for the high level expression of heterologous proteins in the mouse myeloma cell line, J558L, led to the chance discovery of two highly expressing transfectants (C6 and D8). Both these transfectants resulted from plasmids containing a gpt selection marker, a lysozyme reporter gene and the IgH enhancer. Both expressed the lysozyme reporter at around 100 fold higher level than that normally obtained from transfection with the same plasmids, and approximately the same level as the endogenous immunoglobulin genes.

This thesis describes the cloning and characterisation of DNA flanking the plasmids in both the C6 and D8 cell lines.

A genomic library from the C6 cell line was constructed in λ-DASH and screened with plasmid sequence probes. A single clone of 15kb was isolated. This contained 9kb of flanking DNA from one side of the plasmid. Sequencing of the clone revealed that the plasmid had integrated within a B1 repeat element. No homologies to other known sequences were identified.

Similarly, a genomic library from the D8 cell line was constructed and screened. Three independent clones were obtained giving a total of 20kb flanking DNA. Sequencing revealed that flanking DNA on one side of the plasmid was extremely A/T rich. The possible implications of this are discussed.

The A/T rich fragment from the D8 locus was shown to bind to isolated nuclear matrix *in vitro*. However no binding to the nuclear matrix was detected for either locus in an *in vivo* assay. Both the C6 and D8 loci were shown to be DNase sensitive, but no specific hypersensitive sites were identified for either cell line.

Also presented is work which contributes to a project aimed towards using Flp recombinase to place expression constructs and other DNA sequences at the C6 locus by homologous recombination. In order to resolve problems in obtaining Flp activity, an epitope tag was placed at the C-terminus of the protein. Flp is shown to localise to the nucleus. Data is also presented which suggest that Flp (in the expression vectors used in our laboratory) has an apparent negative effect on co-transfected plasmids.

CONTENTS

Chapter 1 INTRODUCTION

1.1 Background to the project....1
1.2 Factors which affect gene expression at the transcriptional level....8
1.2.1 DNA Structure....8
1.2.2 DNA packaging....10
1.2.3 Histone Acetylation and Methylation....12
1.2.4 Promoter and enhancer elements....13
1.3 Locus Control Regions....15
1.3.1 β-globin gene switching....16
1.4 DNase sensitivity as a determinant of chromatin structure....19
1.5 Matrix attachment regions....22
1.5.1 Common sequences in matrix attachment regions....22
1.5.2 The nuclear matrix....23
1.5.3 The effect of matrix attachment regions on gene expression....23
1.6 Other recombinant protein expression systems....25
1.6.1 Bacterial systems....25
1.6.2 Microbial eukaryotes....26
1.6.3 Higher eukaryotic and mammalian systems....27
1.7 Aims of the study....29

Chapter 2 MATERIALS AND METHODS

2.1.1 Chemicals....31
2.1.2 Bacterial Strains....32
2.2 General DNA extraction and Manipulation....32
2.2.1 Handling of genomic and plasmid DNA....32
2.2.2 Quantification of DNA by spectrophotometry....33
2.2.3 Precipitation of DNA....33
2.2.4 Isolation of genomic DNA....33
2.2.5 Isolation of plasmid DNA by rapid mini-preps....34
2.2.6 Commercial mini-prep kits....35
2.2.7 Midi and Maxi preps....35
2.2.8 Restriction digestion....36
2.2.9 Heat Inactivation of restriction enzymes....36
2.2.10 Ligation of DNA....36
2.2.11 Tranformation of E.coli with plasmid DNA....36
2.2.12 Detection of a-complementation....37
2.2.13 Band preparation from agarose gels....37
2.2.14 5′ end labelling....37
2.3 Agarose Gel Electrophoresis....38
2.3.1 Standard Agarose Gel Electrophoresis....38
2.3.2 Pulsed field gel electrophoresis....38
2.4 Library Construction & screening....38
2.4.1 Partial digestion of C6 and D8 DNA....38
2.4.2 Fractionation of partially digested DNA by centrifugation through sucrose gradients....39
2.4.3 Ligation of genomic DNA to λ-DASH vector arms....39

2.4.4 Packaging of λ-DASH libraries 39
2.4.5 Preparation of E.coli for plating of λ-DASH libraries 40
2.4.5 Titering of libraries 40
2.4.6 Primary screening of libraries 40
2.4.7 Secondary Screening. 41
2.4.8 Tertiary Screening. 41
2.5 Purification of λ DNA. 41
2.5.1 Preparation of cleared lysate. 41
2.5.2 Preparation of λ-DNA using Promega "Wizard™" Spin Columns 42
2.5.3 Purification using CsCl gradients 43
2.6 Southern Blotting 43
2.6.1 Alkali blotting 43
2.6.2 Blotting with 20xSSC. 43
2.6.3 Probe labelling with fluorescein 44
2.6.4 Random oligonucleotide labelling with [α32P] dCTP 44
2.6.5 Purification of labelled probes using spin columns 44
2.6.7 Prehybridization and Hybridization 44
2.7 Mammalian Cell Culture Methods 45
2.7.1 Culture of J558L and derived cell lines. 45
2.7.2 Culture of CV1, COS and derived cell lines. 45
2.7.3 Transfection of COS and CV-1 cells. 45
2.7.4 Immunofluorescence. 46
2.7.5 Harvesting of transfected cells 47
2.8 Protein Methods 47
2.8.1 Determination of protein concentration 47
2.8.2 β-galactosidase assays 47
2.8.3 Luciferase assays 47
2.8.4 Separation of protein by SDS-PAGE electrophoresis 48
2.8.5 Electrophoretic transfer of resolved proteins onto nylon membranes. 48
2.8.6 Western blotting. 48
2.9 Matrix Attachment Assays 49
2.9.1 Testing the ability of end-labelled DNA fragments to bind isolated nuclear 49
2.9.2 Testing the retention of DNA restriction fragments to nuclear matrix in treated nuclei. 49
2.10 DNase I sensitivity assays 50
2.11 Polymerase Chain Reaction 52
2.11.1 General PCR practice 52
2.11.2 Oligonucleotides. 53
2.11.3 Purification of oligonucleotides 53
2.11.4 Touchdown PCR. 53
2.11.5 "Long" PCR. 54
2.12 Fluorescent in situ hybridisation 54
2.12.1 Preparation of metaphase chromosome spreads 54
2.12.2 Labelling of probes with dCTP-biotin. 55

Chapter 3 CLONING AND CHARACTERISATION OF DNA FLANKING THE C6 LOCUS

3.1. Introduction 56
3.1.1. Background 56
3.1.2. Potential approaches towards cloning flanking DNA. 57
3.1.3. Bacteriophage λ based genomic cloning 60
3.2. Results 62
3.2.1 Testing the Compatibility of the ColE1 Origin of Replication 62
3.2.2. Contruction of a partial EcoR1 C6 genomic library 63
3.2.3. Screening the library for clones containing plasmid sequence 72
3.2.4. Analysis of positive clone 74
3.2.5. Sequencing of the novel DNA in present in the 5.7kb fragment 79
3.3. Discussion 80

Chapter 4 CLONING AND CHARACTERISATION OF DNA FLANKING THE D8 LOCUS

4.1 Introduction 88
4.2 Results 92
4.2.1 Production of a partial Sau 3A library from D8 genomic DNA 92
4.2.2 Screening the partial Sau3A genomic library with a lysozyme-specific probe 93
4.2.3 Subcloning Strategy 98
4.2.4 Sequencing 101
4.2.5 Attempted Identification of the C6 and D8 loci by Fluorescent in situ Hybridisation 105
4.3 Discussion. 107

Chapter 5 ASSESSMENT OF THE FUNCTION OF THE SEQUENCES FLANKING THE INTEGRATED PLASMIDS IN THE C6 AND D8 CELL LINES 110

5.1 Assaying the DNase I sensitivity of the chromatin flanking the C6 and D8 loci 111
5.2 Assaying for matrix attachment regions 117
5.2.1 Testing the ability of long PCR products from the D8 clones to bind isolated nuclear matrix in vitro. 119
5.2.2 Assaying the binding of the flanking DNA of the D8 and C6 locus to the nuclear matrix of isolated nuclear halos. 125
Discussion 131
Future Work. 132

Chapter 6 ATTEMPTS TO USE THE SACCHAROMYCES CEREVISAE SITE SPECIFIC RECOMBINASE, FLP TO PLACE SELCTED DNA SEQUENCES AT THE ACTIVE LOCUS
6.1 Introduction. 134
6.1.1 Flp recombinase is a member of the integrase family of site specific recombinases 134
6.1.2 The mechanism of FLP recombinase 137
6.1.3 FLP recombinase is active in higher eukaryotic cells. 137
6.1.4 Placement of an FRT site at the active C6 locus 140
6.1.5 Flp activity was undetectable in J558L cell lines and extremely low in CV1 cell lines 142
6.1.6 Epitope tagging of proteins as a means of separating protein expression from activity 142
6.2 RESULTS 145
6.2.1 Placing the Myc-9E10 epitope coding sequence at the 3' end of Flp coding sequence 145
6.2.2 Flp recombinase is localised to the nucleus of transiently transfected COS cells. 153
6.2.3 Co-transfection with Flp recombinase markedly reducuces the expresion level of two second reporter genes. 157
6.3 Discussion 162
6.4 Future Work 168
Chapter 7 General Discussion 170
Chapter 8 References 172
Appendix 1 Primer sequences
Appendix 2 FASTA outputs

Abbreviations

ARS	Autonomous replicating sequence
ATP	adenosine triphosphate
bp	base pairs
BSA	Bovine serum albumin
CMV	Cytomegalovirus
DEAE	diethylaminoethyl
DMSO	Dimethylsulphoxide
dNTP	deoxyribonucleoside triphosphate
DTT	dithiothreitol
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
ECL	enhanced chemiluminescence
EDTA	diaminoethanetetra-acetic acid
ELISA	enzyme linked immunosorbent assay
FACS	fluorescent activated cell sorting
FCS	foetal calf serum
FISH	Fluorescent *in situ* hybridisation
FITC	fluorescein isothiocyanate
Flp	Flp (flip) recombinase
FRT	Flp recombinase target site
GFP	Green fluorescent protein
gpt	xanthine-guanine phosphoribotransferase
HBS	Hepes buffered saline
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HMG	high mobility group
HPFH	heretitary persistence of foetal haemoglobin
HRP	horseradish peroxidase
IgH	Immunoglobulin heavy chain
IMP	inosine monophosphate
IPTG	isopropyl b-D-thiogalactopyranidose
kD	kilo Daltons
LB	Luria Bertani
LCR	Locus control region
LIS	Lithium 3,5-diiodosalisylic acid
MAR	matrix attachment region

mRNA	messenger RNA
NLS	nuclear localisation signal
NMR	nuclear magnetic resonance
NP40	nonylphenyl-polyethylene
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenymethylsulphonylfluoride
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SEAP	secreted alkaline phoshatase
SINE	short interspersed nuclear element
SSC	standard saline-citrate
TAU	tris-acetate-urea
TBS	tris-buffered saline
TE	Tris-EDTA
TEL	telomere sequence
TEMED	N,N,N'-tetramethyl-ethylenediamine
Tm	Melting temperature
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Tween 20	polyoxyethylene sorbitan monolaurate
UPE	upstream promoter element
UTR	untranslated region
UV	ultra-violet
X-GAL	5-bromo-4-chloro-3-indolyl-β-D-pyranoside
YAC	yeast autosomal chromosome
XMP	xanthine monophosphate

Chapter 1
Introduction

1.1 Background to the project

The work presented in this thesis follows on directly from work by Kevin Hudson (Harrison, et al., 1993, Hudson, 1988). In the course of this work a substantial effort was made to generate an optimal expression vector for the stable expression of heterologous proteins in the mouse cell line J558L. This is a myeloma cell line from which the Ig heavy chain genes have spontaneously deleted. There are several reasons for wishing to express proteins in such a cell line:

Proteins are expressed in different systems for a variety of reasons. Vast quantities of heterologous proteins are used in diagnostic kits and biosensors as well as those expressed for structural analyses such as NMR and X-ray crystallography. In many cases expression in simple systems such as *E.coli* or yeast is sufficient for these purposes. However, in some cases it has been shown that post-translational modifications unique to higher eukaryotes such as sulphation and particular glycosylation patterns have been shown to affect the folding and activity of the protein. In these cases there is no alternative but to express the protein in higher eukaryotic systems. In these cases there are several reasons for chosing to express the protein in a mouse myeloma cell line.

Firstly for pharmaceutical purposes, all of the post-translational modifications that occur in heavily modified human therapeutic proteins, such as the clotting factors, occur in mice also. This is not the case in lower eukaryotes which have different glycosylation patterns, and is certainly not the case in bacteria, which have very few post-translational modifications. The fact that J558L is a mouse

cell line means that the risk of transfer of pathogens such as retroviruses to humans treated with proteins of potential clinical relevance is less than it would be if a human cell line were used.

Secondly, from a biotechnological viewpoint the technology for the large scale culture and product purification from this cell type is already well established for the production of monoclonal antibodies. Conversion of equipment and protocols for the production of other proteins should be very simple. Also, the cells have an extrordinary capacity to secrete proteins by a constitutive pathway. This could potentially be exploited for any heterologous proteins expressed, if necessary by incorporation of a cleavable signal sequence. As this work began in the mid-1980s, it pre-dated most of the work on transgenic and cloned animals, at the time when this work started, production of proteins *in vitro* using large scale cell culture technology was the most economical and practical way of producing proteins which could not be produced in yeast or bacteria.

The work by Kevin Hudson led to several conclusions. Firstly, it reiterated the fact that randomly integrated expression constructs are subject to potent position effects in that reporter gene expression varied markedly between different clones carrying the same plasmid (discussed below). Most of the quantitative work on the plasmid constructs generated was done in terms of stable transfection frequency. Transfection frequency reflects the transcriptional efficiency of a given construct in that it gives an indication of the number of transfectants expressing a selectable marker above a threshold level, such that the transfectant is able to grow under selective conditions. If we assume that integration into the genome is a random event, then each constuct will on average be subject to the

same position effects. The number of transfectants (ie surviving clones) therefore reflects the average efficiency of the expression constructs.

The main conclusion of the work was that there is an enormous increase in transfection frequency (in the J558L cell line) with constructs containing the IgH enhancer (figure 1.1). In most plasmid constructs, the IgH enhancer increased the transfection frequency by an order of magnitude. Figure 1.1 also shows considerable variation in promoter efficiency. As would possibly be expected the CMV and SV40 promoters gave best results but even these potent promoters were poorly effective in the absence of the IgH enhancer (Figure 1.1).

The study yielded a set of plasmids which were several fold more efficient (in terms of transfection frequency) than others. These plasmids have several features in common: Firstly they contained the IgH enhancer and CMV or SV40 promoter elements as discussed above. Secondly, they contain expression and selection cassettes in a divergent orientation, with the IgH enhancer positioned between the two promoters (Figures 1.2 and 1.3). Typically the selectable marker used in the plasmids was *gpt* (*E.coli* xanthine-guanine phosphoribosyltransferase)(Mulligan and Berg, 1981). This can be selected for by incubation of transfectants in medium containing methotrexate (to block the *de novo* purine synthetic pathway) and mycophenolic acid (to block the purine salvage pathway whereby IMP is converted to XMP). Under these selection conditions the only way in which cells can generate purines is by the phosphoribosylation of xanthine to XMP by *gpt*. Mammalian cells do not contain this enzyme, so only cells which express *gpt* above a threshold level survive.

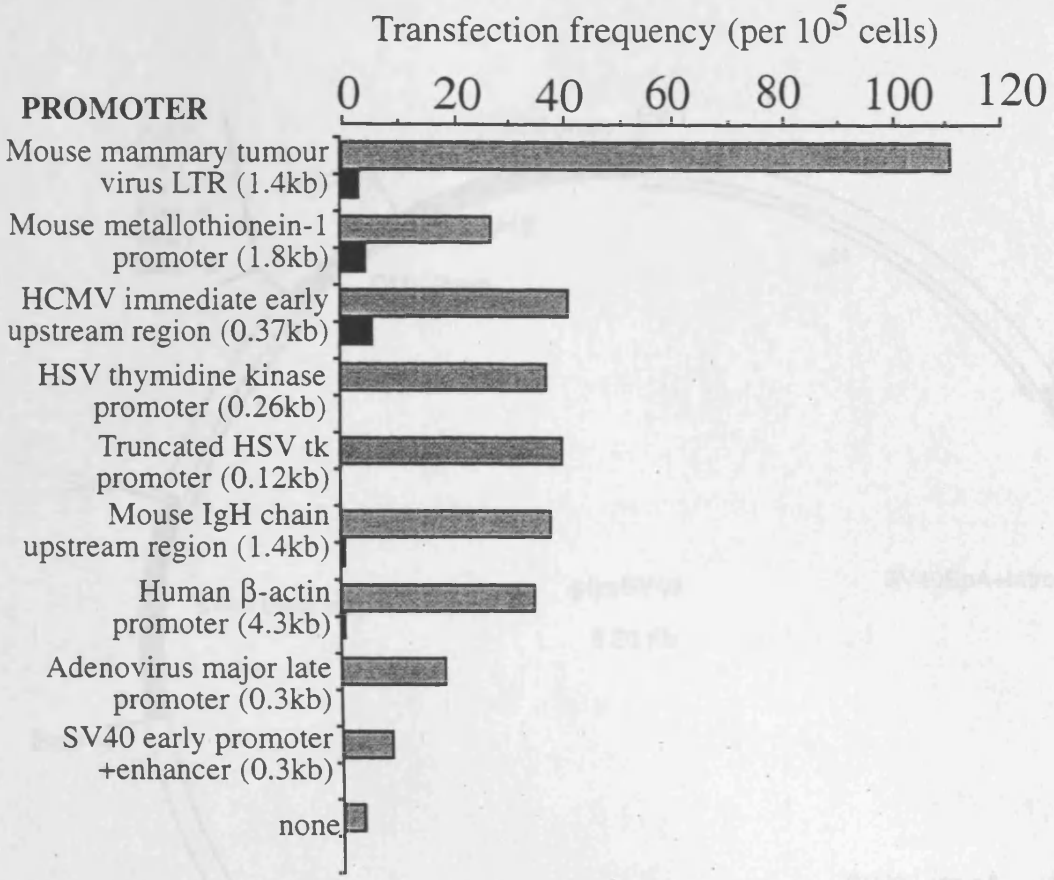


Figure 1.1

Bar Chart summarising data obtained by Kevin Hudson (Leicester University thesis 1988).

Plasmids were constructed containing a E.coli gpt gene (see text) driven by the promoters shown, with (light shading) and without (black shading) the IgH enhancer. Transfection frequency taken as the number of colonies surviving gtp selection per 10^5 cells transfected by electroporation.

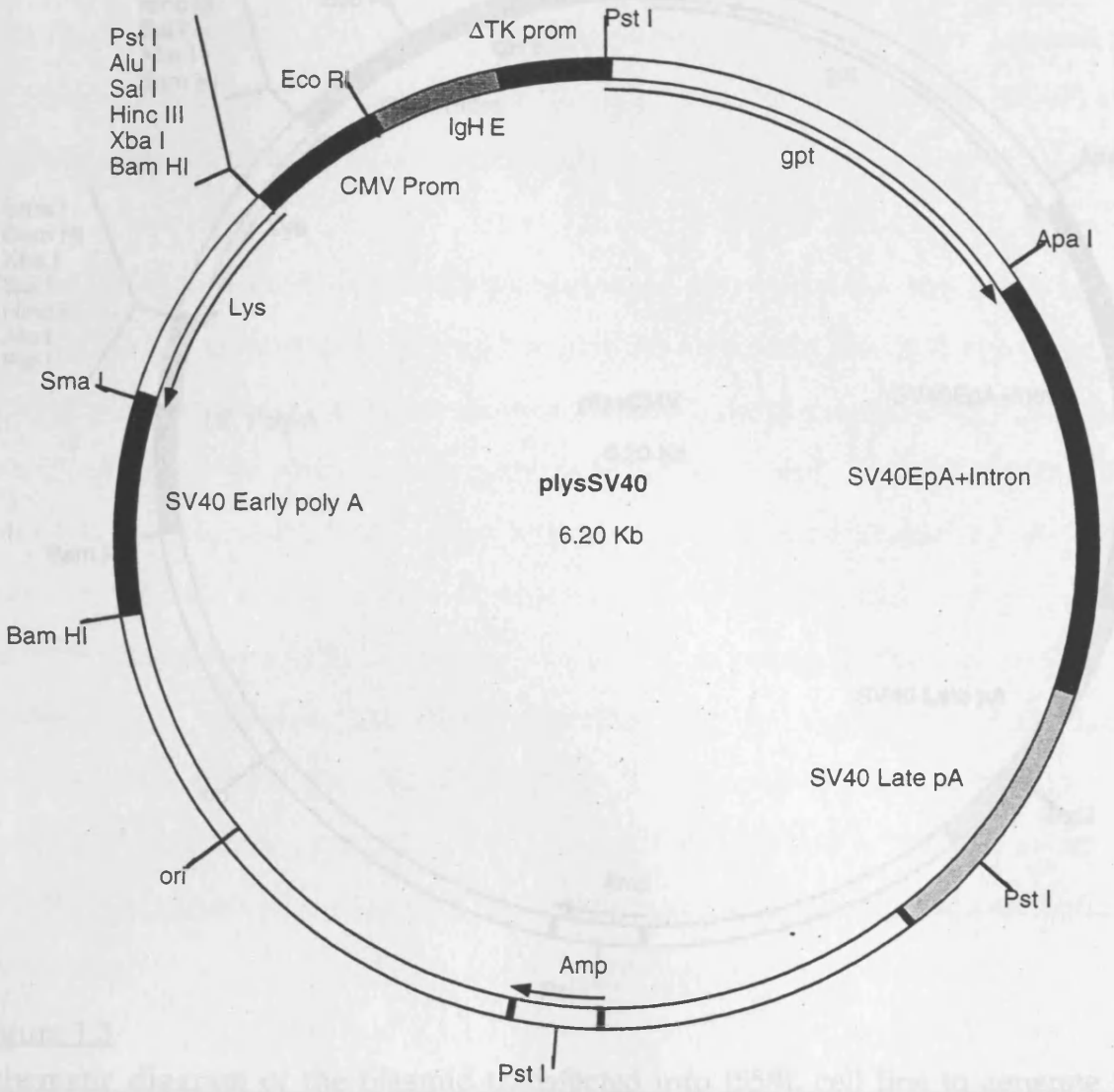


Figure 1.2

Schematic diagram of the plasmid transfected into I558L cell line to generate the C6 cell line.

The plasmid contains the gpt and lysozyme (lys) genes driven by a truncated thymidine kinase (ΔTK Prom) and CMV promoter (CMV Prom) respectively. These are arranged in a divergent orientation around the immunoglobulin heavy chain enhancer (IgH E). The lysozyme expression cassette contains the SV40 early polyadenylation site. The gpt expression cassette contains the SV40 early polyadenylation site plus intron, and the SV40 late polyadenylation site.

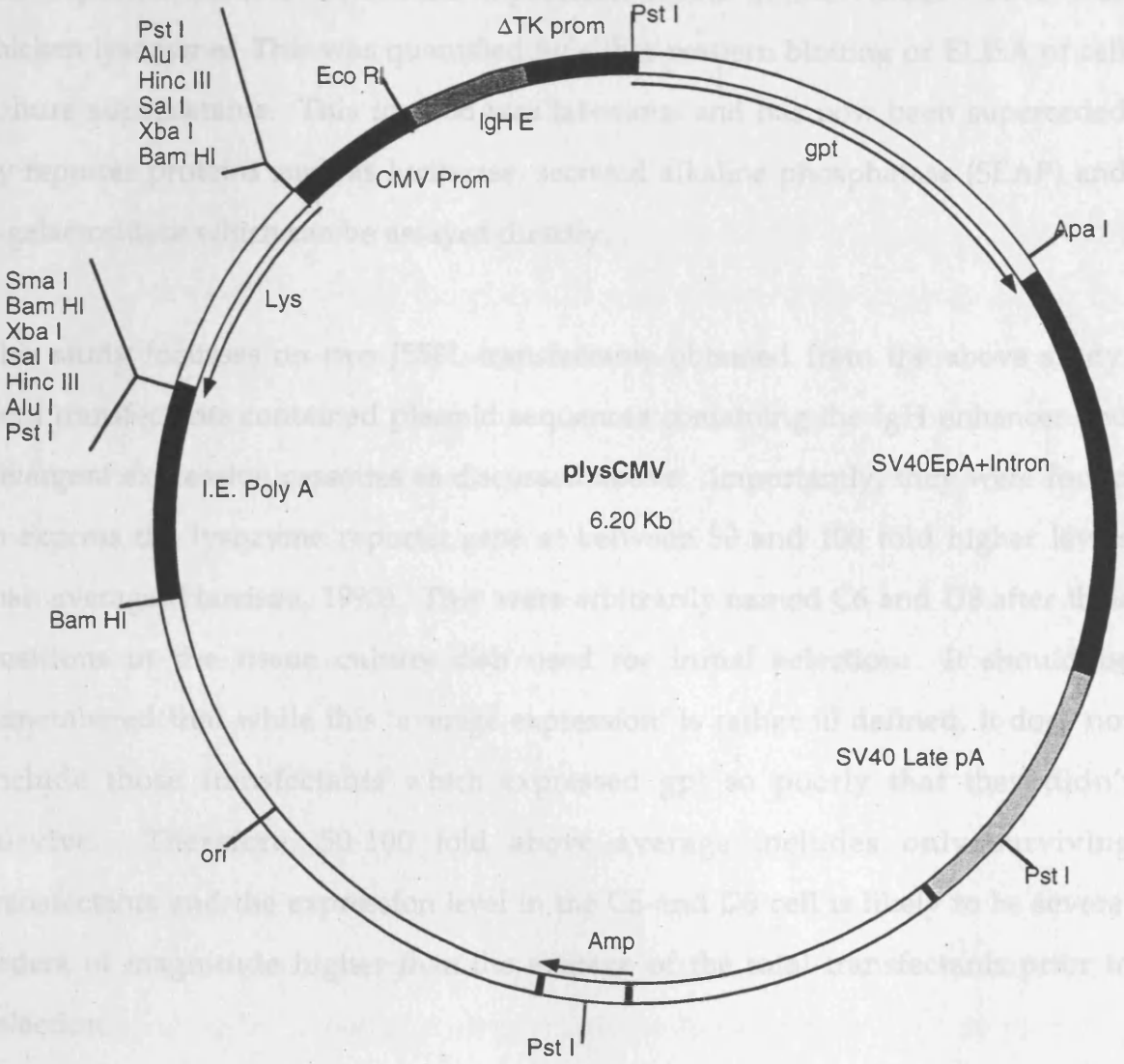


Figure 1.3

Schematic diagram of the plasmid transfected into I558L cell line to generate the D8 cell line.

The plasmid is identical to that in the C6 cell line apart from the polyadenylation signal after the lysozyme coding sequence. plysCMV contains the CMV immediate early (I.E. polyA) polyadenylation site. plysSV40 (figure 1.2) contains the SV40 early polyadenylation site.

The reporter used to determine expression levels in individual clones was chicken lysozyme. This was quantified by either western blotting or ELISA of cell culture supernatants. This method was laborious and has now been superceded by reporter proteins such as luciferase, secreted alkaline phosphatase (SEAP) and β-galactosidase which can be assayed directly.

This study focusses on two J558L transfectants obtained from the above study. Both transfectants contained plasmid sequences containing the IgH enhancer and divergent expression cassettes as discussed above. Importantly, they were found to express the lysozyme reporter gene at between 50 and 100 fold higher levels than average (Harrison, 1993). They were arbitrarily named C6 and D8 after their positions in the tissue culture dish used for initial selection. It should be remembered that while this 'average expression' is rather ill defined, it does not include those transfectants which expressed gpt so poorly that they didn't survive. Therefore, 50-100 fold above average includes only surviving transfectants and the expression level in the C6 and D8 cell is likely to be several orders of magnitude higher than the average of the total transfectants prior to selection.

Clearly there could be several reasons for the overexpression of the lysosyme reporter gene in the D8 and C6 cell lines. This could easily include gross spontaneous phenotypic changes in the cell lines. Translational control at the level of initiation, elongation, RNA splicing and mRNA stability are increasingly becoming recognised as major factors in the control of gene expression. However, nuclear run on experiments by Sarah Glassford (University of Leicester Thesis 1994) indicated that the increased protein expression level in both transfectants correlates with a similarly increased rate of

chicken lysozyme transcription. This clearly showed that the over-expression of chicken lysozyme in the C6 and D8 cell lines originated at the level of transcription.

The main focus of this study was therefore to isolate and characterise the DNA around the integration site of the plasmids and then to attempt to elucidate the cause of the high transcription levels in the two cell lines. Transcriptional initiation is influenced by a large number of factors at many levels of control. The first stages of transciption activation are still poorly understood but clearly involve many epigenetic factors which in turn influence DNA structure, DNA topology, and chromatin structure. There are also several cis-elements and other factors which could potentially facilitate such high transcription levels. These are discussed below.

1.2 Factors which affect gene expression at the transcriptional level

1.2.1 DNA Structure

The structure of DNA, as deduced by Watson and Crick in 1953, is that of two antiparallel strands of complementary polynucleotides that are paired into a right-handed helix, with a helical angle of around 36° per nucleotides pair and therefore around 10 nucleotide pairs per turn. This structure is now refered to as the B-form of DNA. Only relatively recently, two other major forms of DNA were discovered. A-form DNA is again a right-handed helix with a slightly shorter and wider helical turn than B-form DNA. While this form is present at extremely low abundance in eukaryotic double stranded DNA, this structure is more relevant in DNA:RNA hybrids (such in the priming of lagging strand DNA replication by Okasaki fragments) or RNA:RNA duplexes in RNA hairpins. The prevalence of A-form DNA in these hybrids is likely to be due to

the extra hydoxyl group on the ribose sugar of RNA making compaction of the helix to the B-form DNA less energetically favorable. Z-form DNA is a left handed helix with a turn density similar to that of B-form. Again it is generally rare within the chromosome but is more common in DNA containing tracts of alternate purines and pyrimidines (eg GCGCGC). Both these alternative forms of DNA will have different affinities for DNA-binding proteins and will therefore be expected to have widely different biological roles. Despite their relative rarity, these structures are thought to be important factors in general transcriptional control. (Miller, et al., 1985, Naylor and Clark, 1990, Tiesman and Rizzino, 1990)

A potentially more important consideration is the flexibility to generate minor alterations of B-form DNA. In the chromosome, bases can be tilted away from their ideal positions and there can marked alterations of helical turn angle. Similarly the DNA helix can also be bent. Around 200 nucleotide pairs would normally be required for a 90° turn. This length is considerably shorter in DNA containing A-rich tracts (eg.$[AAAAANNNNN]_{10}$)(Griffith, et al., 1986). Protein binding can also cause exaggerated bends in the duplex and is discussed later. An example of this protein-induced DNA bending is that caused by proteins containing a HMG domain (named after the HMG proteins which contain many such domains (Grosschedl, et al., 1994) This domain binds a 20 bp binding site (specific to the particular protein) in DNA and can distort the DNA molecule through approximately 130°(Grosschedl, et al., 1994, Wolffe, 1994). This inherent flexibilty of the DNA molecule can profoundly affect DNA-transcription factor interactions and therefore affect transcriptional activity. Studying the effect of transcription factor binding on the local structure of DNA is also extremely difficult. It is possible therefore, that DNA structure changes markedly around

the sites of transcription factor binding and these changes in structure could affect other factors involved in transcription initiation such as DNA melting.

1.2.2 DNA packaging

A single molecule of naked eukaryotic DNA would span the length of the nucleus several million times. It must therefore be efficiently packaged and reduced in volume. This is accomplished by the binding of histones. There are two types of histone proteins. The nucleosomal histone proteins, referred to as H2A, H2B, H3 and H4, are extremely basic (arginine and lysine rich) between 100-130kDa and are among the most evolutionarily conserved proteins in eukaryotes. Each of the four histones binds DNA as a dimer to form a nucleosome octamer. Under physiological conditions octamers are bound approximately every 200 bp. Around 146bp of DNA is in contact with the histine octamer which leaves a region of 'linker' DNA between each octomer, which is not in contact with the nucleosome (Klug 1982). *In vitro* experiments suggest that nucleosomes bind DNA at preferred points. They rarely fully dissociate under physiological salt conditions although it is becoming increasingly apparent that they can be moved or displaced slightly by modification (phosphorylation and acetylation of histone tails are discussed later), and by interaction with other proteins, such as transcription factors. The binding of nucleosomes causes the compacted DNA to appear as 'beads on a string' under electron microscopy (Paranjape, 1994) A further level of compaction is achieved by binding of H1 histones. H1 histones are bigger than the other histone types (220kDa) and although also extremely basic are far less conserved than other histones. H1 genes indeed appear to be missing altogether in *S. cerevisiae*. There is a subset of around 60 which can be substituted for H1. These include the High Mobility Group (HMG) proteins 15 and 17. Binding of H1 to linker DNA and the

nucleosomes allows the further compaction of DNA into a solenoid with a diameter of approximately 30nm (the 30nm fibre). As a 30nm fibre the average mammalian chromosome would still be around 100 times too long to fit into the nucleus. Certain, specific, regions of the genome are compacted further by folding of the 30nm fibre into a fibre approximately 100nm in diameter. The mechanism of this folding and the precise structure of the 100nm fibre are still undetermined.

The presence of histones and the folding of DNA into tight, highly-ordered structures has an effect on both the initiation and progression of transcription. Activation of trancription requires the disruption of chromatin structure, which occurs in a reverse heirarchy to that of nucleosome assembly described above. In many cases, such as the glucocorticoid receptor-mediated transcription, only removal of H1 from chromatin is necessary for transcription activation. Other studies in yeast have shown that dissociation of the H2A-H2B dimers is required. In yeast, this dissociation has been shown to be effected by a multisubunit general activator complex, a homologue of which will almost certainly be found in mammalian cells. Full dissociaton of the $[H3\text{-}H4]_2$ tetramer rarely occurs and transcription factor accessibility, and more importantly polymerase processivity are maintained by either the histone-DNA contacts moving away from the transcription factor binding site, or rotational positioning of histone-bound DNA so that the transcription factor can access packaged DNA(Vanholde, et al., 1992). Both of these processes require direct input of energy and are dependent on co-operation by other protein factors. Trancription factors have been shown to bind directly to histones, and histones may well take part in the well established cooperativity of transcription complex formation(Owenhughes and Workman, 1996)

1.2.3 Histone Acetylation and Methylation

Acetylation and methylation of the N-terminal tails of histones H3 and H4 (both of which processes are activated by several signal transduction pathways) have been shown to directly affect transcription activation. Although they have no role in DNA wrapping around the histone *per se*, the N terminal tails do bind to DNA. The N-terminal tails of histones are extremely basic and both modifications reduce their overall positive charge and thus weaken the ionic interaction of the tail with the negatively charged DNA (Turner and Oneill, 1995)

Acetylation of histone tails was shown to occur *in vivo* as early as 1981 (Perry and Chalkley, 1981). It has since been demonstrated *in vitro* that the histone tails can be acetylated on a variable number of lysine residues. The degree of acetylation correlates directly with the accessibility of DNA to TFIIIA (as determined by DNA footprint analysis; Lee, et al., 1993 This increase in accessibility is thought to be due to the dissociation of the N-terminal tails (which being highly basic are normally wrapped around the DNA helix) due to repulsion between the phosphate backbone of the DNA and the acetyl groups on the histone tail. Clearly all 'whole cell' experiments whereby histone acetylation is measured by TAU gel analysis (Lee, et al., 1993) only measures the average level of histone acetylation. Some histones at specific loci may be hyperacetylated leading to a large decrease in binding affinity or even total dissociation of the histone. The mechanism by which the acetylation machinery (i.e. the actual enzymes required and any additional factors required for their activation) is targeted to specific genes could involve additional cis sequences which might explain why many apparently redundant sequences are essential for correctly regulated expression of integrated transgenes.

The first stage in activation of transcription may involve the modification of chromatin to allow the transcription machinery access to the DNA. However, as described above, the degree of disruption and also the mechanism by which it is achieved varies markedly between diferent genes. In certain cases, with constitutively expressed genes, a relatively slow rate of nucleosome formation after DNA replication may allow transcription factors to bind DNA ahead of histones (Selker, 1990). In other cases histones appear to cooperate in the binding of transcription factors. In many cases, gross rearrangments of chromatin structure are required. These rearrangements would likely include histone modifications (acetylation and methylation) and dissociation of the H2A-H2B dimer, mediated by the SWI-SNF complex. Components of the mammalian SWI/SNF complexes are associated with both active chromatin directly and with the nuclear matrix(Reyes, et al., 1997) although the means by which they are targeted to active chromatin is undetermined. It seems likely that the targeting of the SWI/SNF complex and the histone acetylation machinery is specified and mediated by additional transcription factors.

1.2.4 Promoter and enhancer elements

All sequences transcribed by polymerase II contain a promoter element directly 5′ to the transcription start site. Promoter activity is absolutely dependent on the orientation and position of the promoter relative to the transcription start site. Promoters range in size from about 300bp (SV40 early promoter) to over 4kb (e.g. the human β-actin promoter) and are both necessary and sufficient for transcription of genes downstream of them. The specificity and efficiency of promoters is achieved by their modular nature. That is, they are composed of many interchangable modules and each module constitutes a binding site for a

specific transcription factor. Most promoters contain two distinct regions. The region proximal to the transcription start site, referred to as the core promoter, generally contains binding sites for general transcription factors such as TBP (TATA binding protein) which binds the TATA box. Upstream Promoter Elements (UPEs) contain binding modules for both tissue-type and temporally-specific transcription factors, and it is largely these elements which confer specifity to the transcriptional activation of particular genes.

Enhancer elements are similar in nature to promoters in that they are modular in nature and contain binding sites for constitutive and general transcription factors. The activities of promoter and enhancer elements are often interchangable, in that promoter elements can be moved relative to the transcription start site to act as enhancers and *vice versa*. Enhancer activity is not normally dependent on position or orientation relative to the transcription start site and they are commonly found both 5′ and 3′ to the transcription start site, and within introns, exons and downstream of the polyadenylation site.

The mechanism of enhancer action

As their name suggests, enhancers augment the transcriptional activity from their affected promoters. There are several models for the mechanism of this activity. The most widely accepted mechanism is one in which cooperative interactions beween transcription factors bound to the promoter and enhancer elements cause a looping of DNA such that the two elements become juxtaposed in space. This leads to a high, local concentration of transcription factors which efficiently recruit RNA polymerase II and associated transcription machinery to the transcription start site. The bending of the DNA between the promoter and

the enhancer may be assisted by the formation of A or Z form DNA, and by histones as discussed above.

Although this model is generally accepted, there are other possibilities. It is possible that some enhancers may operate via other mechanisms such as a local co-operative build up of transcription factors between the promoter and enhancer (commonly known as the stacking mechanism), a transcription factor induced change in DNA structure (particularly melting), or even transcription factors sliding between enhancer and promoter. (Ptashne, 1986, Ptashne, 1988)

It was initially proposed that enhancers increase the level of transcription *via* an increase in the rate of initiation of RNA polymerase II, and hence increasing the absolute level of transcription in each cell. However it has since been shown, by studies at the single cell level, that enhancers in fact increase the proportion of expressing cells at a given time but not the absolute level of expression of individual cells (Walters, et al., 1995). Enhancers should therefore be thought of as elements which increase the probability that in an individual cell a gene will be actively transcribed at a given time. When transcription is active, the level of transcription is independent of enhancer activity.

1.3 Locus Control Regions

Locus control regions (LCRs) are regulatory elements which confer position-independent copy number-dependent expression to linked genes in chromatin. Several LCRs have now been discovered and all of them appear to be tissue specific. The first reported LCR was the β-globin LCR (Grosveld, et al., 1987). This LCR contains four DNase I hypersensitive sites (see later) and is active over a locus of around 50kb. It mediates high level position-independent expression

of the β-globin genes and any other transgenes placed within it provided they are supported by active promoter and enhancer sequences (Assendelft, et al., 1989, Talbot, et al., 1989) As with all other LCRs the β-globin LCR has no effect on expression levels in transient expression (where plasmid sequence does not become integrated into chromatin). Therefore, although the β-globin LCR contains within it potent enhancer elements, its activity and mechanism can be distinguished from that of enhancers. Again, in common with all other LCRs discovered to date the β-globin LCR shows strong tissue specificity in that it is only active in cells of erythroid lineage. In addition to the four DNase I hypersensitive sites which are constitutively active in erythroid cells (but not present in other cells), the locus contains many sites which show sensitivity to DNase I at specific developmental stages. These sites have therefore been implicated in the developmental switching of the β-globin genes during development (Townes and Behringer, 1990).

1.3.1 β-globin gene switching

The β globin locus has for many years been a paradigm for the control of gene expression at the transcriptional level. As discussed above the β-globin genes are expressed in a tissue specific manner. More importantly, they are also expressed in a developmentally specific manner. The ε gene is expressed in the embryonic yolk sac, the Gγ and Aγ genes are expressed in the foetal liver and the δ and β genes are expressed in the adult bone marrow. The genes are expressed through development in the same order as they appear in the locus. This and other evidence from studies on patients with persistent expression of foetal haemoglobin (the haemoglobinopathy HPFH) whereby the γ genes fail to be switched off and continue to be expressed throughout adulthood) led to a model for gene switching where the promoters for each gene compete for the strong

positive influence of the LCR. The ε gene is closest to the LCR and therefore the effective concentration of transcription factors on the ε gene promoter, with respect to the LCR, is higher that that of trancription factors on the more distal genes. The switching from ε to γ expression requires silencing of the ε gene and this ε repression mean that the next available genes (or the next most concentrated with respect to the LCR) are the γ genes(Townes and Behringer, 1990). It is now thought that the cause of HPFH is the absence of γ repression (Grosfeld-Unpublished communication). This model is corroborated by the fact that reversal of the order the β-globin genes leads to expression of the β-gene in all tissues, including the fetal liver (Orkin, 1990)

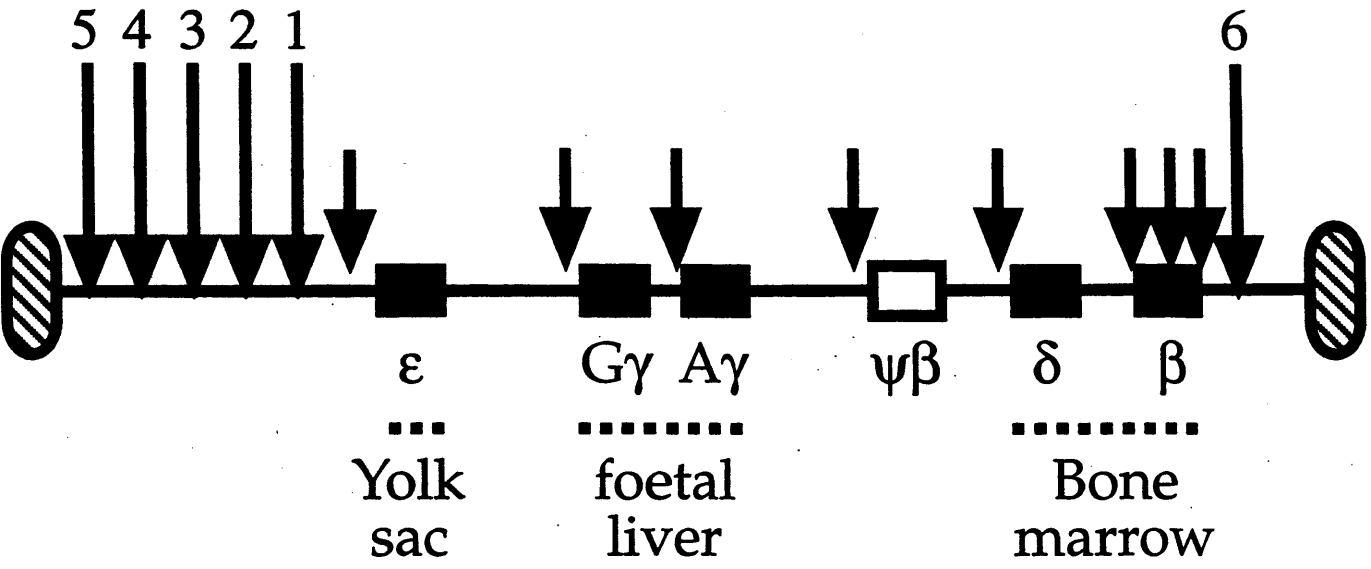


Figure 1.4

Schematic representation of the β-globin locus including coding sequences and the LCR. DNase I hypersensitive sites are depicted with large arrows. Developmental stage specific DNase sensitive sites are depicted by small arrows. Boundary regions for general DNase I sensitivity are shown by hatched ovals. The genes are expressed temporarily in the same order in the same order as they appear in sequence. The ε genes are expressed in the embyonic yolk sac, the γ genes are expressed in the foetal liver and the β and δ genes are expressed in the adult bone marrow. In mammals the entire locus spans approximately 50kb.

Since the discovery of the β-globin LCR many other *cis* DNA elements have been found to confer position independent, copy number dependent expression in other tissue types. The α–globin LCR is active in bone marrow (i.e. also erythroid) (Higgs, et al., 1990), the α(II) collagen LCR is active in fibroblast cells (Slack, et al., 1991). LCR-like activities have also been found flanking the CD2 gene in human T cells (Lang, et al., 1991), the MHC I locus in T cells (Chamberlain, et al., 1991) and the β-lactoglobulin (βLG) locus in mammary glands (Whitelaw, et al., 1992). High level, position-independent expression has also been demonstrated for the chicken lysozyme gene in monocytes (Bonifer, et al., 1990). As discussed below, this is of particular interest because position-independence in this gene has been correlated with the activity of matrix attachment regions at the locus boundaries (Phi-Van, et al., 1990).

1.4 DNase sensitivity as a determinant of chromatin structure

Sensitivity of chromatin to DNase I reflects an open chromatin structure. Hypersensitivity to DNase I (as determined by the presence of discrete band in Southern blotting after incubation of nuclei with low concentrations of DNaseI) was first described in the Drosophila heat shock genes (Wu, et al., 1978) and is now widely taken to indicate that one or more nucleosomes is missing from the chromatin. DNaseI hypersensitive sites have been mapped to a number of specific elements in DNA, including promoters, enhancers, silencers, replication origins, and structural sites in telomeres and centromeres (Gross and Garrard, 1988). Presumably, accessibility of the DNA to DNase I reflects accessibility to other proteins such as transcription factors and DNA replication factors. In the case of DNaseI hypersensitive sites associated with transcription (i.e. enhancers, promoters, and locus control regions) the DNase hypersensitivity is present

before the onset of transcription. This implies that the assembly of a transcription complex requires that the elements required are nucleosome-free. This is also supported by experimental evidence in the mouse β-globin major gene where the promoter is nucleosome free in active cells and inducible cells and is covered by a nucleosome in inactive cells (non-erythroid cells). *In vitro* the nucleosome bound promoter cannot be accesed RNA polymerase, but transcription can activated from the nucleosome free promoter (Benezra, et al., 1986, Lorch, et al., 1987, Losa and Brown, 1987).

Locus	Tissue specificity	Size of flanking DNA required (or distance from control element to transcription unit*	Reference
Human CD2 gene	T cells	5.5kb	(Greaves, et al., 1989)
Rat whey acidic protein (WAP)	mammary gland	949bp 5'and 70bp 3' of the gene	(Dale, 1992)
Human MHC I (HLA B7	T cells	660bp	(Chamberlain, et al., 1991)
Mouse a1(I) collagen	fibroblasts	2.3kb	(Slack, et al., 1991)
Mouse PEPcK	liver	4.8kb	(Cheyette, et al., 1992)
Chicken bA-globin	Erythroid cells	4.5kb	(Reitman, et al., 1990)
Rat LAP-(C/EBPb)	Liver/non-specific	2.8kb	(Talbot, et al., 1994)
Human a-globin	erythroid cells		(Higgs, et al., 1990)
Chicken lysozyme	macrophage	21.5kb	(Bonifer, et al., 1990)
Human Keratin	epithelia	2.5kb 5' and 3.5kb 3'	(Abe, 1990)
Human adenosine deaminase	T cells	*≈15kb	(Aronow, 1992)
Sheep β-lactoglobulin	mammary gland	406bp 5'	(Whitelaw, et al., 1992)

Table 1.1

Summary of 'LCR like' elements mediating copy number dependent position independent expression to genes linked to them. All of the elements are tissue specific except the Rat LAP-(C/EBPb) gene whose expression appears to slightly broader. The size of the construct or distance from the control element to the coding sequence suggest that the l based approach (giving clones between 9-23kb) is an appropriate approach to cloning the sequences flanking the plasmids in the C6 and D8 cell lines.

1.5 Matrix attachment regions

As discussed above, eukaryotic DNA is packaged into chromatin by association with histones and in some cases other non-histone proteins. A further level of organisation is achieved by interaction of the chromatin with the nuclear matrix. The nuclear matrix is a proteinaceous framework which spans the interior of the nucleus and binds to specific regions of chromatin. These regions of DNA (referred to as matrix attachment regions (MARS), scaffold attachment regions (SAR) or more recently S/MARS) interact specifically with the nuclear matrix and have several features in common: They are A/T rich and are frequently, but not always, found in the vicinity of transcribed sequences and, more specifically, elements which control expression such as enhancers (Cockerill and Garrard, 1986, Jarman and Higgs, 1988)

1.5.1 Common sequences in matrix attachment regions

There is no consensus sequence for a matrix attachment region as could be assigned to a transcription factor binding site. MARs do, however have several features in common. Firstly, as mentioned above, they are extremely A/T rich. Almost all contain several topoisomerase II consensus binding sites (GTN(A/T)A(T/C)ATTNATNN(G/A) (Sander and Hsieh, 1985) (Gasser and Laemmli, 1986). Most MARs also contain an A box [AATAAA(T/C)AAA] and a T box [TT(A/T)T(T/A)TT(T/A)TT].(Gasser and Laemmli, 1987) The function of the A box and the T box is still undetermined but the presence of a long stretch of A and T residues is likely to cause distortions in the DNA molecule and may facilitate bending of DNA.

1.5.2 The nuclear matrix

The many proteins which constitute the nuclear matrix have a variety of functions. Proteins with a mainly structural function include matrins (Belgrader, et al., 1991), lamins, and scaffold proteins (previously referred to as SCII) (Saitoh, et al., 1994). Proteins with a more regulatory function which form part of the nuclear matrix are HMG proteins 1 and 2 (Tremethick and Molloy, 1986) topisomerase II (previously refered to as SCI) (Berrios, et al., 1985). Many other proteins are associated with the nuclear matrix via their ability to bind directly to MARS. These include SATB1 (special AT-rich binding protein) (Dickinson, et al., 1992), MAR binding protein (Zong and Scheuermann, 1995), and nucleolin (Dickinson and Kohwishigematsu, 1995). Several transcription factors have also been shown to bind to the nuclear matrix (Dworetzky, et al., 1992) (Vanwijnen, et al., 1993) . This implies that the presence of a matrix attachment region near a promoter might influence the availablity of transcription factors, through their association with the nuclear matrix.

1.5.3 The effect of matrix attachment regions on gene expression

With the exception of a few well characterised systems, the effect of matrix attachment regions on gene expression is still largely undetermined. The best characterised effect of matrix attachment regions on gene expression to date is that of chicken lysozyme. The chicken lysozyme matrix attachment regions co-map exactly with the boundaries of the active chromatin domain, as mapped by DNase I sensitivity (Phivan and Stratling, 1988). These authors showed that when stably transfected into rat-2 cells constructs containing a chloramphenicol acetyl transferase (TC) reporter gene flanked on both sides by the chicken lysosyme 5′ MAR showed 10-fold greater activity per copy than a construct

containing the TC expression cassette alone. Variation in expression levels between transfectants also fell from 100-fold in constructs without the MARs to 6-fold in constructs with the MARs, implying a reduction in the negative effects (position effects) of the chromatin surrounding the integrated plasmids (Phivan, et al., 1990).

The 3'enhancer/MAR of the κ immunoglobulin gene is absolutely required for somatic hypermutation (Betz, et al., 1994). The precise role of the MAR in the mechanism is still undetermined, but the fact that there are known prefered mutation sites (Berek and Milstein, 1987) suggests that the MAR may play a role in either presenting specific regions to the transcription/mutation machinery, or in constraining the DNA strand while recombination takes place to prevent dissociation of strands during the recombination.

Despite their proposed involment in specialised roles such as that described above, it is unlikely that MARs alone can have a key role in the tissue specific or development stage specific activation of genes. There are very few reported cases where a particular MAR is has been identified as active under some conditions and not active in others. This is the only way that MAR could have any significant effects on gene activation. In transgenic mice MARs from the β-interferon gene flanking a reporter gene mediate position-independent, copy number-dependent expression in pre-inplantation embryos but not differentiated tissues (Thompson, et al., 1994).

As is the case with locus control regions, although under some conditions the β-interferon MARs stimulate transcription in stable transfection, they appear to antagonise transcription in transient transfection. (Klehr, et al., 1991). Thus as is

also the case with LCRs their activity can readily be distinguished from that of enhancers.

1.6 Other recombinant protein expression systems

1.6.1 Bacterial systems

The most common recombinant expression systems employ *E.coli*. The growth requirements of *E.coli* are well understood, as are its genetics and and regulatory elements. The design of expression vectors for the use in *E.coli* is therefore relatively straightforward. *E.coli* is cheap to grow and has a short life cycle (doubling time in log phase of growth is typically around twenty minutes) and recombinant protein product can represent up to 25% of total protein product (Old and Primrose, 1989). *E.coli* is therefore an economical and rapid means of producing large quantities of recombinant protein. However, this high level of expression of recombinant protein can cause the formation of inclusion bodies (Williams, 1982). These are aggregates between protein and nucleic acids and are both detrimental to the growth of *E.coli* and cause problems with protein purification. The recovery of proteins from inclusion bodies involves ther use of denaturing agents such as guanidine HCl and SDS. The proteins are then renatured by gradual dilution of these agents. Such a process is clearly not feasible or acceptable for proteins of pharmaceutical or large scale industrial use. The formation of inclusion bodies can often be reduced by the inclusion of an inducible promoter in the expression vector so that expression can be induced at the final stage of bacterial growth. Inducible expression is also useful if the recombinant protein is in some way deleterious to the growth of the cell.

There are also more fundamental problems associated with prokaryotic expression systems. Bacteria do not have many of the post-translational

modifications which are often required for correct folding and activity in eukaryotic proteins. The conditions within *E.coli* cells are reducing, which prevents the formation of disulphide bridges. Also, prokaryotic proteins are synthesised with an N-terminal formylmethionine rather than methionine. This residue can affect protein activity if not removed (Liang, 1985) and may be antigenic if the protein is used for pharmaceutical purposes.

1.6.2 Microbial Eukaryotes

Problems such as the absence of post translational modifications associated with expression of heterologous proteins in prokaryotic cells can often be circumvented by the use of yeast systems. Yeasts have been exploited for biotechnological purposes since ancient times in brewing and baking and hence, their genetics and biochemistry are well understood.

S. cerevisae is the yeast most commonly used for protein expression purposes. It is easy to manipulate genetically and can be grown in culture to high cell densities. The fact that it can be cultured in haploid form make the isolation of mutant strains and transfectants extremely straighfoward. Expression vectors in *S. cerevesiae* can be integrative or autonomous. Integrative vectors commonly integrate in a homologous manner which makes the position effects so commonly encountered in mammalian systems far less problematic in yeast. The construction of autonomous vectors such as YAC (Yeast artificial chromosomes) requires the presence of ARS (autonomous replicating sequence) and telomeric sequences (TELs). YAC vectors can accommodate inserts in excess of 1000kb and are stable throughout the growth of the yeast.

Although some heterologous control elements are active in yeast, efficient expression in yeast requires control elements specific to yeast. Promoters for these purposes are often derived from genes for glycolytic enzymes. Secretion of the protein efficiently also requires yeast specific secretion signal sequences such as those for invertase or α-factor (Brake, 1984, Chang, 1986)

S. cereviae can efficiently post-translationally modify proteins by phosphorylation and acylation. It can also glysoslylate proteins extremely efficiently. A major drawback with all yeast-based systems, however is that yeast only uses mannose residues in the core oligosaccharide. It is unlikely that such a minor difference would have a significant effect on protein folding or activity for most industrial uses, but for medical uses even such minor differences are not desirable due to their potential antigenicity. Therefore, when the proteins of interest require complex glycosylation, there is little alternative but to use a higher-eukaryotic expression system.

1.6.3 Higher eukaryotic and mammalian systems

COS cell expression

COS cells provide an efficient means of heterologous protein expression. The system is derived from early mammalian cell expression vectors based on SV40 where large T replacement vectors could be complemented for large T by cootransfection with native SV40. The SV40 plasmid carrying the transgene is therefore amplified to extremely high level within the cell with a consequent increase in expression level. This technique was used to express many foreign genes in monkey cells including rat preproinsulin (Gruss and Khoury, 1981) and influenza virus haemagglutinin(Gething and Sambrook, 1981).

A key development in this technology came with the discovery that the large T function could be provided by an integrated copy of a replication deficient SV40 virus in CV-1 cells (generating what are now refered to as COS cells) thus eliminating the need for co-transfection of helper virus. Integration of replicatable SV40 quickly renders the cell inviable and the replication activity was disrupted by a 6 base pair deletion in the origin of replication (Glutzman, 1981). Most modern vectors for COS cell expression contain only the SV40 origin of replication. Therefore no virus particles are produced at all and there is no lytic cycle as such, although the cell is still killed by the high level expression of foriegn protein. A major drawback with the COS cell system is the transient nature of the expression, and while the system remains extremely useful for research purposes, it is of little use for the large scale production of recombinant proteins

Baculovirus vectors and insect cells

The baculoviruses *Autographa california* and *Bombyx mori* nuclear polyhedrosis viruses have been exploited extensively for the expression of foreign proteins in insect cells(Smith, et al., 1983). In the normal infection cycle of the virus large amounts of the protein polyhedrin are produced from an extremely active promoter. In laboratory conditions the polyhedrin promoter is not essential for viral replication. Simple and effective expression vectors can therefore be constructed by inserting the foreign gene downstream of the polyhedrin promoter. Recombinant proteins can be expressed in either cultured insect cells, where expression lasts for around three days before cell lysis occurs(Miyamoto, et al., 1985)., or in silkworm larva where yeilds of up to 50μg per larva can be obtained. Although the high yields obtained are an obvious advantage, the size of the viral genome (130kb) renders it difficult to manipulate,

and in the case of expression in cultured cells, the transient nature of the expression gives the same limitations as expression in COS cells.

1.7 Aims of this study

The C6 and D8 transfectants are of interest for several reasons. For reasons described above, a system whereby proteins can be expressed efficiently in mammalian systems is of immense use from a biotechnological and pharmaceutical viewpoint. Although for the mass production of proteins *in vitro* systems have been superceded by transgenic animal technology, the small and medium scale production of recombinant proteins (for use in NMR and crystallographic studies) will still require systems such as the one that our laboratory aimed to achieve.

The two transfectants also have academic potential. The central aim of this study was to identify the factors responsible for the high level of expression in the two cell lines. The biotechnological applications, in terms of large scale expression of heterologous proteins, and the more academic study of the reasons for this high-level expression would both require the isolation of sequences flanking the integrated plasmids. It was shown previously that the high level of expression in both cell lines was transcriptional in origin and would therefore be most likely determined by the chromatin surrounding the plasmids. This flanking DNA could then potentially be incorporated into expression plasmids for use in the stable expression of proteins in myeloma cell lines.

Whichever direction the project was directed, a logical place to start was therefore the isolation of sequences flanking the integrated plasmids. Chapters 3

and 4 describe the isolation of DNA surrounding the C6 and D8 plasmids respectively.

Chapter 5 describes studies using the isolated sequences to determine the cause of the high level of expression. The chapter focuses on assays for matrix attachment regions and DNase I sensitivity assays.

Chapter 6 describes my involvement in work towards the use of FLP recombinase to target specific sequences to the 'active locus' in the C6 cell line. The principles and approach to this work are described fully in the introduction to chapter 6. The use of FLP recombinase was attempted in order to remove any concern at all as to the cause of the high level of expression and to attempt to focus entirely on the biotechnological uses of the highly expressing locus in the C6 cell line. However in the discussion section of chapter 6 some purely academic uses of the FLP system in the active locus are presented.

The central aim of this work was to isolate and identify the loci (or locus) into which these plasmids have integrated.

Chapter 2
Material and Methods

2.1.1 Chemicals

Unless otherwise stated chemicals were of AR grade, purchased from BDH laboratory supplies, Lutterworth, Leicestershire. UK), Fisons (Loughborough, Leicestershire, UK) or Sigma Chemical Company Ltd, (Poole, Dorset, UK). Reagents for bacterial cell culture were obtained from Oxoid (Unipath, Basingstoke, Hampshire, UK). Unless otherwise stated reagents for DNA manipulation and mammalian cell culture were obtained from Gibco BRL Life Sciences. All chemicals were handled in accordance with COSHH regulations and established good laboratory practice (GLP).

2.1.2 Bacterial Strains

The following bacterial strains were used in this study:

Strain	Genotype	Reference
TG1	*sup*E *hsdΔ*5 *thiΔ(lac-pro*AB) F'[*tra*D36 *pro*AB$^+$ *lac*Iq *lac*ZΔM15]	Gibson, 1984
MC1061	*hsd*R *mcr*B *ara*D139 Δ(*ara*ABC-*leu*)7679 D*lac*X74 *gal*U *gal*K *rps*L *thi*	Meissner et al., 1987
TG2	*SupE, hsdΔ5thiΔ(lac-proAB)Δ(sr1-recA)306: Tn10 (tetr), F'[traD36proAB+lacIq], laczDM15Tn10(tet ^{r})]*	Gibson, 1984
XL-1 Blue MRA	*Δ(mcrA)183, Δ(mcr CB-hsdSMR-mrr)173, endA1, SupE44, thi-1, gyrA 96, relA1, lacq*	Jerpseth et al 1993
XL-1 Blue MRA	*Δ(mcrA)183, Δ(mcr CB-hsdSMR-mrr)173, endA1, SupE44, thi-1, gyrA 96, relA1, lac^c, [P2lysogen]*	Jerpseth et al 1993

2.2 General DNA extraction and Manipulation

2.2.1 Handling of genomic and plasmid DNA

Wherever possible DNA was stored in TE buffer (10mM Tris HCl ; 1mM EDTA; pH8) rather than water, to protect from contaminating nucleases. When in use DNA solutions were kept on ice where possible. DNA solutions were stored at -20°C. DNA samples and items in direct contact with DNA samples were handled with gloves and in accordance with generally accepted good laboratory practice.

2.2.2 Quantification of DNA by spectrophotometry

Where necessary DNA was quantified by measuring absorbance at 260nm. An estimate of the purity of the DNA was obtained calculating the A_{260}/A_{280} ratio. (A_{280} is the peak extinction wavelength of the amino acids with aromatic side chains and therefore reflects the amount of protein present.) DNA was considered pure if it had an A_{260}/A_{280} ratio greater than 1.8.(Sambrook, et al., 1989)

2.2.3 Precipitation of DNA

Routine precipitation of DNA was carried out by the addition of sodium acetate to 0.3M followed by two volumes of absolute ethanol. The mixture was then cooled to -70°C for 15 mins and centrifuged at 13,000rpm in a standard benchtop microfuge. Where necessary an equal volume of isopropanol was substituted for 2 volumes of ethanol. For precipitation of small DNA fragments and oligonucleotides, sodium acetate was replaced with ammonium acetate (to 4M). Where small quantities of DNA were precipitated a co-precipitant of either E.coli tRNA or Pellet Paint (Applied Biosystems; as described in the manufacturer's instructions) were added to the solution before the salt. Pellet paint is a proprietary glycogen-based co-precipitant which gives the added advantage that the resulting DNA pellet is pink.

2.2.4 Isolation of genomic DNA

DNA was isolated from cultured cells by a previously established method (Sambrook, et al., 1989) with minor modifications. Breifly, cells were recovered from culture medium by centrifugation at 1500g for 10 minutes at 4°C, washed in an equal volume of ice-cold TBS, recovered again by centrifugation and

resuspended in TE (10mM Tris HCl pH^8, 1mM EDTA) to a final density of $5x10^7$cells/ml. 10 volumes of extraction buffer (10mM Tris HCl pH^8 0.1M EDTA, 20µg/ml pancreatic RNase, 0.5% SDS w/v) were added and the suspension was incubated at 37°C for 1 hour with occasional swirling. Proteinase K was added to a final concentration of 100µg/ml and the mixture incubated at 37°C for a further 3 hours. The mixture was extracted twice with phenol and once with chloroform and dialysed for 16 hours against >100 volumes of TE buffer at 4°C. DNA was then quantified by spectrophotometry and where necessary concentrated by ethanol precipitation as described in section 2.2.2.

2.2.5 Isolation of plasmid DNA by rapid mini-preps

For the purpose of screening bacterial colonies the classical method of alkaline-lysis followed by phenol choroform extraction proves extremely laborious and the use of proprietory kits becomes unduly costly when large numbers of colonies are screened. Thus, a crude but effective mini-prep protocol was employed exclusively for screening colonies. Colonies were picked into 5ml of media containing the appropriate antibiotic and incubated overnight at 37°C with shaking. 500µl of the resulting bacterial suspension were added to an equal volume of phenol/chloroform/isoamylalcohol (25:24:1), mixed by gentle inversion for 5 minutes, and then centrifuged at 13,000rpm in a microcenrifuge for 5 minutes at room temperature. 450µl aqueous phase was immediately added to an equal volume of isopropanol at room temperature. After gentle mixing DNA was recovered by centrifugation at 13,000rpm in a microfuge at room temperature. The DNA pellet was washed twice with 70% ethanol v/v, air dried for 5 minutes and resuspended in 50µl TE buffer. For colony screening purposes, plasmid DNA was generally digested with a single enzyme expected to give a different band pattern to the parent plasmid(s). Due to the presence of a

large amount of contaminating chromosomal DNA in plasmid DNA prepared by this method, once a potentially correct clone was identified, DNA was prepared by one of the cleaner methods detailed below for confirmation or subsequent manipulation.

2.2.6 Commercial mini-prep kits

Small scale preparation of DNA was performed using Promega "Wizard™" kit in accordance with the manufacturer's instructions. Briefly, cleared lysate was obtained from 3mls of overnight bacterial culture by standard alkaline lysis-neutralization (above). A proprietary resin (in 8M guaninidium isothiocyanate) was then added to the lysate, mixed thouroughly, applied to a supplied column and washed with a solution of 80mM potassium acetate, 8.3mM Tris-HCl, pH 7.5, 40μM EDTA, 55% ethanol. The resin was dried by centrifugation at 13,000rpm (microfuge) for 2 minutes and the DNA eluted with 50μls TE buffer (the resin binds DNA with extremely high affinity in buffers of high ionic strength and has low affinity in buffers of low ionic strength).

For later work miniprep kits were obtained from Qiagen. This is another resin based kit which functions by a similar principal.

2.2.7 Midi and Maxi preps

Large scale plasmid preparations were performed using a kit supplied by Qiagen. Again these kits utilize standard alkaline lysis-neutralization steps, followed by binding and elution of DNA from a proprietary resin , as described above.

2.2.8 Restriction digestion

Restriction enzymes were purchased from Gibco BRL. Enzymes were used in the buffer and digestion conditions recommended by Gibco and where double digestion was required the most compatible buffer for the two enzymes was chosen. If the two enzymes were incompatible the restriction digests were performed sequentially. (DNA being cut with the first enzyme, ethanol precipited and then cut with the second enzyme). To prevent non-specific activity the final glycerol concentration never exceeded 5% (v/v).

2.2.9 Heat Inactivation of restriction enzymes

Where necessary and possible, restriction endonucleases were heat inactivated according to the manufacturer's instructions, normally by heating to 65°C for 25 minutes. Enzymes resistant to heat inactivation were removed by phenol-chloroform extraction followed by ethanol precipitation of DNA.

2.2.10 Ligation of DNA

Plasmid ligations were carried out in standard ligation buffer supplied by the manufacturer (50mM Tris-HCl pH 7.5, 7mM $MgCl_2$, 1mM DTT, 1mM ATP, 5% w/v PEG 6000) at 16°C. Ligation was generally allowed to proceed for more than 16 hours and never less than 2 hours.

2.2.11 Tranformation of E.coli with plasmid DNA

E.coli cells were made competent by the Calcium Choride method (Hanahan et al 1983) as modified as described (Sambrook, et al., 1989)and were stored in aliquots at -70°C until required.

Cells were thawed rapidly and incubated on ice for 5 minutes. Ligated DNA was diluted 5 fold prior to addition to cells to reduce the inhibitory nature of PEG in the ligation buffer. DNA and cells were gently mixed and incubated on ice for 30 minutes. Heat shock was achieved by incubation at 42°C for exactly 2 mins followed by immediate placement on ice. Cells were allowed to recover by incubation at 37°C in non-selective medium for 30 mins and then spread onto appropriate selective media and incubated at 37°C for 16-20 hours.

2.2.12 Detection of α-complementation

For Blue-White selection of recombinant pBluescript (Stratagene) clones, plates were supplemented with 400pM IPTG and 0.004% X-Gal w/v. After incubation for 16 hours at 37°C the blue colour of non-recombinant colonies was enhanced by incubation for at least 2 hours at 4°C.

2.2.13 Band preparation from agarose gels.

Agarose gels were run and stained as described (Section 2.3.1). Bands were excised under low intesity uv illumination using a scalpel. DNA was extracted from the gel slice using a proprietory kit (Qiagen). Briefly, the gel slice was solubilised by incubation in Sodium Iodide (10%) at 50°C. The solution was then passed through spin columns, which bind DNA in a manner identical to that described for the mini-preparation columns (section 2.2.6).

2.2.14 5′ end labelling

DNA fragments were radio-labelled at the 5′ end by incubation with [γ^{32}P]ATP in the presence of polynucleotide kinase of (1X buffer was 100mM NaCl, 50mM KCl, 5mM $MgCl_2$) as described (Sambrook, et al., 1989).

2.3 Agarose Gel Electrophoresis

2.3.1 Standard Agarose Gel Electrophoresis

Unless stated otherwise, all gel electrophoresis was carried out in 1x E buffer (0.04M Tris, 5mM NaOH, 0.001M EDTA; pH7.8) using 1% agarose (Seekem HGT) gels. Gels were run at 8-10 volts/cm (between electrodes) and subsequently stained by soaking for 20-30 minutes in ethidium bromide (0.5µg/ml in water). Excess ethidium bromide was washed out by soaking the gel in water for 5-30 minutes. DNA was then visualised by illumination with UV light at 254nm.

2.3.2 Pulsed field gel electrophoresis

Pulsed field gels were run in 0.5x TBE (0.045M Tris-borate, 1mM EDTA; pH8) using 1% agarose (Seekem) gels. Unless stated otherwise gels were run at 10°C at 200V with a 0-5 minute pulse ramp over 5 hours. Gels were subsequently stained with ethidium bromide as described above.

2.4 Library Construction & screening

2.4.1 Partial digestion of C6 and D8 DNA

Conditions for the partial digestion of C6 and D8 DNA were determined by serial 1:1 dilutions of EcoR1 as described (Sambrook, et al., 1989) Optimal conditions for a modal fragment size of 20kbp were assayed by pulsed field gel electrophoresis against λH3 markers.

Once optimal conditions for partial digestion were determined, the reaction was scaled up to accommodate 50µg of DNA. Typical reaction volumes were around 1ml, so all components of the reaction were pre-warmed to 37°C before the enzyme was added.

2.4.2 Fractionation of partially digested DNA by centrifugation through sucrose gradients

After digestion DNA was loaded on to sucrose gradients (5-20% w/v sucrose in 1M Tris. HCl pH7.5, 5M NaCl, 0.5M EDTA) and centrifuged for 6 hours at 26,000rpm in an SW27 rotor (MSE) at 15°C. These condition were pre-determined by pilot centrifugation of bacteriophage Lambda DNA cut with HindIII (data not shown).

1.5ml fractions were taken by taking drops from a needle inserted into the bottom of the tube, and every third fraction was examined by pulsed field gel electrophoresis. Fractions containing DNA of 9-23Kb were pooled and precipitated with isopropanol. The subsequent pellet was dissolved in 500µl TE buffer and concentrated using Promega "Wizard" DNA clean up kit. This step also removed any residual sucrose which co-pelletted with the DNA DNA was finally resuspended in 30µl TE buffer.

2.4.3 Ligation of genomic DNA to λ–DASH vector arms

0.4µg of fractionated genomic DNA was ligated to 1µg of EcoR1-BamH1 digested λ–DASH (Stratagene) vector DNA in a total volume of 5µl according according to the manufacturer's instructions. Ligation was allowed to proceed for 16 hours at 4°C. Since efficient packaging of λ-DNA requires the DNA to be concatameric rather than circular, an aliquot of ligated DNA was routinely checked by agarose gel electrophoresis.

2.4.4 Packaging of λ-DASH libraries

Libraries were packaged in either Promega Packagene® or Stratagene Gigapack® II Gold packaging extracts according to the manufacturer's instructions.

2.4.5 Preparation of *E.coli* for plating of λ-DASH libraries

Cells were grown to an OD_{600} ≤0.6 in LB medium supplemented with 0.2% w/v maltose and 10mM $MgSO_4$. They were then pelleted by centrifugation 2000rpm (MSE) and resuspended to OD_{600}=0.6 in 10mM $MgSO_4$ and stored at 4°C for up to 24 hours.

2.4.5 Titring of libraries

Libraries and picked plaques were titered by plating serial 10-fold dilutions of packaged library on a lawn of XL-1 Blue [P2] *E.coli*. The merits of this type of titering when using λ vectors such as λ-DASH are discussed in chapter 3. 1µl of diluted library was mixed with 200µl *E.coli* (prepared as in section 2.4.5) and incubated for 20 mins at 37°C with shaking. The cells were then mixed with 3mls LB Medium containing 0.6% w/v Bacto-agar at 42°C, mixed gently and poured onto a 9cm petri dish containing a bottom layer of LB medium with 1.5% Bacto-agar, prewarmed to 37°C.

2.4.6 Primary screening of libraries

For primary screening the library was plated at 50,000 colonies per 150mm petri dish essentially as described above. 50,000 pfu were added to 600µl *E.coli* cells, prepared as describeed above, and plated in 8ml LB medium (0.6% agarose) over a prewarmed LB base (in 1.5% agar). After setting, plates were incubated for 10-16 hours at 37°C.

Plates were chilled for at least 2 hours at 4°C before plaques were transferred to duplicate nylon membranes (Amersham Hybond N). Transfer was allowed to proceed for 2 mins and 4 mins for the first and second filters respectively. Filters

were aligned using waterproof ink in a syringe needle. The membranes were then denatured by submerging for 2 mins in 1.5M NaCl, 0.5M NaOH, neutralized for 5 mins in 1.5M NaCl, 0.5M Tris-HCl (pH8) and finally rinsed for 30 seconds in 0.2M Tris-HCl (pH 7.5), 2xSSC. Membranes were air dried and fixed by exposure to UV light (254nm) for 20 seconds. Pre-hybridization, hybridization and washing were carried out as described below.
Putative positive plaques (duplicate positives only) were picked into 1ml SM buffer and rescreened.

2.4.7 Secondary Screening.

Picked plaques from the primary screen were titered and plated at a density of 1000 plaques per 150mm plate. Positive plaques were picked in 100µl SM buffer and further tested in a tertiary screen.

2.4.8 Tertiary Screening.

Picked plaques from the secondary screen were titered and plated at a density of 50 plaques per 90mm plate. Positive plaques were picked into 200µl SM buffer and stored at 4°C before being expanded for DNA preparation.

2.5 Purification of λ DNA.

2.5.1 Preparation of cleared lysate.

Isolated positive plaques were grown to high density overnight on a single 9cm plate. Phage were diffused into 3mls of λ buffer (10mM Tris HCl pH 7.4, 5mM MgSO4, 200mM NaCl, 0.1% gelatin), added to 250mls *E. coli* MRA strain (OD_{600}=0.2 in LB medium supplemented with 2% maltose and 10mM $MgSO_4$) and cultured overnight During this time the O.D.600 of the culture is expected to increase to around 1 and then fall back down to around 0.2 as the cells lyse.

10ml of chloroform were added to lyse infected but non-burst cells and the suspension was shaken for a further 30 minutes. DNase I and RNase A were added to a final concentration of 1µg/ml each and incubated at room temperature for 1 hour. The lysate was cleared of cell debris and chloroform by centrifuging at 6000rpm for 10 minutes (Sorval SS37 rotor). PEG_{6000} was added to the supernatant to a final concentration of 10% (w/v) and phage precipitated by standing on ice for 2 hours. Phage particles were then spun down (10,000rpm, 10 minutes in a Sorval SS37 rotor) and resuspended in 1ml of λ buffer. Excess PEG was removed by 2 extractions with an equal volume of chloroform.

2.5.2 Preparation of λ-DNA using Promega "Wizard™" Spin Columns

In the work described in Chapter 3, a modification of a method by Charnock-Jones et al (1994 Promega Profiles 18) was employed to recover λ DNA. This modification has been published, Walker and Harrison, 1994 Promega Profiles 19.

Residual bacterial genomic DNA and RNA were removed by passing the cleared lysate through a 1ml syringe containing 1g DE 52 cellulose pre-equilibriated with λ buffer. Bromophenol blue was used to monitor the adsorption of negatively charged matter to the cellulose. The column was washed with an equal volume of λ buffer. Guanidinium isothiocyanate was added to the eluant to a final concentration of 0.8g/ml followed by 1ml of Promega "Wizard" miniprep resin. The mixture was shaken at room temperature until the guanidinium isothiocyanate dissolved and then loaded onto a miniprep column. The resin was washed with 2mls 80% v/v isopropanol followed by 1ml column wash buffer (100mM NaCl, 10mM Tris-HCl pH 7.5, 2.5mM EDTA, 50% v/v ethanol). DNA was finally eluted in 100µl followed by 50µl TE (at 70°C). Yield was around 10µg phage DNA from 250ml lysed culture.

2.5.3 Purification using CsCl gradients

CsCl gradients were carried out as previously described (Sambrook, et al., 1989) Briefly, caesium chloride was added to the crude suspension of λ particles to a final density of 1.15g/ml. This mixture was then loaded on to a step density gradient of (from bottom of tube) 1.7g/ml, 1.5g/ml, 1.45g/ml CsCl (in λ buffer). Densities were checked to be accurate to 2 decimal places by refractive index analysis. After centrifugation at 22,000rpm for 2 hours at 4°C, phage particles were clearly visible as a blue band in the interface between ρ1.5 and ρ1.45. The band was collected by piercing the centrifuge tube and DNA prepared by essentially the same method as for genomic DNA (section 2.2.4). Phage particles were first dialysed against 1000 volumes TE to remove the CsCl.

2.6 Southern Blotting

Two methods of Southern blotting were employed:

2.6.1 Alkali blotting

After photographing, gels were depurinated by soaking for 5 mins in 0.25M HCl. They were then denatured in 1.5M NaCl, 0.5M NaOH. Capillary transfer to Hybond N+ nylon membrane (Amersham) was carried out overnight in the same denaturing buffer. Membranes were then rinsed briefly in 2xSSC, air dried, and fixed by 20 seconds exposure to uv light (254nm).

2.6.2 Blotting with 20xSSC.

For Fluorescein-based DNA detection (section 2.6.3) gels were depurinated and denatured as described above. They were then neutralized for 30 minutes in 1.5M NaCl, 0.5M Tris-HCl (pH8). Capillary transfer was carried out overnight in 20xSSC. Membranes were then air dried and fixed as described above.

2.6.3 Probe labelling with fluorescein

DNA fragments were labelled with fluorescein-congugated dCTP (Amersham) by the random hexanucleotide method in accordance with the manufacturers instructions.

2.6.4 Random oligonucleotide labelling with [α^{32}P] dCTP

10ng template DNA was heated at 100°C for 10 mins together with 0.1nmoles random hexanucleotides (Pharmacia). The solution was then snap-cooled by incubation on ice for 5 minutes. 2µl 10x OLB (0.5M Tris-HCl pH8, 50mM $MgCl_2$, 0.2mM dNTPs, 100µM β-mercaptoethanol, 2M Hepes pH6.6), 1µl BSA (1mg/ml), 2µl [α^{32}P] dCTP (0.37MBq/µl) were added sequentially with gentle mixing. The volume was made up to 20µl with sterile H_2O and 5 Units Klenow polymerase added. The solution was incubated at 37°C for 1 hour.

2.6.5 Purification of labelled probes using spin columns

^{32}P labelled probes were separated from free nucleotides using spin columns of Sephadex G50. A 1ml syringe was plugged with sterile polyallomer wool and filled with a suspension of Sephadex G50 in 1xTE buffer. The syringe was centrifuged for 5 mins at 100g. The volume of probe mix was increased to 200µl with TE and added the the top of the resin. The column was spun dry for 10 minutes at 1000g. Eluant was denatured by boiling for 10 minutes and cooled in ice water for 5 minutes before hybridization.

2.6.7 Prehybridization and Hybridization

For ^{32}P based hybridization techniques, pre-hybridization and hybridization were carried out in Church-Gilbert buffer (0.5M $NaHPO_4$ pH7.2, 7% SDS, 0.5mM

EDTA, 1%BSA) at 65°C. Hybridization was allowed to proceed for 16-24 hours. Unless otherwise stated, non-specifically bound probe was removed by washing the filters 2x30 mins in 3xSSC, 0.1%SDS at 65°C and then 1xSSC, 0.1%SDS at 65°c. Filters were stored at -70°C while exposing to X-ray film (Fugi). For fluorescein-based detection, hybridization and detection of bound probe were carried out exactly to the manufacturer's instructions.

2.7 Mammalian Cell Culture Methods

All mammalian cells were cultured at 37°C in a humidified atmosphere of 5% v/v CO_2.

2.7.1 Culture of J558L and derived cell lines.

J558L and derived cell lines (C6 and D8) were cultured in Dulbecco's modified Eagles medium (DMEM; Gibco) supplemented with 4.5g/l glucose and 10% foetal calf serum (FCS). FCS was obtained from various sources (mainly Sera Lab and Flow Lab and was batch tested before use. C6 and D8 cells lines were routinely cultured in gpt-selective media (1.44mM xanthine, 0.1mM hypoxanthine, 6µg/ml mycophenolic acid) to preserve the integrated reporter plasmids.

2.7.2 Culture of CV1, COS and derived cell lines.

CV1 and COS 8(Glutzman, 1981) cell lines were grown as monolayers in DMEM (4.5g/l glucose) supplemented with 10% FCS. CV#5 cells (which carries an integrated copy of pFRTNeo; see chapter 5) were routinely cultured in 400µg/ml G418 to preserve the integrated plasmid.

2.7.3 Transfection of COS and CV-1 cells.

COS and CV1 cells were transfected using a modified version of the method of Cullum et al., 1987. 5-50µg plasmid DNA were added to 1ml DEAE-dextran (1mg/ml) in TBS and heated to 60°C. The mixture was allowed to cool slowly to room temperature and added to 4mls Tris-buffered DMEM (4 parts DMEM : 1 part 0.25M Tris-HCl (pH7.5). Chloroquine was added to 100nM. Cells were washed briefly with serum-free DMEM. The DNA mix was poured onto the cells and incubated at 37°C for 6-8 hours. It was then removed by aspiration and the cells were shocked by covering with 10% DMSO in HBS for exactly two minutes. DMSO was quickly removed by aspiration and the cells rinsed thoroughly with serum-free DMEM. Cells were then cultured as normal in DMEM with 10% FCS.

2.7.4 Immunofluorescence.

For immunofluorescence, cells were cultured on coverslips which had been coated with poly-D-lysine by soaking for 1 hour at 37°C in a 1M solution in sterile water.

Cells were washed once with PBS (4°C) and then fixed for 15 mins in 3.7% formaldehyde in PBS at room temperature. After washing twice with PBS at 4°C cells were extracted in 0.2% Triton X-100 in PBS at room temperature. The cells were quickly washed twice further in ice cold PBS. Cells were incubated at 37°C with primary antibody diluted into TBS-T, 1% w/v BSA. Cells were washed with 3 changes of PBS over 30 minutes. Cells were then incubated with Texas Red-conjugated sheep-anti-mouse secondary antibody diluted (1:100) in TBS-T, 1% w/v BSA for 30 minutes and then washed as before with PBS.

Counterstaining for F-actin was achieved by incubating the cells with FITC conjugated phalloidin (Sigma) for 20 minutes at room temperature in the dark. After mounting in 50% v/v glycerol, 100mM Tris (pH7.4) and sealing with clear

nail varnish, cells were visualized on a Ziess Axiophot epifluorescence microscope. Photographs were taken on Iford HP5 Plus film (ASA 400).

2.7.5 Harvesting of transfected cells

Cells were harvested by scraping with a rubber policeman into 200µl per 90mm plate of lysis buffer (100mM potassium phosphate pH7.8, 0.2% Triton X-100, 1mM DTT, 0.2mM PMSF). Cell debris was removed by centrifugation at 13,000rpm in a benchtopmicrofuge.

2.8 Protein Methods

2.8.1 Determination of protein concentration

Protein concentrations were determined by the Bradford method using reagent purchased from Pierce. 30µl of various dilutions of cell extract was added to 1ml reagent and OD_{595nm} measured. Standard curves were obtained using dilutions of BSA between 0-400µg/ml.

2.8.2 β-galactosidase assays

β-galactosidase assays were carried out using the Galactolight Plus™ chemiluminescent system obtained from Tropix Inc (Massachusetts USA) following the manufacturer's instructions. 100ng total protein was added to each assay. Luminescence was initially measured using a scintilation counter, as described in the manufacturer's instructions, and later using a Bioorbit luminometer. Relative β-gal activities were calculated using the formula previously described (Fulton and Van Ness, 1993)

2.8.3 Luciferase assays

Luciferase assays were carried out using a commercial kit obtained from Promega, following the manufacturer's instructions. 10μg total cell lysate were used in each assay and luminescence was measured in a Bioorbit luminometer.

2.8.4 Separation of protein by SDS-PAGE electrophoresis

Unless otherwise stated, SDS-PAGE gels were 10% (w/v) polyacrylamide in resolving buffer (375mM Tris-HCl pH8.8, 0.1% (w/v) SDS). Stacking gel consisted of 4% w/v polyacrylamide in stacking buffer (125mM Tris pH6.8, 0.1% SDS). Polymerisation was induced by the addition of TEMED and ammonium persulphate to 0.06% (v/v) and 0.1% (w/v) respectively.

50-100μg cell lysate were boiled for 5 minutes in 1x Laemmli buffer (50mM Tris pH6.8, 4% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 0.005% bromophenol blue) before being added to the gel.

Gels were run at 40mA in 25mM Tris, 192mM glycine, 0.1% (w/v) SDS (pH8.3).

2.8.5 Electrophoretic transfer of resolved proteins onto nylon membranes.

Proteins were transferred onto HybondC$^+$ nylon membranes by the method of Tobin et al., 1979. Transfer buffer consisted of 25mM Tris, 200mM glycine, 20% (v/v) methanol. Transfer was carried out at 85V in for 90 minutes with cooling. Transfer was checked by soaking the membrane for 5 minutes in 10% Ponceu-S solution and rinsing briefly with water.

2.8.6 Western blotting.

Membranes were blocked with 5% (w/v) powdered milk in TBS-T buffer (10mM Tris-HCl pH8, 0.9% (w/v) NaCl, 0.1% (v/v) Tween-20). They were then incubated with the appropriate dilution of primary antibody (diluted in the same blocking buffer) for 1.5 hours before washing with 3 changes of excess TBS-

T over 30 min. Horse radish peroxidase-conjugated secondary antibody (Amersham) was then applied (again diluted 1:1000 in blocking buffer) for 40 mins. The filter was again washed with 3 changes of excess TBS-T over 30 mins. HRP was detected using an ECL based system made and generously supplied free of charge by Dr Mike Murray (Dept. of Genetics, University of Leicester).

2.9 Matrix Attachment Assays

2.9.1 Testing the ability of end-labelled DNA fragments to bind isolated nuclear matrix *in vitro*.

The ability of ^{32}P end-labelled DNA fragments to bind extracted nuclear matrix in vitro was assayed as exactly as previously described (Cockerill and Garrard, 1986)

2.9.2 Testing the retention of DNA restriction fragments to nuclear matrix in treated nuclei.

Retention of DNA fragments on the nuclear matrix of isolated nuclei was assayed by the method first described by Mirkovitch et al (1984) with a few minor modifications. Approximately $1x10^8$ cells were washed twice in excess PSB (4°C) and then three times with isolation buffer (3.75mM Tris-HCl pH7.4, 0.05mM spermine, 0.125mM spermidine, 0.5mM EDTA/KOH pH7.4, 1% (v/v) thiodiglycol, 20mM KCl). They were then resuspended in ice-cold isolation buffer containing 0.1% digitonin and broken in a Dounce homogenizer (20 cycles). Nuclei were pelleted by centrifugation at 900xg (4°C) and washed once in isolation buffer. They were then resuspended in 100µl isolation buffer containing 0.1% digitonin and lacking EDTA. (To reduce protease activity all buffers during and after the washing stage were supplemented with 10,000IU Aprotinin and 0.1mM PMSF.) The nuclei were stabilised by heating to 37°C for

20 mins. Histones were extracted by gradual addition of 7ml extraction buffer (5mM Hepes/NaOH pH 7.4 0.25mM spermidine 2mM EDTA/KOH pH7.4, 2mM KCl, 0.1% digitonin, 25mM lithium 3,5-diiodosalisylic acid (LIS)) and incubation for 5 minutes at room temperature. Extracted nuclei were then recovered by centrifugation at 2400g for 20 minutes. LIS was then removed by repeated washing in excess digestion buffer (20mM Tris-HCl pH7.4 0.05mM spermine, 0.125mM spermidine, 20mM KCl, 100mM NaCl, 10mM $MgCl_2$, 0.1% digitonin). Nuclei were resuspended in a convenient volume of digestion buffer for a time-course digestion and restriction enzymes added to a final concentration of 1000U/ml and incubated at 37°C with gentle shaking. Aliquots of the digestion reaction were removed at various time intervals and solubilized DNA separated from matrix-bound DNA by sedimentation at 2400xg for 10 mins at 4°C.
DNA was extracted by overnight incubation with 0.5% SDS and 100µg/ml proteinase K. The solution was then extracted twice with equal volumes of phenol and once with choroform. DNA was then precipitated, digested to completion with the same enzyme and quantified by extinction at 260nm. Where posssible equal quantities (5µg) of pellet and supernatant DNA were run on a 1% agarose gel and Southern blotted as described above.

2.10 DNase I sensitivity assays

DNase I sensitivity assays were performed essentially as described previously (Siebelist, et al., 1984) with minor modifications. Firstly, cells were washed with ice-cold PBS rather than Hanks balanced salt solution. Cells were lysed by resuspension of the cell pellet in 10% saponin (rather than NP40) in nuclear isolation buffer: 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 1mM EGTA, 15mM Tris HCl (pH7.4), 5mM dithiothrietol, 1mM PMSF, 2M Sucrose. Nuclei were isolated by centrifugation through 2M sucrose (in nuclear isolation buffer) at 12,000 rpm

in an SW27 (Sorval) rotor for 20 mins at 4°C. Nuclei were resuspended in nuclear isolation buffer supplemented with 5% glycerol. DNase I was added to the nuclei as described in Chapter 5. Just prior to addition of DNase I the buffer was supplemented to 5mM $CaCl_2$, 1mM $MgCl_2$.

2.11 Polymerase Chain Reaction

2.11.1 General PCR practice

To avoid contamination between plasmid positive controls and genomic PCR, and to prevent new PCR reactions becoming contaminated with PCR products from previous reactions, filtered pipette tips were used and wherever possible PCR reactions were set up in a different area from that where they were analysed. A negative control in which no template DNA was added to the reaction was always included even where none is shown. All PCR reactions were Hot-Started, to prevent the extension of spuriously-primed template, by the addition of polymerase when the reaction had reached 94°C in the denaturation step of the first cycle.

Unless otherwise stated PCR was carried out with "Red Hot" Taq polymerase (Applied Biosystems). PCR reactions were carried out in a total volume of 10µl. The denaturing step was 94°C for 30 secs. The annealing step was the Tm for the oligonucleotide, as calculated below (section 2.11.2), for 1 minute. Extension was at 72°C for 30-180 secs depending on the size of the desired products. The total number of cycles was between 15 and 30, and PCR products were 'cleaned up' with a final incubation at 72°C for 20 minutes. This final incubation has been shown both to enhance the integrity of the PCR products, as defined by the sharpness of the bands obtained in agarose gel electrophoresis, and to enhance the efficiency of A/T based cloning of PCR products.

2.11.2 Oligonucleotides.

Oligonucleotides were synthesized as a service by the Protein and Nucleic Acid Chemical Laboratory at the University of Leicester using an Applied Biosystems, model 394 machine. They were delivered as crude stocks in dilute ammonium hydroxide with the 3' DMT group pre-removed.

The Tm of the oligonucleotide was calculated empirically as the sum of individual base interactions. The Tm of a single G:C and A:T bond was taken to be 4°C and 2°C respectively.

2.11.3 Purification of oligonucleotides

100µl 8M ammonium acetate was added to 200µl crude oligonucleotide stocks followed by 600µl absolute ethanol. The solution was mixed by vortexing and cooled to -70°C for 30 minutes. Precipitated oligonucleotide was pelleted by centrifugation in a benchtop microfuge for 15 minutes at 13,000rpm. The pellet was washed twice with a large excess of 70% ethanol at room temperature and dried by vacuum desiccation. The pellet was resuspended in 100µl H_20 and a 10µl aliquot quantified spectrophotometrically. The oligonucleotide concentration was determined using the formula:

$$\text{concentration(M)} = \frac{OD_{260} \times 1 \times 10^{-5}}{10 \times \text{length}}$$

2.11.4 Touchdown PCR.

Where multiple products were obtained or smearing was apparent after agarose gel electrophoresis, a touchdown PCR protocol was employed (Don, et al., 1991) The annealing temperature is dropped by 1°C every 1 or 2 cycles until the

calculated Tm is reached. The rationale behind this method is that correct binding of the primers at the correct site will occur at a higher temperature than spurious binding and therefore when the calculated Tm is reached, the correct product will have a head start in the exponential amplification over spurious products.

2.11.5 "Long" PCR.

Long-range PCR was carried out using kits purchased fron either Takara (Expand™) or Boehringer Mannheim (Expand™), following the manufacturer's instructions. Both of these kits employ a mixture of Taq polymerase and the proofreading thermostable polymerase *Pwo* and are supplied with a range of buffers to facilitate rapid optimisation of conditions.

2.12 Fluorescent *in situ* hybridisation

2.12.1 Preparation of metaphase chromosome spreads

C6 and D8 cells were split at least 24 hours prior to harvesting for the production of metaphase spreads.

Colcemid (Sigma) was added to culture medium to a final concentration of 0.05mg/ml and the cells were placed at 37°C for 1 hour. Cells were pelleted by centrifugation at 1100rpm for 10 minutes at room temperature and resuspended in 0.075M KCl (prewarmed to 37°C). The cell suspension was incubated at 37°c for a further 12-15 minutes. Cells were then pelleted as above and fixed by the dropwise addition of Carney's solution (glacial acetic acid:methanol 1:3) at 4°C. The cells were incubated at 4°C for 10 minutes, pelleted as above and resuspended in fresh Carney's solution and kept at 4°C for 1-2 hours.

Microscope slides were washed in Carney's solution and dried with tissue paper. To make the cell spreads, the microscope slides were held at an angle of approximately 45° while a drop of the cell suspension (maintained on ice and sealed from moisture as far as possible) was dropped onto the slide from a height of approximately 1 metre. The slide was held at an angle until all of the solution had evaporated. The optimum density of cells in Carney's and the height from which the cells were dropped were determined empirically, by examination under phase microscopy, and varied markedly depending on atmospheric humidity. Useful metaphase spreads appear mat-black in phase contrast microscopy. The presence of a halo around them indicates the presence of cytosol which renders the chromosomes inaccessible to probe. After drying, metaphase spreads were stored at 4°C for 1-4 weeks before being denatured , neutralised, and dehydrated.

2.12.2 Labelling of probes with dCTP-biotin.

Probes for FISH were labelled with dCTP-biotin by the random hexanucleaotide method using a kit purchased from Amersham. Labelling was checked and hybridisation carried out according to the manufacturer's instructions.

Chapter 3
Cloning and characterisation of DNA flanking the C6 locus.

3.1. Introduction

3.1.1. Background

When reporter genes integrate into the genome of mammaliam cells the position effect that they succumb to is normally negative. That is, the level of transgene expression obtained is generally less than that obtained when the gene is expressed in a transient, non-integrative manner and, where the comparison is possible, at a far lower level than that of the endogenous gene. In the case of transfection of plasmids containing the IgH enhancer into plasma cells, the average level of transgene expression is several orders of magnitude lower than that of the IgH genes.

This facet of the position effect reflects the fact that in any given cell type only around 5% of genes are expressed and that only a fraction of the genome represents coding sequence. Thus, if for the purposes of this argument we assume non-homologous integration to be a random event, there is relatively little chance of reporter genes integrating into a region of the genome which is actively expressed. It is therefore extremely rare that transfectants such as the D8 and C6 cell lines, in which expression of the reporter gene is comparable to expression of endogenous imunoglobulin genes, are obtained.

Isolation of the sequences flanking the plasmids in such cell lines is important for several reasons. Firstly, there are academic questions which can be answered by cloning and obtaining sequence data from the the flanking sequences. Sequence data should yield information as to whether the plasmid has integrated into a known locus. Potentially, the effect of the cloned sequences on transfection efficiency and expression levels can then be assayed with a view to determining any enhancer or LCR activity within the cloned sequences. Secondly, if the flanking sequences were found to mediate higher level or position-independent expression to genes around them, they could potentially be useful biotechnologically. The technology for the large-scale culture of hybridoma/myeloma cells lines is already extremely sophisticated due to the demand for monoclonal antibodies. It would not be difficult to transfer the technology developed specifically for monoclonal antibody production to the production of recombinant proteins in similar cell lines. Thus, isolation of the flanking sequences from the C6 and D8 cell lines was the primary aim of this project.

3.1.2. Potential approaches towards cloning flanking DNA.

Several rapid approaches towards obtaining DNA flanking the C6 plasmid have been attempted. These include plasmid rescue, inverse PCR, and anchor mediated PCR (Sutcliffe 1995; Glassford 1993)

Plasmid Rescue

Plasmid rescue is a potentially extremely rapid and efficient means of isolating DNA flanking integrated plasmids. It is entirely dependent, however, on the plasmid's origin of replication and antibiotic selection marker remaining intact during all manipulations. DNA is digested with a suitable restriction enzyme

such that a substantial amount of flanking DNA is present on the same restriction fragment as the plasmid's origin of replication and antibiotic selection marker. The digest is then incubated with DNA ligase in dilute conditions and in the absence of polyethylene glycol, to favour intra-molecular rather than inter-molecular ligation. The ligation mixture is then transformed into competent *E.coli* which are then plated onto suitable selective media.

Theoretically, the only way in which colonies can be obtained is by the re-circularisation of the restriction fragments containing the plasmid origin of replication together with the flanking DNA. While in theory this technique should provide a convenient and rapid method of isolating flanking DNA, there are several potential limitations. Firstly, if the plasmid sequence is present as a single copy in large genomes such as that of the mouse, its relative abundance in a ligation of total genomic DNA is extremely low. Therefore so much DNA would need to be ligated and transformed into *E.coli* that the purely chemical nature of DNA as a polyvalent anion is likely to reduce the transformation efficiency. This was demonstrated by experiments by Sarah Glassford showed that the presence of 10μg genomic DNA in a standard bacterial transformation (Materials and Methods) caused a fall in transformation efficiency of 10ng plasmid DNA from $1x10^7$ cfu/μg (without genomic DNA) to $1x10^5$ cfu/μg. This dramatic fall in efficiency of transformation renders the approach of plasmid rescue very difficult. Even without these considerations the transformation efficiency required for this technique needs to be on the limit of present technology. Most importantly, the length of flanking DNA that can realistically be obtained by this method is relatively low since the efficiency of the technique would arguably fall in proportion to size of the fragment to be rescued, due to the

fact that larger fragments are more difficult to circularise and transfect into *E.coli*i.

Several variations of this method have been attempted, without success, by previous workers attempting to clone the flanking DNA around the plasmids in the C6 and D8 cell lines. The technique is also entirely dependent on the plasmid origin of replication and antibiotic selection marker remaining intact and funtional throughout the integration and selection process.

Inverse PCR

Many of the difficulties encountered with plasmid rescue also apply to the technique of inverse PCR. Here, divergent primers are raised to known sequence and the genomic DNA is again cut with restriction enzymes such that the known sequence and flanking DNA are present on the same restriction fragment. The DNA is then ligated in conditions favouring circularisation rather than concatenation (conditions discussed above). If the restriction fragment of interest is successfully circularised, then the primers which were divergent in the linear duplex DNA molecule then become convergent within the circle and a PCR product can be formed, encompassing all of the DNA between the 3' ends of the primers and the restriction sites. The length of DNA which can be amplified by such a method is fairly short and the considerations in terms of required efficiency and available technology which apply to plasmid rescue and anchor-mediated PCR (discussed below) apply equally to this method. Again this method has been attempted unsuccessfully by previous workers in the laboratory.

Anchor PCR

Another potentially rapid method which was tried without success was that of anchor-mediated PCR (Pfeifer, et al., 1989). Exhaustive variations of this approach were attempted by previous workers (S Glassford; N Suttcliffe; University of Leicester Theses 1993; 1995). In the most basic variation of this method the DNA is again digested with appropriate restriction enzymes such that plasmid and flanking DNA are present on the same restriction fragment. Short oligonucleotide linkers are then ligated to the restriction fragments and PCR is carried out using primers specific to both the linker oligonucleotide and plasmid sequence. Although the only product to be exponentially amplified should be the plasmid and flanking sequence, there is always a problem with background amplification in such appproaches, due to the non-specificity of the primer to the oligonucleotide sequence. Again, the efficiency of the ligation and PCR reaction demanded for the the success of this approach were at the absolute limit of the technology available at the time, and indeed commercial kits based on this approach have only recently become available.

3.1.3. Bacteriophage λ based genomic cloning

For the reasons described above it was decided early in the project that such attempts towards obtaining flanking sequence from the C6 and D8 cell lines should, if possible, be avoided. Even with the substantial advancement of required technology since these methods were attempted, such as the advent of super-competent *E. coli* cells and refined Long-PCR systems, the experiences of previous workers suggested that an alternative approach was required.

It was shown previously that the most promising technique to clone these apparently recalcitrant flanking DNA sequences was that of phage Lambda based cloning (N Sutcliffe; Leicester University Thesis, 1996). While being considerably more time consuming and laborious than the techniques described above, λ cloning has the advantage that the clones obtained will be considerably larger than those which can reasonably be expected from plasmid rescue or anchor-mediated PCR or inverse PCR described above.

λ replacement vectors allow the cloning of DNA fragments of between 9 and 23kb. Table 1.1 shows that the size of a typical LCR locus (or the flanking DNA around the gene necessary for position independent expression) is typically far less than 20kb. Thus, given that Lambda replacement vectors contain between 9-23kb, if the loci of interest contain a 'typical' LCR, then at least some of the clones from the C6 and D8 cell line would be expected to contain the control element (or elements) of interest. The commercial vector, λ–DASH (Stratagene) was chosen for the production of a C6 genomic library. λ–DASH is a replacement vector based on the early λ2001 strain (Karn, et al., 1984). Like all lambda replacement vectors λ–DASH contains a stuffer fragment, between the two inverted multi-cloning-sites which is non-essential for phage replication under laboratory conditions. The λ–DASH stuffer fragment contains the genes *red* and *gam*. The *red* gene products (exo and bet) are involved in recombination during the conversion between early (θ-form) and late (rolling circle) DNA replication. The *gam* gene products block the activity of exonuclease V, encoded by the host genes *rec B* and *rec C*. Exonuclease V normally inhibits late bacteriophage λ replication by degradation of the concatenated DNA molecules produced by rolling circle replication (Enquist and Skalka, 1973).

red$^+$ *gam*$^+$ phages are inhibited by the presence of P2 phage. Thus plating the packaged library on a *recA*$^+$ P2 lysogenic strain facilitates a one step elimination of all non recombinant λ–DASH. However, the fact that the *red* and *gam* genes are missing means that all DNA replication is dependent on the host's *REC A* protein and therefore recombinant λ–DASH cannot grow in *RecA*$^-$ strains. Their dependence on their host's *Rec* system renders λ–DASH and relatated vectors extremely inefficient in the absence of a *chi* site. Therefore, to prevent massive over-representation of recombinants containing, by chance, a consensus chi site, a chi site has been engineered into the right arm of λ-2001 and most related vectors.

λ–DASH also contains within each multicloning site, a *Not I* restriction site. *Not I* has an 8 base recognition site that, together with a base composition that is rare in mammalian genomes, means that it rarely cuts mammalian genomic DNA. Thus entire clones can normally be exised from the λ arms using this single restriction enzyme.

3.2. Results

3.2.1 Testing the Compatibility of the ColE1 Origin of Replication

A potential problem with the λ-based cloning of this locus is that any clones obtained by this method are likely to contain the pBR322 origin of replication (Figure 3.1). The compatibilty of the pBR322 origin with REC A mediated λ replication is not documented and therefore, given the ease of the experiment and the dire prospect of working towards a high-titre genomic library from which almost all desired clones were by definition excluded, this compatibility was tested before construction of the library began. The plasmid pSV-Vμ1 (Neuberger, 1985) was cut with *EcoR1* to give a linear fragment of 18kb. This was

ligated to λ–DASH arms, precut with *EcoR1* and *BamH1*. *Eco RI* and *Bam HI* restriction sites are adjacent in the DASH multi-cloning. Digestion with *Eco RI* and *Bam HI* therefore leaves incompatible ends on the λ arms and stuffer fragment (the small fragment between the two restriction sites is lost in ethanol precipitation). Therefore, this procedure should, in theory at least, eliminate the re-formation of non-recombinant λ–DASH.

The ligation mix was transformed by heat shock into competent *E.coli* XL–1 MRA [P2] and plated with a feeder layer of cells (Materials and Methods). This yielded an efficiency of transformation of about one order of magnitude less than normal (10^6pfu/ng DNA would be expected in transformation of plasmid DNA into competent cells). This reduction in efficiency can be accounted for entirely by the size of the transformed DNA (≈50kb). The plaques were of normal size which suggests that the pBR322 origin does not interfere in any way with RecA mediated λ–DASH replication. The presence of pSV-Vμ1 in the clones was tested by probing with pBR322 sequence. Since pSV-Vμ1 contains the pBR322, plasmid backbone,as do the plasmids in the C6 and D8 cell lines, some of the plaques obtained with the pSV-Vμ1 probe were picked into SM buffer and retained for use as positive controls in the screening of the C6 and D8 libraries.

3.2.2. Construction of a partial *EcoR1* C6 genomic library

The restriction map of the locus generated by Southern bloting (S Glassford; University of Leicester Thesis 1993) revealed that the plasmid in the C6 cell line had integrated as a single copy, and also showed *Eco RI* sites on both sides of the plasmid (figure1.2). A partial *Eco RI* genomic library would therefore be expected to contain clones spanning the whole *Eco RI* restriction map, and therefore clones from both sides of the plasmid (Figure 3.2)

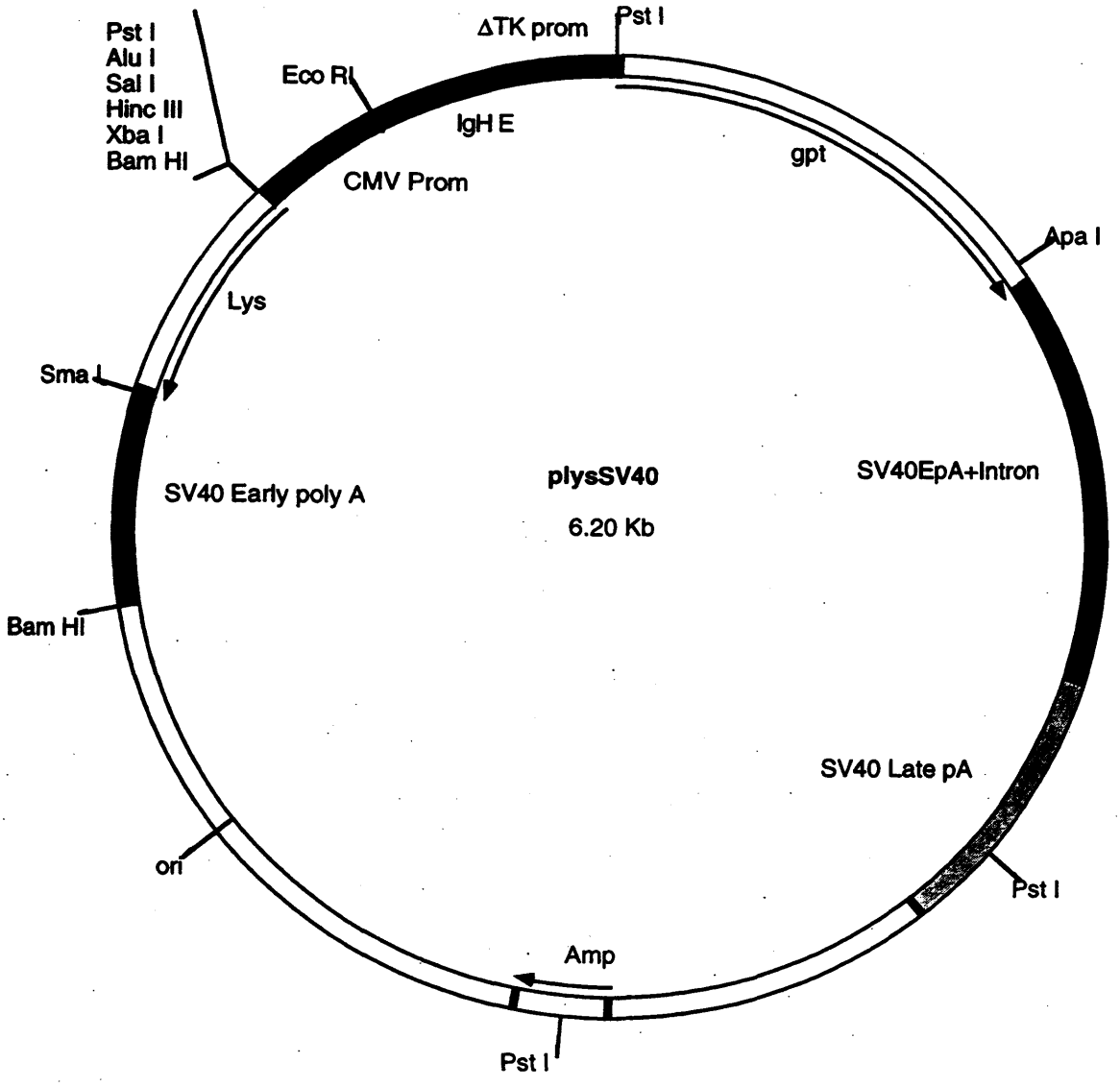


Figure 3.1
Schematic diagram of the plasmid transfected into the C6 cell line. Reproduced from Figure 1.2 (Introduction).

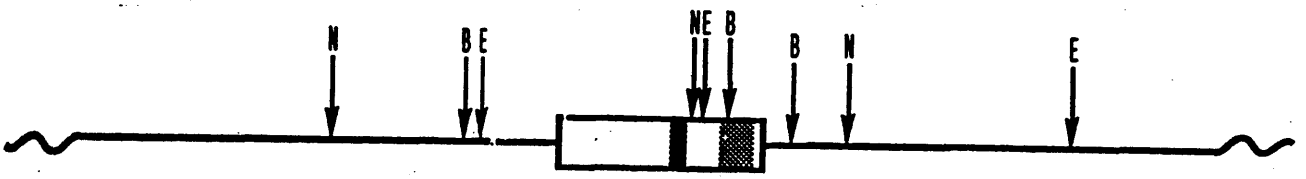


Figure 3.2
Restriction map of the C6 locus as determined by Southern hybridisation.
Light and dark shading represent gpt and lysozyme coding sequence respectively.
E=*Eco RI*, B=BglII, N=NcoI. (Courtesy of S Glassford; Leicester University Thesis 1993.)

Conditions for the digestion of C6 DNA with *EcoRI* to give a modal fragment size of 20kb were determined empirically by digestion of 1μg of C6 DNA with serial two-fold dilutions of *EcoR1*. Digestion was allowed to proceed for one hour and then stopped by the addition of EDTA. The digests were then run on an agarose gel by pulsed field electrophoresis. The digestion conditions giving the modal fragment size closest to 20kb were determined as 1/6 unit of *EcoRI* for 1μg DNA in a total volume of 30μl (Figure 3.3). These conditions were scaled up stoichiometrically to accommodate 50μg (DNA and enzyme concentration were kept constant). A 1μg aliquot of the digest was tested by pulsed field gel electrophoresis before ligation to λ arms.

As with the pSV-Vμ1 ligation the λ–DASH was precut with EcoR1 and BamH1 (Figure 3.5). In this initial experiment, the ligation ratio was 4μg C6 DNA to 1μg λ–DASH DNA. This and minor modifications of this ratio failed to give a sufficiently high packaging efficiency to produce a representative genomic library. The most likely cause of this was reasoned to be the preferential ligation of fragments smaller than 9kb. The partially restricted C6 DNA was therefore further enriched for DNA of 9-23kb by centrifugation through sucrose gradients.

Optimum centrifugation conditions for the separation of 9-23kb fragents from larger and, more importantly, smaller fragments were determined by pilot runs with *HindIII* digested λ DNA. Conveniently, this contains fragment of 9.1 and 23.0kb. Optimal separation conditions were determined as 26,000rpm for 6 hours at 15°C in a Sorval SW27 through gradients of 5-20% sucrose (data not shown). 50μg of partially digested C6 DNA was loaded onto each of two gradients.

Fractions of 1.5mls were collected and 500µl aliquots from every third fraction were precipitated and analyzed by pulsed field gel electrophoresis (Figure 3.4A)

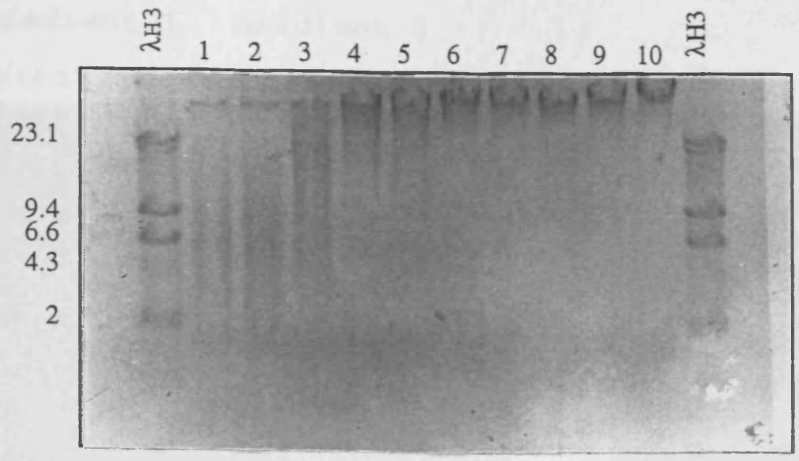


Figure 3.3

Determination of optimal digestion conditions to give a modal restriction fragment size of 20kb. Serial 1:1 dilutions of 2U *Eco RI* were made in tubes 1-9. Tube 10 is a zero enzyme control. 1µg C6 DNA was present in each dilution in a total volume of 30µl. Digestion was allowed to proceed for exactly 1 hour and then stopped by addition of excess EDTA. The gel shows that optimal conditions for modal fragment size of 20kb are between tubes 3 to 4. This represents 1/6 units of enzyme for 1µg DNA in a total volume of 30µl. (Markers sizes on left: kb)

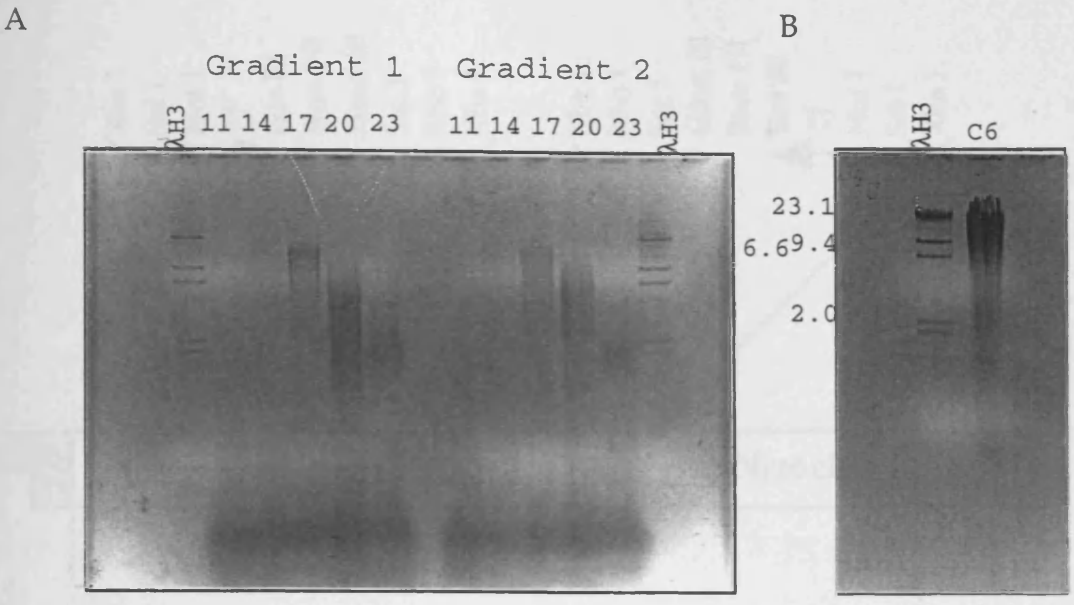


Figure 3.4A

Fractionation of partially digested C6 DNA on sucrose gradients.

DNA was digested as described in figure 3.3 and sedimented through a sucrose density gradient (5-20% sucrose, Sorval SW27 rotor, 24Krpm for 6 hours at 15°C). Gradients were subsequently fractionated into 1.5ml aliquots. 500µl of every third fraction was precipitated with isopropanol and run on a pulsed field agarose gel (200V 1-4sec ramp over 5 hours).

Figure 3.4B

Fractions 14-18 were pooled, precipitated with isopropanol and further separated from residual sucrose using a DNA spin column (materials and methods). The C6 track represents 2/25µl of recovered, size selected DNA.

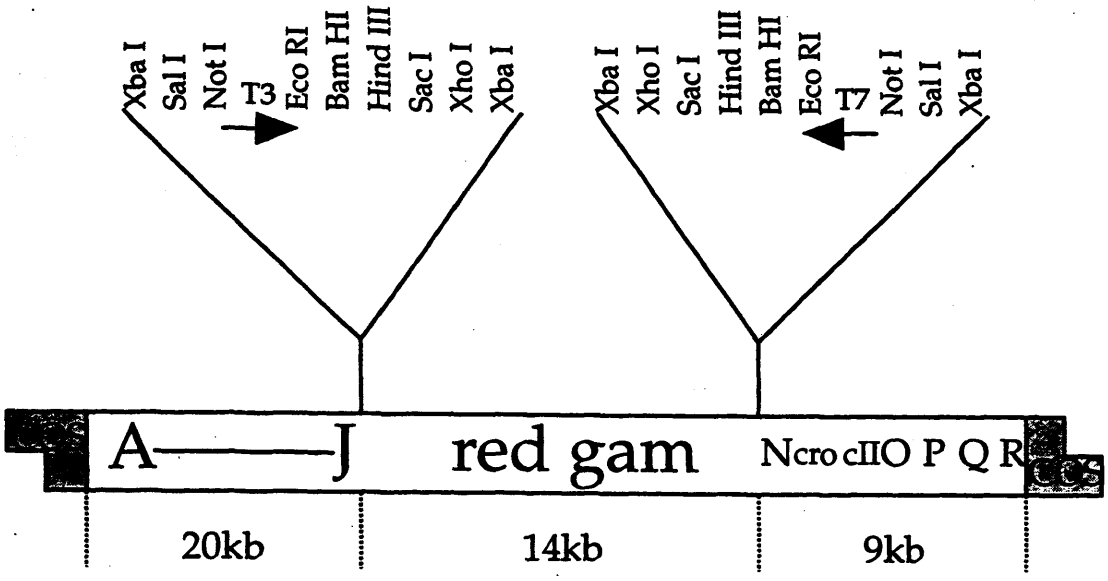


Figure 3.5

Genomic map of λ–DASH replacement vector.

Replacement of the 14kb stuffer fragment leads to the *red*$^-$ *gam*$^-$ phenotype, which can be selected for by plating the phage on P2 lysogenic *E.coli* host (see text).

Double digestion with the desired restriction enzyme and one close to it on the polylinker generates a fragment short enough to be lost in ethanol precipitation. This renders the stuffer fragment non-ligatable to the λ–arms and therefore eliminates the need for them to be purified prior to library construction.

Note that the promoters are omitted from the map for clarity.

(Figure is reproduced with slight modification from Stratagene catalogue 1994)

Fractions shown to contain fragments between 9-23kb were pooled and precipitated with isopropanol. DNA was further purified away from residual sucrose by concentration on a spin-column (Materials and Methods). Figure 3.4B shows an agarose gel analysis of this recovered material. The ligation was repeated with this size selected DNA, again using a 1μg:4μg ratio of λ:Genomic DNA. After ligation, the DNA was packaged using Packagene™ packaging extract (Promega) and titered as described (Materials and Methods) A library of $3x10^5$ independent clones was obtained. An indication of the potential success of screening such a library can be determined using the following equation (Clarke, 1976):

$$N = \frac{\ln(1-P)}{\ln\left(1-\frac{1}{n}\right)}$$

N represents the number of independent clones in the library, n is the ratio of total genome size to average fragment size and P is the probability that a given DNA sequence is present. Rearrangement of this equation for P and assuming the modal insert size to be 20kb and the total size of the diploid C6 genome to be $4.6x10^5$kb, gives a 99.99% probability that a given single copy sequence would be present in the library of $3x10^5$ independent recombinants. It is normal practice in genomic cloning to consider only the size of the haploid genome, because normally the sequence of interest with be present on both alleles. In this case, however, the plasmid will only have integrated into one allele so the size of the diploid genome must be allowed for. It should also be emphasised that the genome of the J558L cell line is likely to be bigger than the expected $4.6x10^5$kb due to the tendency of the cell line to take on an unstable pseudotetraploid state.

This would reduce the probability slightly, although it is unlikely that such an equation would ever be used as more than a guide as to how representative a given genomic library would be.

Since previous workers had shown that the locus was so recalcitrant to plasmid rescue and PCR based cloning techniques, it was decided that the library should not be amplified. It is well documented that library amplification leads to differential representation of clones, since some clones grow more slowly than others due to the size and nature of the insert. It is possible, therefore, that the amplification process could lead to any clones from the C6 locus which were present in the original ligation being underrepresented or effectively eliminated by the amplifiction process. The packaged library was therefore plated directly onto a lawn of XL–1 MRA [P2] *E.coli* (Materials and Methods).

3.2.3. Screening the library for clones containing plasmid sequence

The library was screened with fluorescein-labelled pTK2lys plasmid. This probe would hybridise clones containing either pBR322 or lysozyme sequence. In each round of screening the efficiency of DNA transfer to nylon membranes and hybridization was checked by spiking a single plate with a known number of pfu containing pSV-Vμ1 insert. Screening of the entire library of $3x10^5$ clones revealed a single positive plaque (figure 3.6). This was picked into SM buffer and and purified further in secondary and tertiary screening as previously described (Materials and Methods; results not shown).

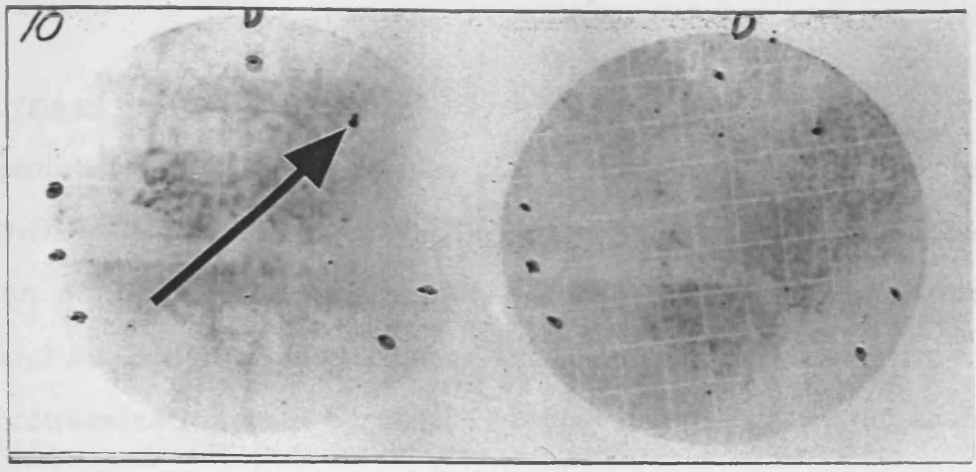


Figure 3.6
Single putative positive clone obtained from the primary screen of the C6 library with fluorescien labelled pTK2Lys plasmid. The library was plated at a density of 50,000 plaques per 150mm plate, transfered to nylon membranes and screened as described (materials and methods).
A single spot is visible in the same position on both duplicate lifts. This plaque was picked and re-screened as described.

3.2.4. Analysis of positive clone

A single, isolated positive plaque was picked from a plate from the tertiary screen, grown in liquid culture and DNA prepared using spin columns by a modification of the method of Charnock-Jones et al. (Walker and Harrison; Materials and Methods). Figure 3.7 shows that the method yields pure λ DNA with little contamination with bacterial DNA (as determined by the absence of smearing even after over exposure of the photograph).

Digestion with *Not I* released a single insert fragment of 14kb. An attempt was made to subclone the 14kb *Not I* fragment into pBluescript. However, all white colonies picked and digested with *Not I* showed a single 14kb fragment of identical size to that obtained by digestion of the original λ clone. No restriction fragment corresponding to pBluescript was present in any of the clones. This suggested that the 14kb fragment contains the origin of replication and the ampicillin resistance gene from the integrated plamid pLysSV40. Digestion of the 14kb fragment with *Eco RI* produced three restriction fragments of 6.2, 5.7 and 2.7kb (data shown in figure 3.8). The 5.7kb fragment was found to be autonomous and therefore contained the plasmid origin and ampicillin resistance gene (Figure 3.9). The 6.2kb and 2.7kb framents were subcloned into pBluescript. These data suggest that the plasmid had integrated into the active locus without deletion or significant rearangement of the origin of replication or the ampicillin resistance gene. They also suggest that the attempts at plasmid rescue in this cell line by previous workers were theoretically sound, and that attempts by this method to clone the flanking sequence at this locus failed only due to the practical difficulties discussed above.

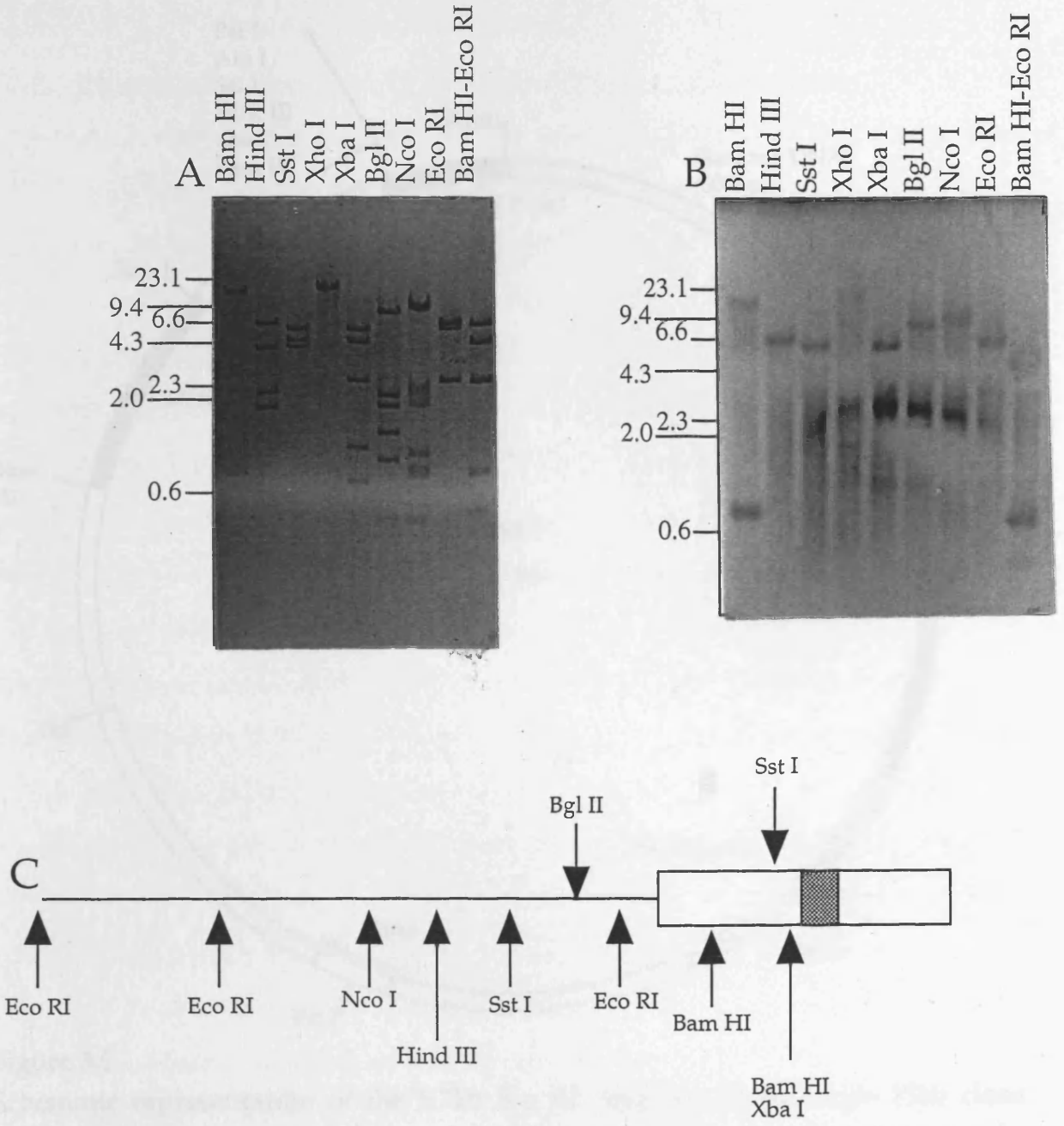
Figure 3.7
Isolation of DNA from clone.
An isolated positive plaque was grown to high density on a single 9cm plate. Phage were then diffused into 3mls SM buffer for 2 hours at 4°C. The phage suspension was then titered and expanded in liquid culture as described (materials and methods). DNA was prepared by a modification of the method of Charnock-Jones et al., 1994 (Materials and Methods). DNA on the gel represents 10µl of the 150µl obtained by this method.

Figure 3.8

Restriction analysis of the clone obtained from the C6 cell line.

The 14kb plasmid excised from the λ-clone was linearised by digestion with Not I. The linear fragment was then digested with the enzymes shown and electrophoresed through a 1% agarose gel (A). The gel was subsequently blotted on nylon membrane and Southern blotted as described (2.5.3) and hybridised with Fluorescein-αdCTP labelled probe of pTK_2Lys plasmid (B). C. Schematic representation of the restriction map generated from the above data. Positions of the Nco I, Bgl II and Eco RI are in full agreement with those mapped by Southern blotting with C6 genomic DNA (figures 3.1 and 4.1).

(Marker sizes: kb)


A
Bam HI
Hind III
Sst I
Xho I
Xba I
Bgl II
Nco I
Eco RI
Bam HI-Eco RI
23.1
9.4
6.6
4.3
2.0
2.3
0.6
B
Bam HI
Hind III
Sst I
Xho I
Xba I
Bgl II
Nco I
Eco RI
Bam HI-Eco RI
23.1
9.4
6.6
4.3
2.0
2.3
0.6
C
Sst I
Bgl II
Eco RI
Eco RI
Nco I
Hind III
Sst I
Eco RI
Bam HI
Bam HI
Xba I

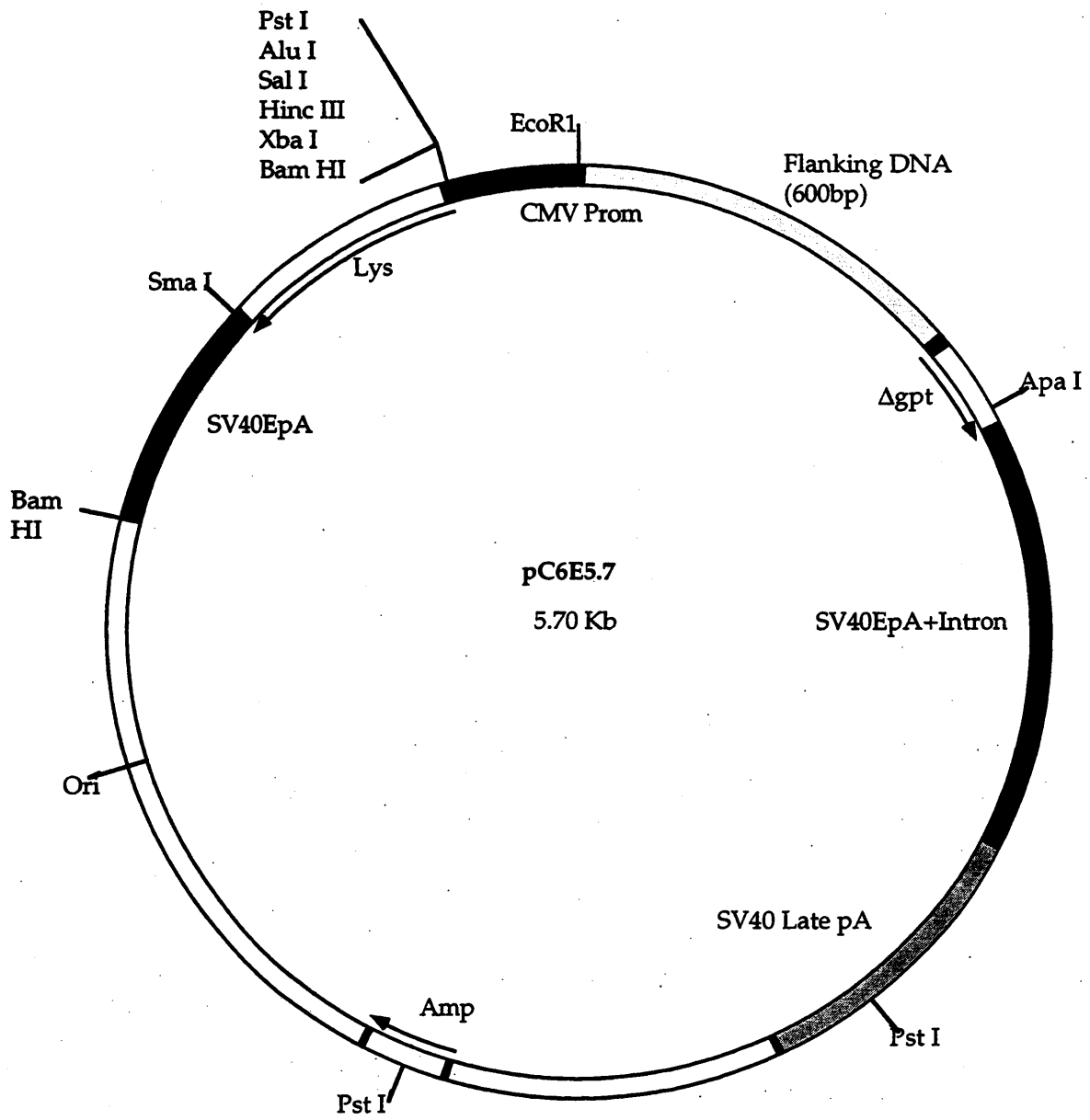


Figure 3.9

Schematic representation of the 5.7kb *Eco RI* fragment from single 15kb clone (*Not I* fragment) obtained from the C6 library. The *Eco RI* site represents the true ends of the fragment as they are found in the *Not I* fragment.

3.2.5. Sequencing of the novel DNA in present in the 5.7kb fragment

Sequencing was initially carried out on the 5.7kb plasmid using the primers CMV1 and β-lac1 (figure 3.10). Sequence obtained with CMV1 contained CMV promoter sequence which read through an EcoR1 site directly into novel sequence (flanking sequence). The *Eco RI* site after the CMV promoter represents the end of the clone. The fact that it reads directly into novel sequence is due to the fact that the 5.7kb fragment is circularised. Sequence obtained with β-lac1 primer yielded only β-lactamase sequence which read directly into pBR322 and SV40 sequence and then, into gpt sequence. The presence of gpt sequence was unexpected because the map of the pLysSV40 plasmid and the Southern blotting data (Sarah Glassford) suggested that lysozyme and gpt sequences were separated by an *Eco RI* site. It was reasoned therefore that the gpt sequence was due to duplication of the gpt sequence when the plasmid integrated. Further sequencing with primer gpt3 gave sequence data which read through the entire partial gpt repeat into novel flanking sequence. Further primers to the sequence obtained from primers gpt3 and CMV1 (gptnov and CMVnov) allowed overlapping, double-stranded sequence to be generated to all 537bp of flanking sequence present on the 5.7kb plasmid. The full sequence is presented in figure 3.10. The sequence contains two repetitive elements. Reading from the gpt end of the sequence (the gpt repeat reads directly into repetitive sequence) there are 8 repeats of the heptamer GGAGGAC. This is followed directly by 16 repeats of the hexamer GGAGAC. This repeat then reads into non-repetitive sequence with an A-T/G-C content of around 1:1.

It was never the intention of this study to generate large quantities of sequence data. Of far more interests was an investigation into the biochemical nature of

the high level of expression of the C6 and D8 plasmids. For this reason, after the boundary between the plasmid and flanking DNA had been mapped (figure 3.10) only the ends of the 6.2kb *Eco RI* fragment were sequenced. Sequencing was again carried out using the ABI 377 automated sequencer (Materials and Methods) with M13 forward and reverse (pBluescript specific) primers.

After initial sequencing of the 6.2kb *Eco RI* fragment (sub-cloned into pBluescript) *via* M13 universal primers specific to pBluescript, a PCR primer was raised in order to test the contiguity of the 6.2kb and 5.7kb fragments in C6 genomic DNA. It is feasible that the 6.2kb Lambda clone obtained could be a chimera of *Eco RI* fragments from many physical locations which ligated together during the construction of the library. This possibility was tested by PCR. Convergent PCR primers were raised to gpt sequence (gpt 3) and novel sequence from the 6.2kb *Eco RI* fragment (6.2rev). Touchdown PCR was carried using a Long PCR kit (Boehringer Mannheim) on C6 genomic DNA, J558L genomic DNA, and J558L genomic DNA spiked with 50pg clone DNA (the 14kb *Not I* fragment). Figure 3.12 shows a PCR product of predicted 600bp for C6 genomic DNA and 14kb clone DNA but not for J558L genomic DNA alone. This shows that the λ clone obtained from the library is representitive of the C6 locus and is not the result of gross rearrangements during the cloning process.

3.3. Discussion

Pearson-Lipman searches (FASTAs) using Wisconsin GCG software revealed 100% homology of the repetitive region with mouse B1 elements. These elements are similar but non-homologous to Alu repeat elements in humans. Both B1 and Alu elements are members of the SINE (Short Interspersed Nuclear Element) family of repetitive sequences. Like other interspersed repetitive

elements SINEs result from retrotransposition, insertion of RNA sequence into the genome via its conversion to DNA by a reverse transcriptase. B1 and Alu elements are homologous to and are thought to originate from 7SL RNA, a structural RNA which forms part of the SRP complex.

Alu elements in primates and B1 elements in rodents appeared independently after the divergence of rodents and primates from their common ancestor (Schmid and Jelinek, 1982). 7SL Alu and B1 genes are transcribed by DNA polymerase III (Pol III) and contain the Pol III specific, internal A-box and B-box promoters within the transcribed unit. For this reason, newly retrotransposed Alu and B1 members are themselves capable of being transcribed. This is not the case with retrotransposed Pol II transcripts. In Pol II transcribed genes the promoter elements are upstream of the transcription start site and are therefore not included when the transcript is retrotransposed. This is almost certainly how many of the known 'pseudogenes' originated.

The fact that B1 and Alu elements evolved separately in rodents and primates, and have been maintained for so long in evolution would suggest that they confer considerable selective advantage. Although they are transcribed by Pol III, Alu elements have been shown to bind Pol II specific transcription factors. A non-consensus Alu within an intron has been shown to have enhancer activity (Hambor, et al., 1993) SINEs have also been implicated in genomic imprinting. They are hypermethylated in the female germline and hypomethylated in the male germline This methylation is Alu specific and is opposite to the general pattern of methylation of the germline in which sperm DNA is generally hypermethylated and oocyte DNA is hypomethylated. However this general (non-Alu) methylation is rapidly altered in somatic cells. Thus their wide

dispersion throughout the genome and their specific methylation patterns makes them ideally suited for gametic imprinting.

A potential function of Alu and B1 transcripts has also been demonstrated. SINE transcripts are barely detectable in normal cultured cells. However there is a massive increase in their transcription in response to cell–stress, viral infection or translation inhibition (treatment with cyclohexamide). From this evidence it is thought that SINE RNAs may have a role in preventing the activation of eIF2 kinase. This prevents the general switch from cellular to viral gene translation in early viral infection.

SINES are thought to constitute around 10% of the mammalian genome, and there are around 10^6 B1 homologues in the mouse genome. Therefore the presence of a B1 element in the flanking sequence of the C6 plasmid is not surprising. Although, as discussed above, Alu and B1 elements have been shown to bind DNA polymerase II transcription factors and to alter chromatin structure (Schmid and Jelinek, 1982) it is not clear that the presence of a B1 element in the vicinity of the plasmid is likely to be of little functional consequence in term of the high expression level of the plasmid. Any conclusion or discussion concerning the presence of the B1 repeat is entirely speculative although it is reassuring that B1 elements are often located around transcribed sequences and are enriched in R bands on chromosomes.

The sequence of the flanking DNA beyond the B1 element is non-repetitive. Apart from minor stretches of high homology (typically around 20bp of 90-100% identity), Pearson-Lipman searches yielded little in the way of homologous matches from existing databases to the non-repetitive flanking DNA. The

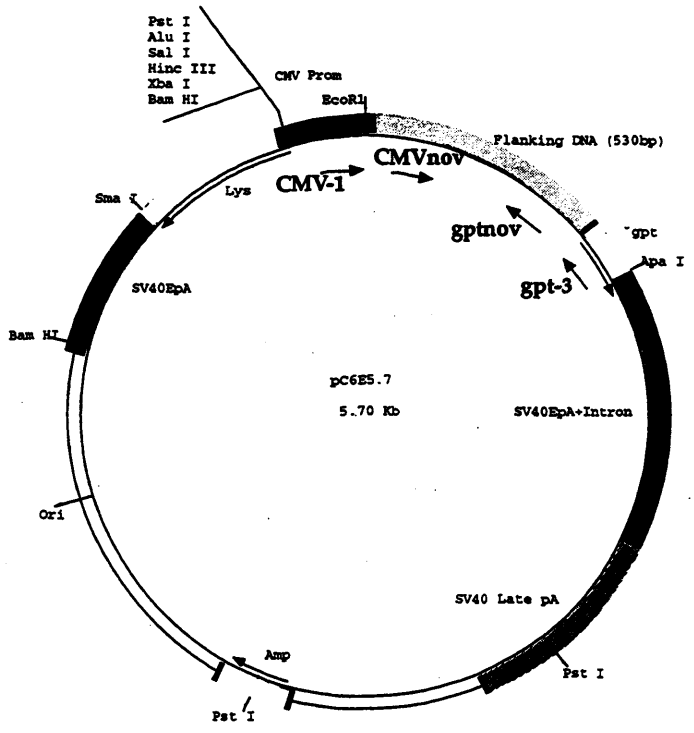
longest stretches of homology obtained from database scans was 293bp (53.2% identity) and 231bp (53.7% identity) from the ends proximal and distil to the plasmid respectively. Both of these are to cosmid clones from C.elegans. (Wilson, et al., 1994). No other significant stretches of homology were found. A full list of sequences similar to the flanking DNA is available.

Although complete double strand sequencing will be required to confirm any conclusions regarding the sequence data, it is interesting in itself that the only sequence homologies found for both ends of the 6.2kb *Eco RI* fragment were of low homology to cosmids from C.elegans. The function, if any and location of these homologous sequences is at present ill defined, so detailed comparison between the two sequences would be largely fruitless. This fact also vindicates the approach whereby obtaining full sequence data for the genomic clone was given relatively low priority. It would be pointless to generate huge quantities of sequence data for clones whose function if any in generating the high level of expression of the C6 plasmid is undefined, and if there are no characterised homologous sequences with which to compare them to. It was considered far more beneficial to first obtain clones from the D8 cell line and secondly, to concentrate on determining the mechanism of the high cell of expression in the C6 and D8 cell lines.

Figure 3.10

Top: Strategy employed in sequencing the 500bp flanking DNA in the 5.7kb Eco RI fragment. Sequencing was initially carried out using primers CMV-1 and gpt3. These generated DNA sequence which read through CMV promoter and Δgpt sequence respectively into novel (flanking) sequence. Further primers (CMVnov and gptnov) were then generated. Bottom: Sequence data from all four primers was fitted together using 'Bestfit' and 'Overlap' algorithms on theWinconsin GCG package. Hexamer and Heptamer repeat regions with homology to mouse B1 elements are underlined.

A



B

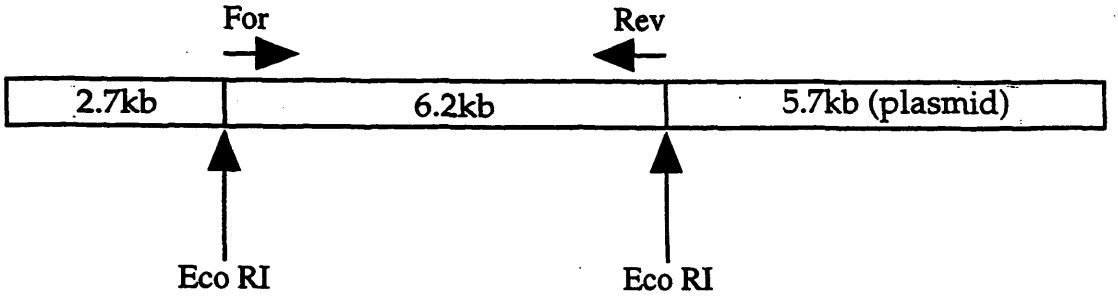
CMV

```
1   CTTTCAGTGC ACAGCCACTG GTGNAAGCTA TCATGCTCCT GTGCGGACCT
51  TANCAAAACT CAGCCTGTTG AAAAGAACAA ACAAAAGAAA AGAAGCCCGG
101 AGGAAAGGAG ATGCAGGGCT GGAGGGGGGG GGGGGCTCCG NCTCACCCCT
151 GCAGTCCCTT GGGAACAGGG CTGGAGGACT CAGGAATCTC ACTGTGTGCC
201 AAGGCTCTCT CTTGTCTTCA ACAAATCCAA AAAAGAAAGC AGCCCGGCCT
251 GTGGCTCCGA GGGTAGAGTG CTAGCCTATG GGACCTTNTC TCAAAGCAGC
301 AATGATAACA GTAATAAAAT TAACAGTATT TANNGATGGC TAGTCNGGGT
351 CAGGCATGNC AGCACATNNC TTTAGTCCCA GNAGNCAGAG GCAGAGNCAG
401 AGGCAGAGGC AGAGNCAGAG GCAGAGGCAG AGNCAGAGGC AGAGGCAGAG
451 GCAGAGGCAG AGGGAGAGGC AGAGGCAGAG GCAGGAGGCA GGAGGCAGNA
501 GGCAGGAGGC AGNAGGCAGG AGGCAGGAGG CAGGAGG
```

Δgpt

Figure 3.11
Sequencing strategy for the 6.2kb EcoR1 fragment
Top: Schematic diagram showing the orientation of the pBluescript M13 primers with respect to the whole λ clone. This was determined by restriction analysis using double digesion with enzymes unique to one side of the pBluescript multicloning site plus enzymes with sites mapped in figure 3.8 (data not shown).

Bottom: Sequence obtained from the ends of the 6.2kb EcoRI fragment. Multicloning site sequence has been removed.



rev

```
  1  TTCCTTTCTC TGTGGTNGTG GNAACATAGT CTTTGCTTCC TTGCTTGGTT
 51  AAGATCTACT TCCTGTTTCT AGTTACTACC TCATACTTGG GAGGCATCTT
101  CAAAGCTCGA GCCTTTCTCA TGCTGATCTC CTGATTTTCA GAGCTGATGT
151  ATATAGGTTC TACAATAAGG TAATTGGAAC CACAGCTATC CAGAATGGCC
201  TCTTCTTTGT ATATCATTGA NGTTCTGGNT TTCACTCTAT CCCAAACCTC
251  AAACTCTCTT AATTTGGGCA TTTGGATATG TATGGCATGT CTTTGGGTGA
301  ACCTATTAAG TCTGCCACGT TTTAATCTTT TTCAGAAAAC CCNCAACCCA
351  TTTANGATTA GAGGGTTCC
```

for

```
  1  TTCCCTGCTT AGCTATTGGG GCATCATTAG ACCTCTTCTG GGGCTCAGCT
 51  GTCATTATGC CAGATATCTT TATTTTAGAA ATACTGTATG GACCTTTCTC
101  AGACATTATA GCACTCAGAG TGGGATAATC TTCCTTTCTC CCTTTGTTAG
151  TATTCGATGG CAATGTCACA TTTCTGAGCA TGTCACTACT ACCTAACCAT
201  CAGTCCTCCC CTCGACTCTG GAAACTTGTG TCCACCAGTG GTCTCTGTCA
251  TTTTAAATTT CTAGTTCAAG TCAAATTAAA AAAAAAAATT AACAGTCTTT
301  ACCAAGGTAT CCTGGAGGNT CTAGGGAAGA CCCTGATCTC TAGANCANCT
351  GTCTACTATC CAATGTAGGN CTCAATCTAC ANAGGATG
```

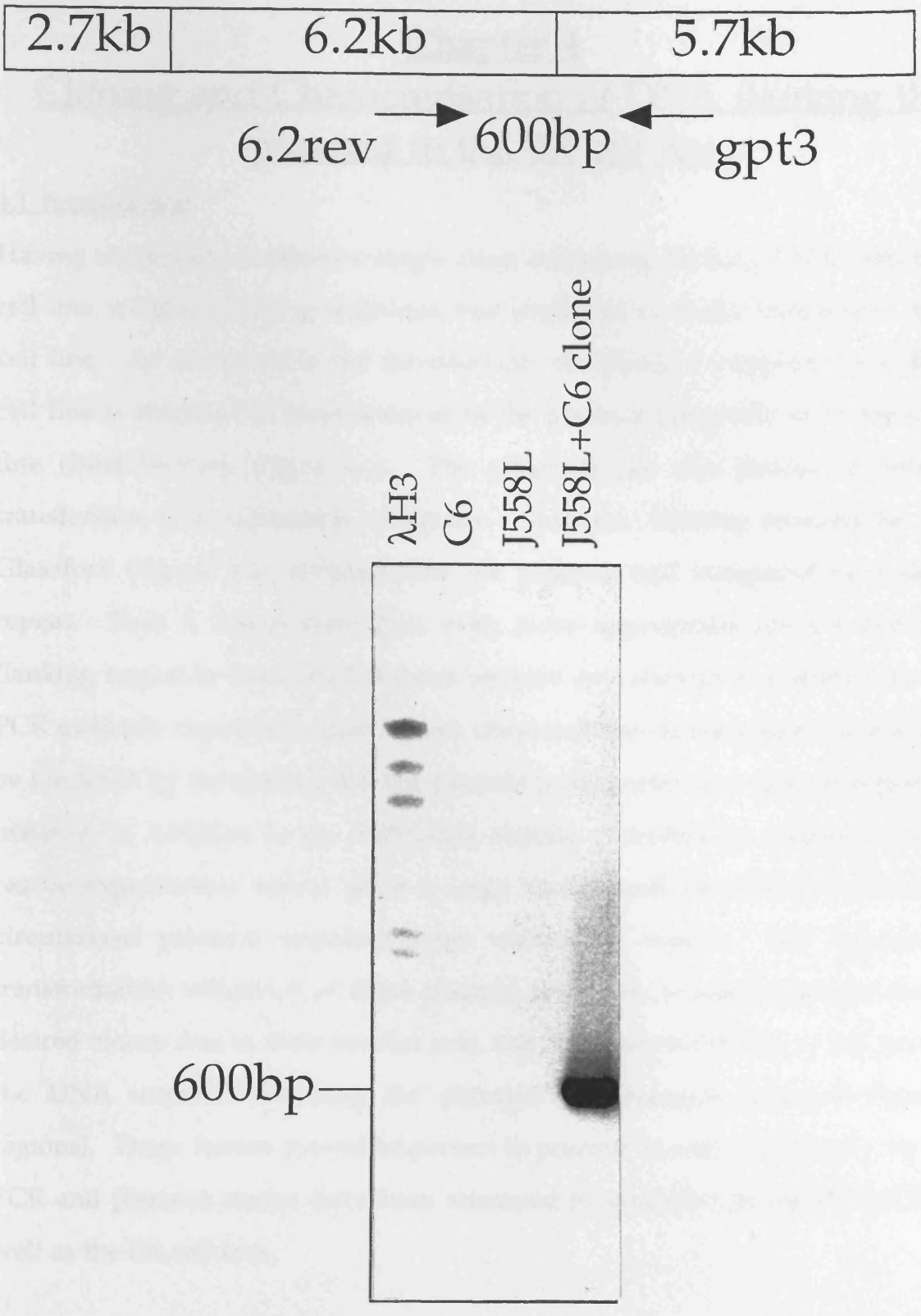


Figure 3.11.
PCR to confirm the integrity of the C6 clone with respect to the 5.7 and 6.2kb Eco RI fragments. PCR was carried out on 500ng C6 genomic DNA, 500ng J558L genomic DNA (negative control) and 500ng J558L DNA spiked with 50pg C6 λ clone DNA (positive control) using a Long PCR kit (Boehringer Mannheim) in buffer containing 1% (final v/v) DMSO. Touchdown PCR was performed as described (Materials and Methods). Gel represents 1/10th total reaction product

Chapter 4
Cloning and Characterisation of DNA flanking the plasmid in the D8 cell line.

4.1 Introduction

Having succesfully obtained a single clone containing flanking DNA from the C6 cell line, a similar cloning technique was employed to obtain clones from the D8 cell line. As discussed in the introduction, the plasmid integrated into the D8 cell line is identical in most features to the plasmid integrated in to the C6 cell line (Introduction; figure 4.1). The plasmids are also similar in terms of transfection and expression efficiency. Southern blotting analysis by Sarah Glassford (figure 4.2) revealed that the plasmid had integrated as a tandem repeat. Thus λ based cloning is even more appropriate for the cloning of flanking sequence from the D8 locus because any attempt at plasmid rescue or PCR methods dependent upon known characteristics of the plasmid are likely to be hindered by the the fact that the plasmid is integrated as a tandem repeat. For instance, in addition to the difficulties already described in chapter 3, plasmid rescue experiments would yield a large background of clones containing re-circularised plasmid sequence from within the repeat. The ligation and transformation efficiency of these plasmid fragments would be greater than the desired clones due to their smaller size, and also potentially due to the nature of the DNA sequence flanking the plasmid (for example potential repetitive regions). These factors proved important in practice as well as in theory, as again PCR and plasmid rescue have been attemped to exaustion in the D8 cell line as well as the C6 cell line.

The restriction map of the D8 locus (Figure 4.2) shows that a partial *Eco RI* digest, as was used to generate the C6 genomic library would limit the potential overlap

of clones within the library and, because of the tandem repeat, any clones obtained by screening with a plasmid specific probe would be likely to contain an excess of plasmid sequence compared to flanking sequence. Therefore, a D8 genomic library was constructed using a partial *Sau 3A* digest. This enzyme has the short recognition sequence, GATC. Overlapping restriction fragments from partial digestion with *Sau 3A* should therefore be more representitive of the whole locus and should contain considerably more flanking sequence compared to plasmid sequence than fragments obtained from a partial *Eco RI* digest.

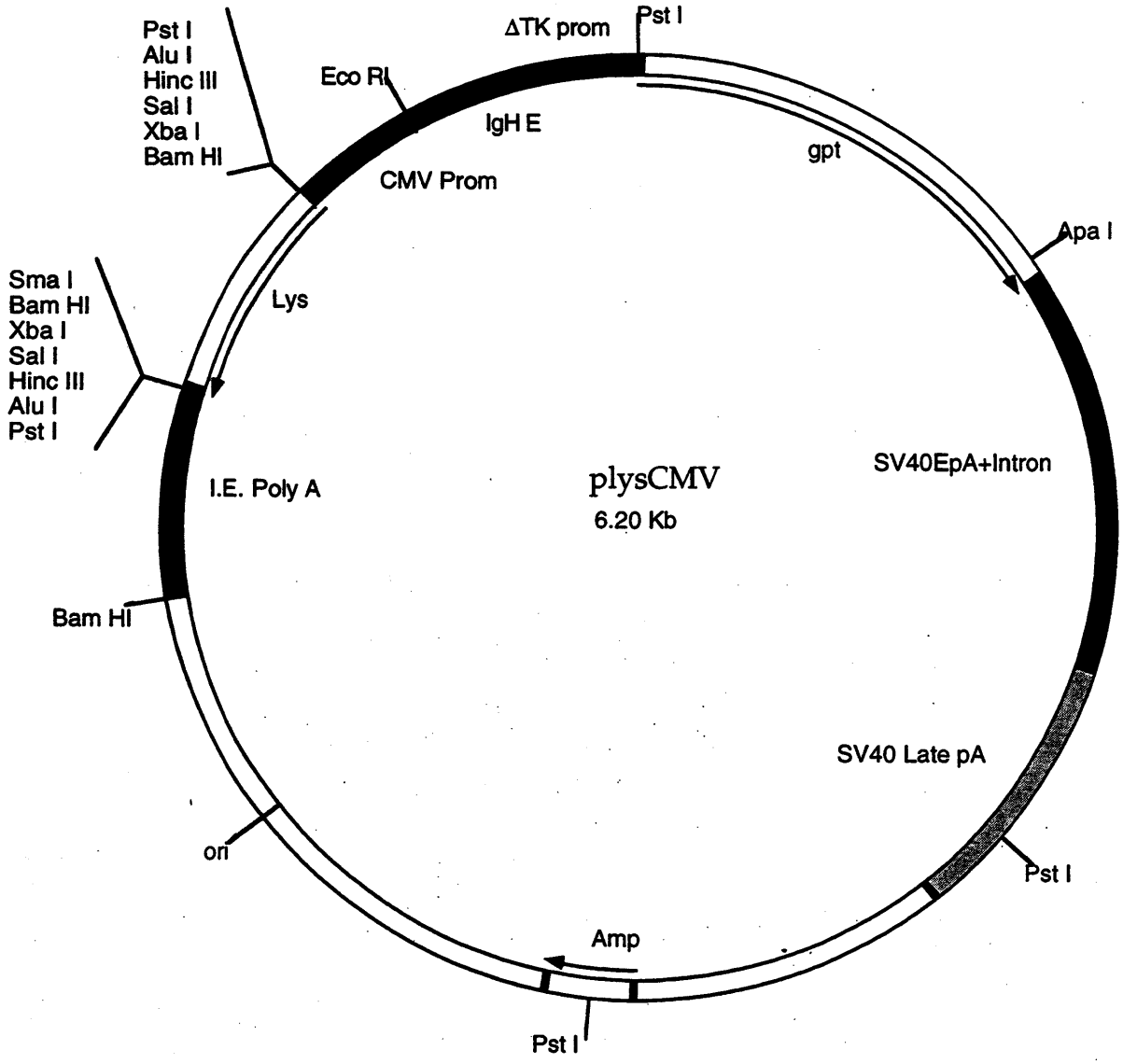


Figure 4.1
Schematic representation of the plasmid transfected into the D8 cell line. The plasmids in the C6 and D8 cell lines differ only in the polyadenylation signals after the lysozyme coding sequence. pLysCMV contains the CMV immediate early polyA signal and an intron following the lys coding sequence.

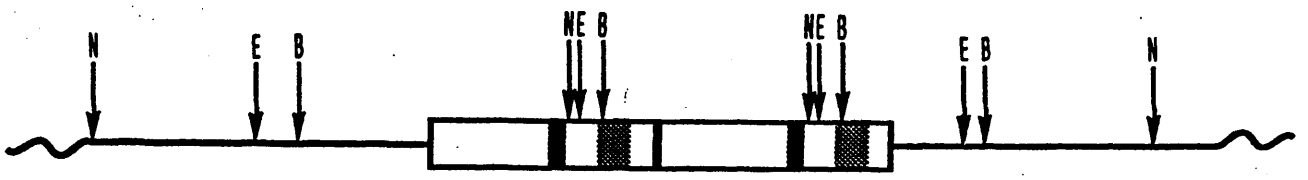


Figure 4.2
Restriction map of the DNA flanking the plasmid in the D8 cell line generated from Southern blotting data obtained by Sarah Glassford. The data showed that the plasmid integrated as a tandem repeat. Quantitative analysis suggested that two copies are present. A single band was obtained in Southern blotting experiments with rare cutting restriction enzymes (data not shown) suggesting a single integration site (Harrison, et al., 1993)

4.2 Results

4.2.1 Production of a partial *Sau 3A* library from D8 genomic DNA.

A partial *Sau 3A* genomic library was generated by the method described for the C6 library in chapter 3 with minor modifications. Optimal conditions for the partial digestion of D8 genomic DNA to give a modal fragment size of 9-23kb were determined by the serial dilution of enzyme as previously described (chapter 3; Figure 4.3). Once optimal conditions were determined they were scaled up stochiometrically to accomodate 50µg DNA. The resulting fragments were enriched for fragments between 9-23kb by centrifugation on a sucrose gradient. Fractions containing fragments between 9-23kb long, as determined by pulsed-field agarose gel electrophoresis (Figure 4.4), were pooled and precipitated with isopropanol.

In addition to the library construction method described in chapter 3, the size fractionated D8 DNA was also treated with HK™ phosphatase (Epicentre Technologies). It would be expected that phosphatase treatment would increase the window of optimal ratios between insert and vector DNA by eliminating ligation between insert fragments. Thus ligations between two or more insert fragments to form unpackagable clones containing inserts greater than 23kb are eliminated. Phosphatase treatment should also reduce the risk that any clones obtained from the D8 locus are mosaics of *Sau 3A* fragments from different genomic loci. HK™ phosphatase presents advantages over standard calf intestinal alkaline phosphatase preparations in that it can be totally inactivated by heating at 65°C for 30 minutes. In pilot runs with digested D8 DNA, phosphatase treatment for two hours (as described in the manufacturer's instructions) yielded an average number of plaques per microgram of λ-DASH

DNA of $3.9x10^5$, compared with $2.7x10^5$ from the same quantitiy of non-phosphatase treated D8 DNA.

The ligation was packaged using the commercial packaging extract Gigapack Gold™ (Stratagene), and titered as described (materials and methods; chapter 3). A library of 550,000 independent clones was obtained. It was likely that a library of this size would contain several clones which spanned the locus of interest. Given the relative ease with which a clone was obtained from the C6 genomic library, the library was amplified as described by the manufacturer (Materials and Methods). A $1x10^5$ fold amplification of the library was obtained.

4.2.2 Screening the partial Sau3A genomic library with a lysozyme-specific probe

The library was screened as previously described (Materials and Methods) using a α [^{32}P] dCTP labelled PCR product raised from primers lys3 and lys4 (figure 4.1). Again putative positive plaques were picked and re-screened as described (Materials and Methods; chapter 3). Three independent positive plaques remained at the end of the tertiary screen. These were arbitrarily named after the primary screening plates from which they were derived; 13, 14 and 18 (figure 4.5). All three clones were expanded in large scale liquid culture and the DNA isolated as described (Materials and Methods section 2.5.3).

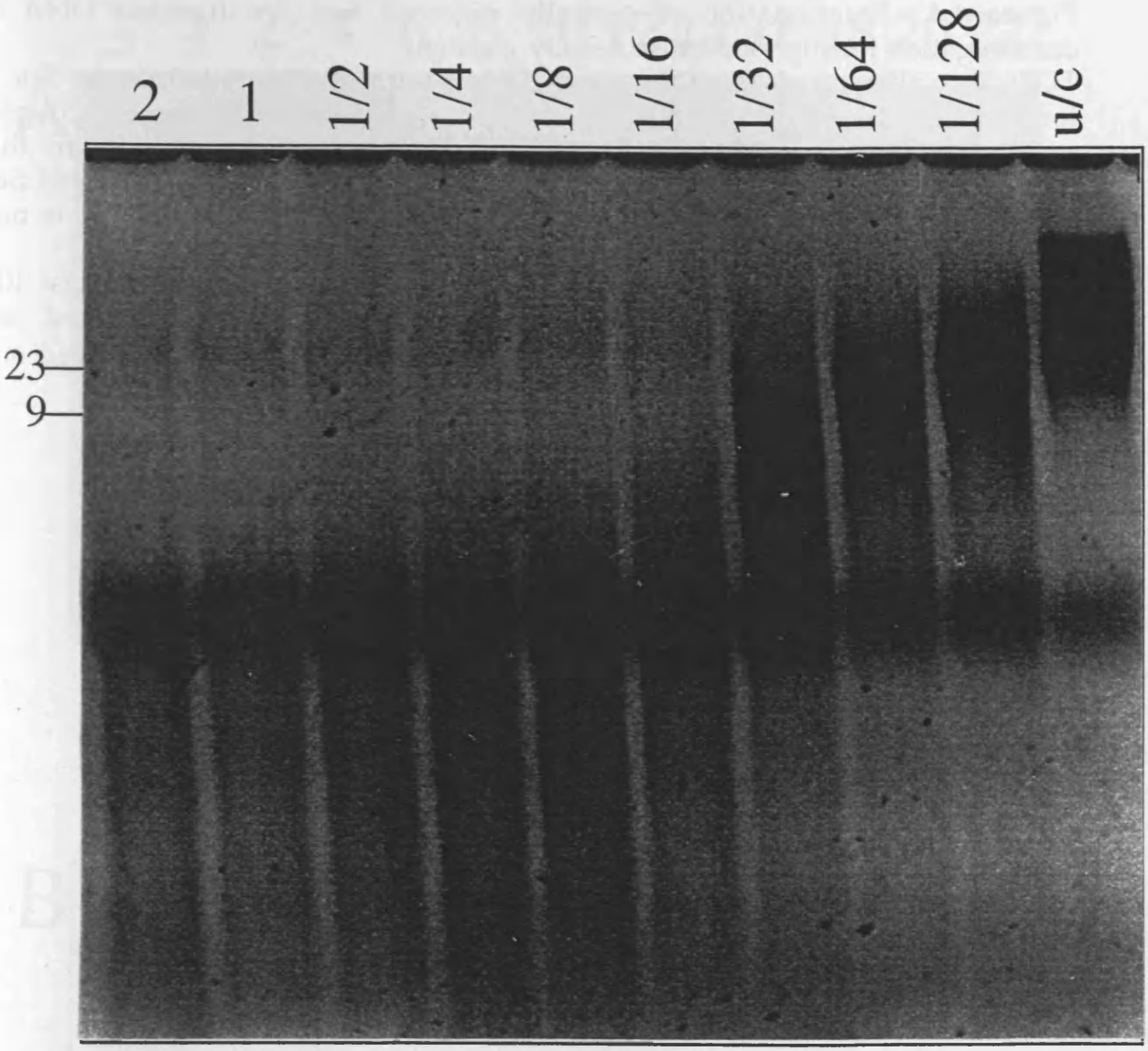
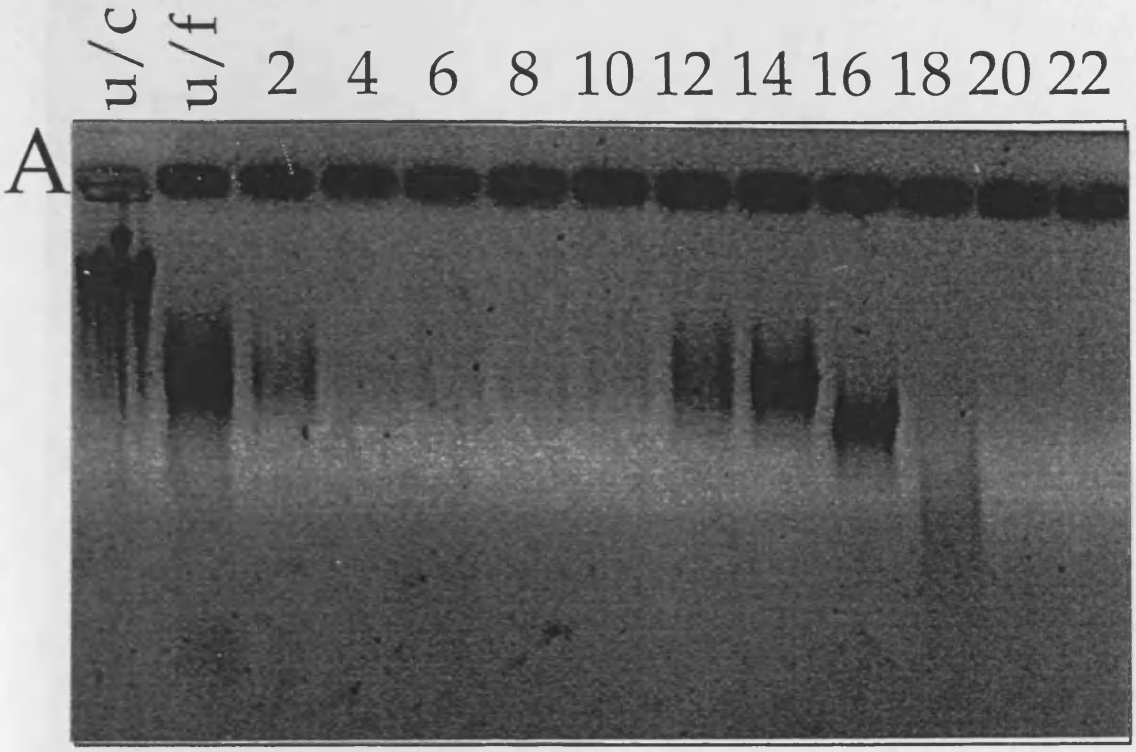
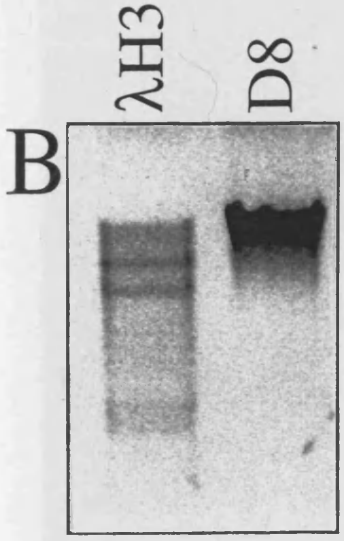


Figure 4.3 Partial digestion of D8 genomic DNA with *Sau 3A*.
1µg of D8 genomic DNA in a total volume of 30µl was digested with serial two fold dilutions of *Sau 3A* and examined by pulsed field gel electrophoresis. (200V, 1-4s ramp over 5 hours at 10°C. The gel suggests that the optimal digestion conditions are around 1/100 units of enzyme for 1µg DNA in a total volume of 30µl. These conditions were scaled up 50µg DNA in digested with 0.5 units Sau 3A in a total volume of 1500µl.

Figure 4.4 Fractionation of partially digested *Sau 3A* digested DNA by centrifugation through a sucrose density gradient
TOP: Two aliquots of 50µg D8 genomic DNA were digested with 0.5 units *Sau 3A* for 1 hour. The digested DNA was then centrifuged at 26,000rpm through a sucrose gradient (5-20%) for 6 hours at 15°C. The gradients were then fractionated into 1.5ml aliquots and every second aliquot analysed by pulsed field gel electrophoresis (200V 1-4s ramp over 5hours at 10°C). u/c represents non-digested D8 DNA, u/f digested non-fractionated D8 DNA.
BOTTOM: Fractions 15,16 and 17 were pooled, dialysed overnight against 1000 volumes 1X TE buffer, ethanol precipitated, phosphatased treated and redissolved in 1XTE at 1mg/ml. $^{1}/_{50}$th of the recovered DNA was loaded onto the gel.


A
u/c
u/f
2
4
6
8
10
12
14
16
18
20
22


B
λH3
D8

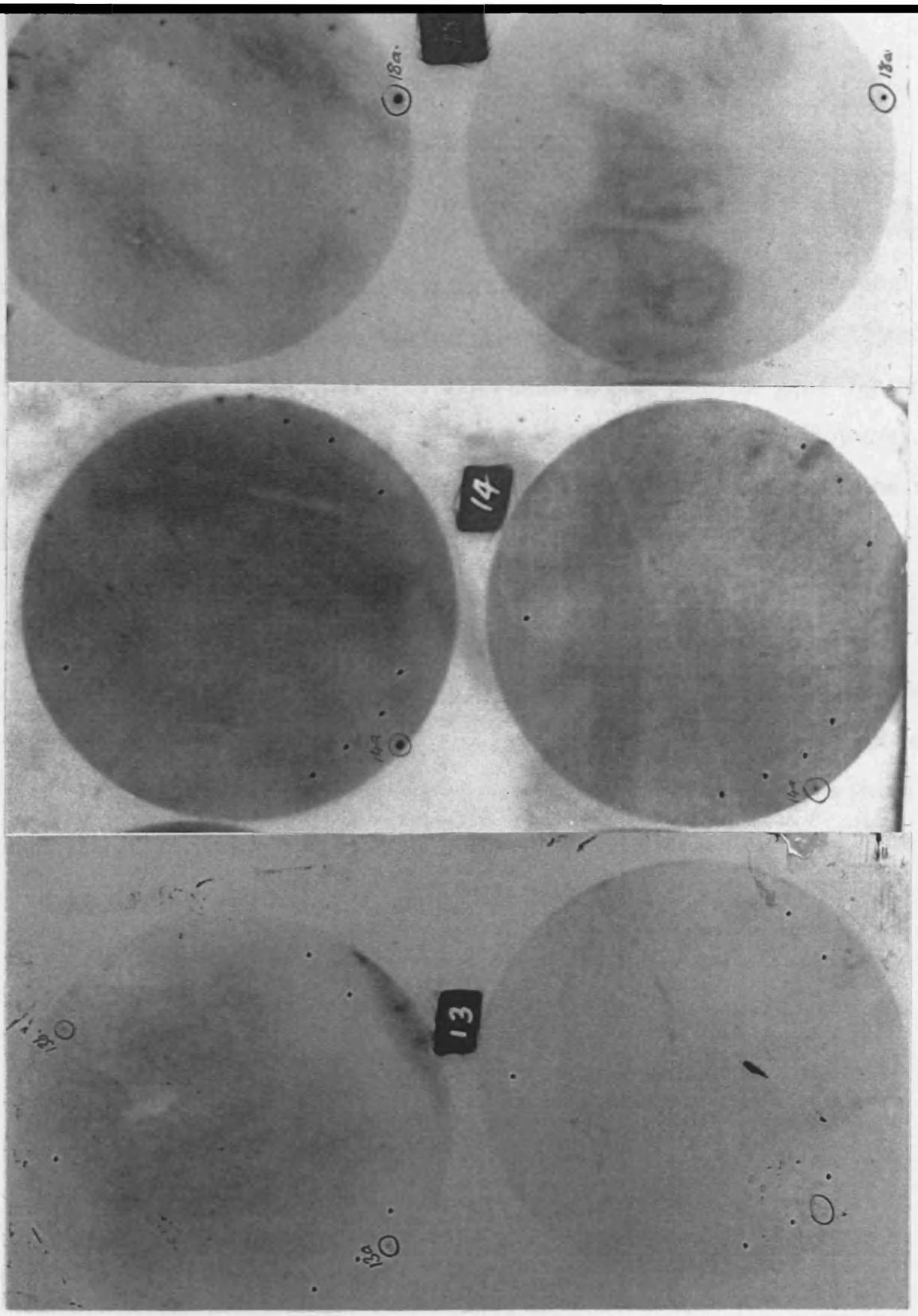


Figure 4.5

Primary screening of the partial Sau3A genomic library

The library was plated at a density of 50,000 plaques per 150mm plate and screened with a lysozyme specific probe generated from a 600bp PCR product from primers lys3 and lys4 as described (Materials and Methods). 9 putative positive plaques were obtained from this primary screening 3 of which (13, 14, and 18; shown above) remained positive through the tertiary screening.

4.2.3. Subcloning Strategy

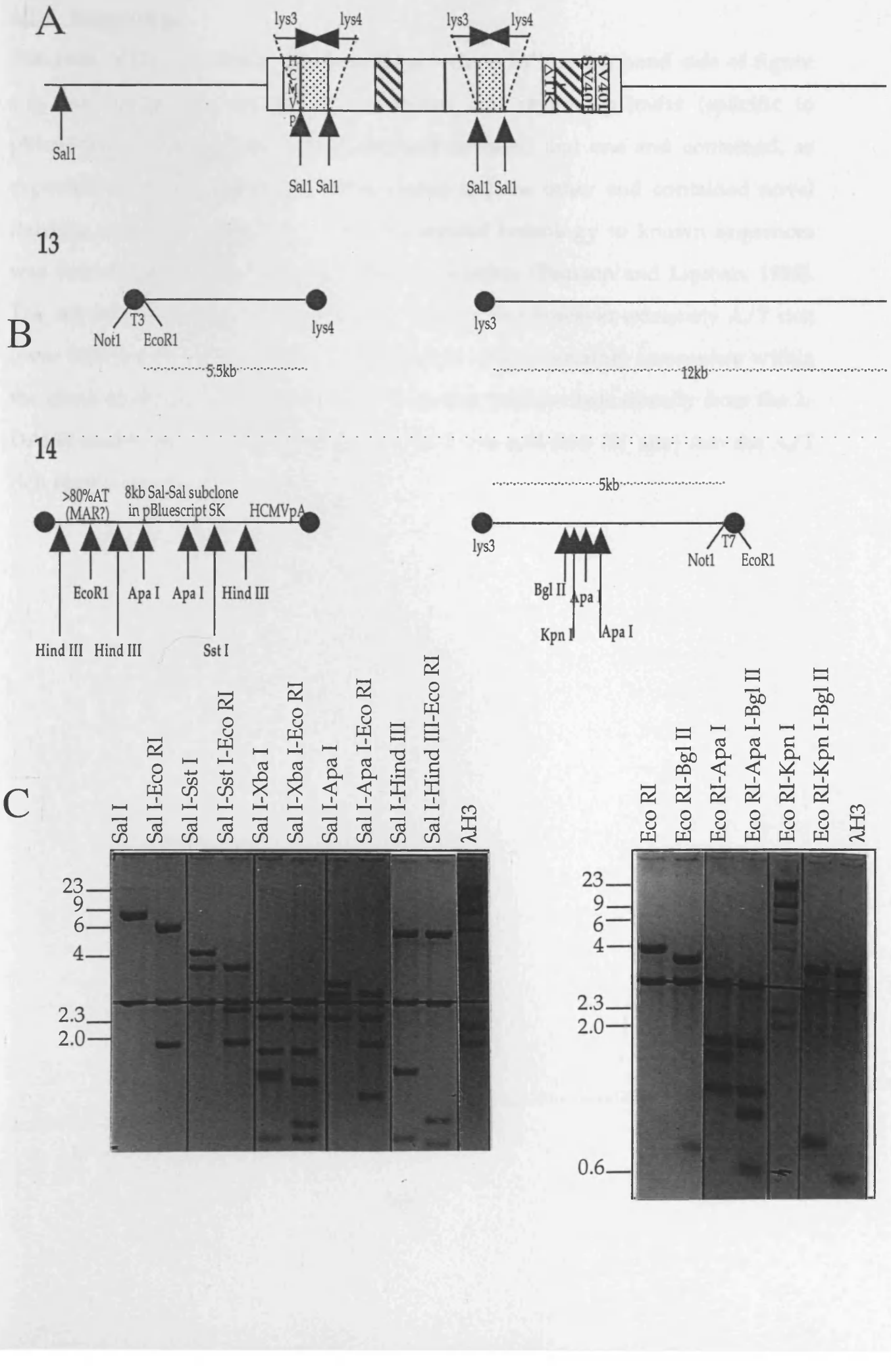
Restriction digestion with *Not I* revealed that inserts in clones 13 and 14 were 23 and 20kb respectively. Clone 18 was considerably shorter at 15kb. An initial sub-cloning strategy was based on ligation of *Sal I* fragments from clones 13 and 14. *Sal I* restriction sites are present on both sides of the lysozyme gene (figure 4.1) and at the end of the λ-DASH multi-cloning site (figure 3.5). However, the clones proved recalcitrant to this stratetegy and only one fragment, the right hand *Sal I* fragment of clone 14 was sub-cloned in this way (figure 4.6).

A Long PCR method was employed to subclone the remaining fragments from clones 13 and 14. Primers Lys3 and Lys4 were used against T7 and T3 respectively (figure; figure 3.5). These primer combinations were determined empirically. Long PCR on clone 18 showed that it was shorter in both directions (from lys3 and lys4) than clones 13 and 14 and was therefore not further analysed. The long PCR products were subcloned into pBluescript™ (Stratagene) by digestion with *Eco RI* as depicted in figure 4.6.

Figure 4.6
Characterisation of clones 13 and 14 from the D8 genomic library
A. General structure of the D8 locus, showing the tandem repeats of plasmid plysCMV. Known, key restriction enzyme sites are marked with vertical arrows, primer binding sites are depicted with horizontal arrows.

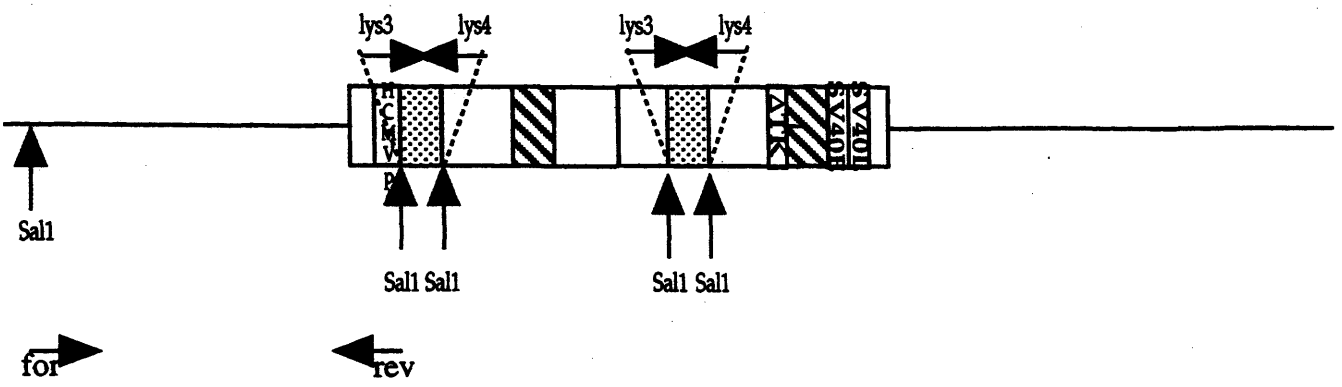
B. Diagramatic representation of subcloned ends.
Left side: An 8kb fragment corresponding to the left side of the locus was subloned directly out of λ clone 14 by digestion with *Sal I.* A 5.5kb fragment could be generated by long PCR out of λ clone 13 using primers lys 4 (plasmid sequence) and T3 (λ DASH sequence)
Right side: A 12kb and 5kb fragment could be obtained by long PCR out of λ clones 13 and 14 respectively. These were subcloned into pBluescript by digestion with *EcoRI.*

C. Restriction analysis of the subcloned fragments using combinations of restriction enzymes. For ease of interpretation, data for restriction enzymes which did not cut within the clone were removed from the photograph. Restriction sites for Eco RI and Bgl II within the clones correspond to exactly to those mapped by previous workers (Glassford, 1993)


A
lys3
lys4
lys3
lys4
Sal1
Sal1 Sal1
Sal1 Sal1
13
B
T3
Not1
EcoR1
lys4
lys3
5.5kb
12kb
14
>80%AT
(MAR?)
8kb Sal-Sal subclone
in pBluescript SK
HCMVpA
EcoR1
Apa I
Apa I
Hind III
Hind III
Hind III
Sst I
5kb
lys3
Bgl II
Apa I
Kpn I
Apa I
T7
Not1
EcoR1
C
Sal I
Sal I-Eco RI
Sal I-Sst I
Sal I-Sst I-Eco RI
Sal I-Xba I
Sal I-Xba I-Eco RI
Sal I-Apa I
Sal I-Apa I-Eco RI
Sal I-Hind III
Sal I-Hind III-Eco RI
λH3
23
9
6
4
2.3
2.0
Eco RI
Eco RI-Bgl II
Eco RI-Apa I
Eco RI-Apa I-Bgl II
Eco RI-Kpn I
Eco RI-Kpn I-Bgl II
λH3
23
9
6
4
2.3
2.0
0.6

4.2.4. Sequencing

The ends of the *Sal I-Sal I* subclone from λ-clone 14 (the left hand side of figure 4.5) was sequenced using M13 forward and reverse primers (specific to pBluescript). The sequence data obtained revealed that one end contained, as expected, the SV40 polyadenylation signal and the other end contained novel flanking sequence (figure 4.7). No substantial homology to known sequences was found with Pearson-Lipman (FASTA) searches (Pearson and Lipman, 1988). The sequence obtained with the reverse primer was however extremely A/T rich (over 80% across 500bp). This A/T stretch is almost certainly incomplete within the clone as the sequence from the M13 reverse primer reads directly from the λ-DASH multi-cloning site (between the *Sal I* site and *Bam HI* site) into the A/T rich region (figure 4.8).

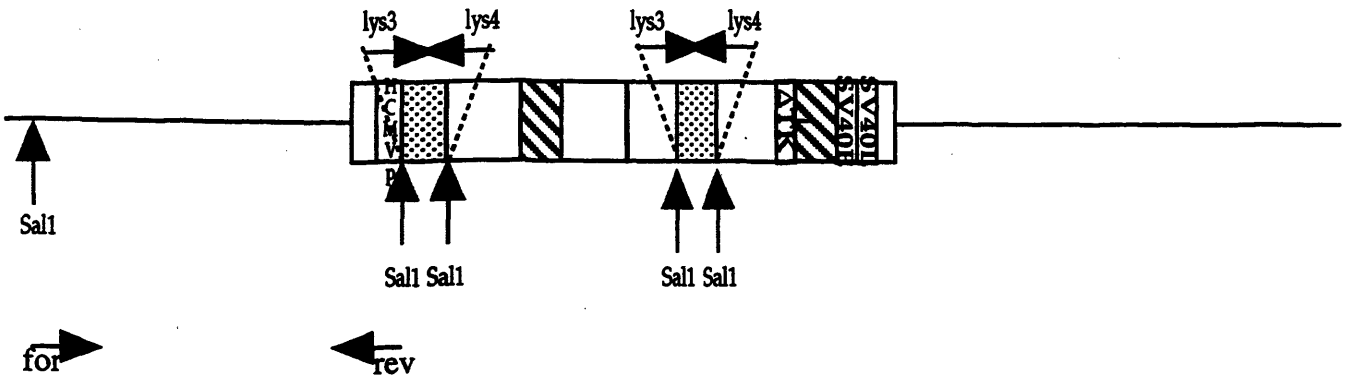



```
   1  GACTCTAGAG GATCCCACGT CACTATTGTA TACTCTATAT TATACTCTAT
  51  GTTATACTCT GTAATCCTAC TCAATAAACG TGTCACGCCT GTGAAACCGT
 101  ACTAAGTCTC CCGTGTCTTC TTATCACCAT CAGGTGACAT CCTCGCCCAG
 151  GCTGTCAATC ATGCCGGTAT CGATTCCAGT AGCACCGGCC CCACGCTGAC
 201  AACCCACTCT TGCAGCGTTA GCAGCGCCCC TCTTAACAAG CCGACCCCCA
 251  CCAGCGTCGC GGTTACTAAC ACTCCTCTCC CCGGGGCATC CGCTACTCCC
 301  GAGCTCGAAT TTATGCGGTG TGAAATACCG CACAGATGCG TNAGGAGAAA
 351  ATACCGCATC AGGCGCTCTT CCGNTTCNCN NNNANTNAAT CNTGNGCNCG
 401  GTCGTTCGGN TGCNGNNAAA CNGTATNNCT CNTNNNANGG GGTNATTNGG
 451  TNNTCCCNNA ATNNNGGGAN AANNNGAAAA AAAATNTNAN NNNNAGGNCG
 501  NNNNNGGNCN GGAANNNNNN NGGNNNGTTG NNGNGTTTNC CNNNGNNCCN
 551  NCCCCCCNNA NNAA
```

Figure 4.7

Sequence of the proximal end of the subcloned 8kb Sal-Sal fragment (using M13 forward primer). As expected the sequence is 100% homologous to CMV polyadenylation signal.

Figure 4.8
Sequence of the distal end of the subcloned 8kb Sal-Sal fragment. (using M13 forward primers) The sequence reads through remaining λ-DASH polylinker directly into A/T rich sequence.




```
   1  GACGCGGCCG CGGAATTAAC CCTCACTAAA GGGAACGAAT TCGGATCCAA
  51  ATCATTTTTA AGCCCTGGGA CTCTTATCTA CTTCTCACTG TTCAGTGGTC
 101  CCATCAGTAT GTNCAAGGTG CATACCTAAA TATAATGTAA GCAATATCAC
 151  AGCAAGCCAA TAGCCAGCAT CAAACTAAAT GGGGAGAAAC TTGAAGCAAT
 201  CCCACTAAAA TCAGAGACAA GACAAGGGTG CCCACTCTCT CCCTATCTAT
 251  TCAATATAGT ACTCGAAGTT CTAGCAAGAG CAATTAGACA ACAAAAGGAG
 301  GTCNACGGGA TACAAACTGG AAAGGAGGAA GATATCNCTA TTTGCAGATG
 351  ATATGATAAT ATACATAAGT GACCCNAAAA AATCCACGAG AGAACTCCTG
 401  TTGGAGCAAA TNAACNACCT CANCAAAGTG GCTGGACACA AAATTAACTT
 451  GAACAAATCA AAATGTTTAA ATATATTAAT TTTTGCAATT CTNAGGAATC
 501  TGCCTCAAGG AAAAATATCT TATGTTTCAT TGAATNTGTG TTCATAATGA
 551  TATTAATTAC NGTATTTCAA ATTATATAAT ACTTGTTNNG AAGGGGCTGC
 601  CGTTNGGATG TTAANTTAAT AAATTAATNA AAGAAAGTCN TTNNTTGAAT
 651  NATGGGATGA ATATNAANCT GTNTTAAANT TTNANTTCAA AACNTNCNAN
 701  ANTTTGCNTT NTTGCNNNNC CANNTNNAAA AATTATNGGA CCNCCTTGNA
 751  GGCTGGGGAA AAAAACGGGC TCANTTTTTT CAAAANACCN TTGNNTGNTT
 801  TTTGTCAAAA AAGGANCCAA GTTTNAAATT TNCCNNCNCC CCCCCGGNTT
 851  NTCCNCCTTG NNNNTTTTNG GGGGNTTCTN GGCCTCCTCT TTTTCCCCCC
 901  NNNNNCNCTG GNGGNCNNNN GGGNAAAAAA TTTNNNAANT NCNTCTNNNN
 951  NGAAANNNTT GGGGCCNNTT TTNCCCCCCC CNNGGGGNAA AATNNANTTG
```

There was no significant homology between the sequenced flanking DNA (the ends of the subclones) and known sequences on the Genbank/EMBL databases. Sequencing of the entire length of the clones in the hope of finding known sequence was both beyond the scope of this study and would in any case have been premature, in that no defined function had yet been assigned to the flanking DNA isolated. The subcloned fragments were therefore partially digested with *Sau 3A* and cloned 'shotgun' into pBluescript (digested with *Bam HI* and phosphatase treated). Recombinant colonies (determined by blue-white selection) were picked and grown as a library in 96 well plates. Random clones were picked for sequencing and the sequence compared with known sequences using Pearson-Lipman searches.

Again, with the exception of plasmid sequence, no significant homologies were found between DNA sequence obtained and known sequence. It can therefore be reasonably concluded that the D8 plasmid, as with the C6 plasmid has integrated in a region which has not been previously defined.

4.2.5 Attempted Identification of the C6 and D8 loci by Fluorescent *in situ* Hybridisation

As partial sequencing of the clones obtained from the D8 and C6 cell line yielded no clue as to the identification of the locus into which the plasmids have integrated an attempt was made to determine the nature of the integration sites by Fluorescent *in situ* Hybridisation (FISH). Colcemid (Sigma) was added to exponentially growing D8 and C6 cells and cells were incubated at 37°C for

exactly 1 hour. Cells were then incubated in hypotonic buffer, fixed and dropped onto cover slips to form metaphase spreads as described (Materials and Methods).

Inspection of the metaphase spreads obtained under phase-contrast revealed that both cell lines contained considerably more chromosomes than the 39 expected in a mouse cell line. This aneuploidy, deletion of some chromosomes and duplication of others is extremely common where cells are maintained in culture over long periods. This would be unlikely to affect the locus of integration since cells were routinely passaged through gpt selection medium in order to eliminate any variants from which the plasmids were deleted. More importantly, the chromosomes were also all very small, acrocentric and virtually impossible to tell apart from one another by eye.

The slides were probed with D8 flanking DNA from clones 13 and 14 and also plasmid sequence. It was hoped that in the C6 cell line the plasmid sequence would hybridise to the C6 locus and the D8 flanking sequence would hybridize to the D8 locus. Thus, by comparison to the signal obtained for the D8 cell line (in which plasmid and D8 flanking sequence would be expected to hybridize to a single locus) the chromosomal location of the two plasmids could be compared. The comparative use of commercially available chromosome paints should allow identification of the chromosome(s) of interest in the likely even that they could not be identified by their size and shape.

This experiment was attempted several times and in each case no signal apart from background staining was ever detected. This is almost certainly due to a combination of two factors. Firstly, the probes were not optimal, in that the D8 probes contained at least one A/T rich tract and possibly some satelite sequence

(appendix) which would have potential to hybridize to many loci. Additionally, the plasmid probe (6kb) is at the lower length limit of detection by present technology. Secondly, the resources available for FISH were oriented toward detection of multicopy telomeric repeats. Successful detection of a single copy 6kb fragment would probably require confocal microscopy and computer assisted background subtraction. None of this state of the art technology was readily available, so no further attempt was made to analyse the transfectants by this method.

4.3 Discussion.

The use of λ-DASH has yielded three independent genomic clones containing DNA flanking the D8 cell line and a single genomic clone from the C6 cell line. These clones, all greater than 15kb in length, are far longer than could be expected from the more rapid methods described above and exonerate the conclusions of Nick Sutcliffe (Leicester University Thesis 1995) that lambda cloning vectors would be the most productive and reliable method available for obtaining genomic clones from the D8 cell line.

Restriction mapping of the clones shows that they are fully representative of the loci within the two cell lines from which they were derived and that no rearrangment either within the loci or between heterogenous DNA fragments to form chimeras occurred during the cloning process.

The high A/T content of the 9kb Sal-Sal subclone (figure 4.6) is of interest for several reasons. A/T rich 'linkers' of between 300 and 3000bp are distributed throughout avian and mammalian DNA at average distances of 10-30kb apart (Moreau, et al., 1981). A/T stretches are found at the end of the α and β-globin

gene clusters (Dolan, et al., 1981) and also flanking the actin (Firtel et al 1979) and histone (Grosschedl and Morf, 1981) genes.

For many years it has been widely thought that the function of A/T rich stretches is to punctuate the genome into discrete functional domains. This was initially demonstrated by electron microscopy of partially denatured genomic DNA. Bubbles in the duplex strand corresponding to A/T rich tracts, which melt more easily than random sequence, were seen at extremely regular intervals along the genomic DNA. The same denatured regions were also seen in clones of the Drosophila P1 protein and chicken β-globin DNA. In both cases A/T rich regions co-map with the limits of pre-mRNA and in the case of the α-globin cluster, co-map with regions of DNase I hypersensitivity and Msp I restriction site methylation (Moreau, et al., 1982)

A/T richness is also a consensus feature of regions of the genome which associate with the nuclear matrix (also known as the nuclear scaffold when referred to in metaphase chromosomes (Paulson and Laemmli, 1977) and also depending on the assay used to detect them). It has been widely postulated that this is a major mechanism by which the genome is separated into functional domains. Theoretically, anchorage to the nuclear scaffold allows delimitation of regions of DNA prone to unwinding, melting and torsional strain.

As is the case with the B1 element in the clone obtained from the C6 cell line, the presence of the A/T tract, or indeed a matrix attachment region, is largely inconsequential in that it cannot in itself reasonably explain the high expression levels of the D8 plasmid. If an A/T tract or a Matrix Attachment Region *per se* caused the effect then occurrence of the high expression level in stable

transfectants, as seen in the D8 and C6 cell lines should be far more frequent than that observed. The fact that A/T tracts are spread uniformly throughout the genome also means that few conclusions can be drawn from the sequence data *per se* as to the genetic or physical location of the plasmid, or to the function of the A/T rich tract. It is clear that not all A/T rich stretches of DNA have the same function, specifically genetic punctuation marks or matrix attachment regions and indeed many A/T tracts which probably have no effect on transcription at all. Thus an analysis of the function of the clones obtained from the D8 and C6 cell lines (Chapter 5) was required before any further speculation was possible as to the cause of the high level of transcription in the C6 and D8 cell lines.

Chapter 5
Assessment of the function of the sequences flanking the integrated plasmid in the C6 and D8 cell lines.

A central aim of this study was to assess the biochemical nature of the extremely high level of lysozyme expression by the plasmids in the C6 and D8 loci. To this end, once clones and partial sequence data for the DNA flanking the plasmids were obtained, an attempt was made to decipher the cause of the unusually high lysozyme expression levels in the two cell lines.

The mechanisms of control of eukaryotic gene expression can be viewed as a interconnecting and arguably interchanging hierarchy. DNA methylation is unquestionably at the top of the hierarchy and has been described as the 'bulk control' of gene expression (Gasser and Laemmli, 1987) In somatic cells methylated DNA is transcriptionally inactive (heterochromatin). The methylation pattern is passed on accurately to all daughter cells. Methylation, however, appears only to be involved in the maintainence of allele specific gene expression patterns, whereby only maternal or paternal alleles of specific genes are expressed, and plays only a minor role in the temporal or tissue specificity of gene expression. Thus, in the study of the C6 and D8 transfectants, investigation of the methylation patterns of the loci was never considered, and it was always reasonably assumed that both loci were hypomethylated.

A central mechanism of gene expression control is at the level of chromatin structure (Chapter 1). It had been suspected since their discovery that histones play a key role in the control, or at least the repression of gene expression.

Regions of transcriptionally active chromatin generally have a more open chromatin structure than transcriptionally silent regions. This is reflected in sensitivity to nucleases such as Micrococcal nuclease and DNase I (as characterised by a diminishing signal in Southern blotting when prepared nuclei are exposed to nucleases). Regions of chromatin spanning *cis*-acting control elements such as Locus Control Regions and many promoter and enhancer elements are, when activated, hypersensitive to DNase I. This is demonstrated by cleavage of the DNA at a mappable site (similar to that for a restriction enzyme) when isolated nuclei are incubated with very low concentrations of DNase I. This is widely assumed to reflect the fact that nucleosomes are slightly displaced from their normal conformation or are completely missing from the chromatin fibre. Thus, since DNase I sensitivity and DNase I hypersensitivity are diagnostic features of transcriptionally active chromatin and cis-activating elements respectively, DNase I sensitivity assays present a logical first step in the identification of the putative *cis*-elements involved in the the high expression of the D8 and C6 plasmids.

5.1 Assaying the DNase I sensitivity of the chromatin flanking the C6 and D8 loci

DNase I sensitivity assays were performed as previously described (Siebelist, et al., 1984). Briefly, nuclei were isolated as previously described (Materials and Methods) and incubated with DNase I in the presence of 5mM $CaCl_2$ and 1mM $MgCl_2$ The difficulty in judging the amount of DNase I to add to the nuclei (due to differences in activity between batches of DNase I), meant that pilot experiments were required to determine the quantity of DNase I to add to the reaction. Serial two-fold dilutions of 1mg/ml DNase I were added to separate aliquots of nuclei and were incubated at 37°C for 30 minutes. The concentration of DNase I which caused the almost total disappearance of the lysosyme specific

Nco I fragment in the C6 cell line was chosen for a time-course experiment (data not shown). In subsequent experiments the concentration of nuclei was kept constant and the volume of DNase I added was increased in proportion with the volume of nuclei.

A time-course digestion of nuclei was carried out as described (Figure 5.1; Materials and Methods). DNA was subsequently isolated (Materials and Methods) and digested to completion with *Nco I*, electrophoresed through a 1% agarose gel and transfered to nylon membranes.

Initially, hybridisation was attempted with probe sequences flanking the D8 and C6 loci. This would have been of considerable benefit for this assay because the whole experiment could have been carried out using the same samples, obtained from a single cell line (C6, D8 or J558L) using probes which would hybridise at each of the relevant loci. This would mean that DNase I sensitivity assays for both the C6 and D8 loci could be carried out in a single experiment. However, interpretation of the Southern blots obtained proved difficult owing to background bands and smearing

The source of this problem was almost certainly the presence of multicopy sequence and possibly also due to the A/T richness of parts of the D8 locus. Attempts to remove multi-copy sequence from the probe by pre-reassociation with sheared mouse genomic DNA (C_0t suppression) failed to alleviate this problem, in that there appeared to be little compromise between a high background smearing and extremely low signal in Southern blotting (data not shown). Similarly, attempts to use small Sau3AI subclones of the D8 Lambda clones proved equally unsuccessful. The only subclones to yield suitable

signal:background levels in Southern blotting gave identical patterns as plasmid probes, suggesting that they contained only plasmid DNA. These data suggest that, at least in the D8 case, that the clones obtained contain little single copy sequence. It therefore proved necessary to use probes specific to lysosyme and gpt sequences.

Figure 5.1 shows the result of the time course DNase I digestion on the C6 cell line. No DNase I hypersensitive site can be identified, as shown by the presence throughout the time-course of single bands for each probe, of identical size as non-DNase I treated naked DNA (Figure 5.1). It should be noted that the presence of the second, larger band obtained with the gpt probe is due to the repeated gpt sequence at the end of the plasmid insert (Chapter 3). After exposure to X-ray film the nylon filters were stripped and re-hybridized with a probe specific to the mouse β-globin LCR hypersensitive site III, generated by touchdown PCR out of mouse genomic DNA using primers mHS3for and mHS3rev appendix. In the adult mouse, the β-globin LCR is only active in erythroid cells and would therefore be expected to be DNase I resistant in mouse myeloma cells. It therefore provides a control of DNase I sensitivity, against general DNA degradation by DNase I.

There is clearly a reduction in band intensity for both gpt and lysozyme probes as the time course proceeds. This reduction in intensity is less marked with the mouse β-globin hypersensitive site III probe (mHS3). This is best seen by comparison of band intensity from naked DNA and band intensity of DNase I digested bands. Both the lysozyme and gpt bands at 30 minutes are virtually invisible when compared to the naked DNA, whereas after 30 minutes digestion the band obtained with the mHS3 probe is of comparable intensity to that of

naked DNA. This suggests that the diminishment of signal as the digestion proceeds genuinely reflects an open chromatin structure of the C6 locus and does not represent general DNA degradation due to artifical denaturation of chromatin during preparation of the nuclei.

The same case exists with the D8 cell line, although the data is arguably less convincing (Figure 5.2). Less reduction in gpt and lysozyme band intensity were obtained through the time-course, although the pattern obtained for the mHS3 probes is similar, if not more reduced through the time-course than in the C6 cell line. Again, comparison of the ratios between band intensities at 30 minutes and naked DNA suggest that the D8 locus is more DNase sensitive than the mouse β-globin HS3, although it could reasonably be argued that the D8 locus is less DNase I sensitive than the C6 locus. Such a comparison can only be made if it is assumed that the β-globin HS3 is identical in both cell lines.

It should be emphasised at this stage however that the mHS3 probe gave substantial background smearing in Southern blotting. Since gpt and lys probed filters were stripped and re-probed with mHS3, this background can only be due to the probe itself. While it can be argued, as above, that the mHS3 probe demonstrates as expected less DNase I sensitivity than the C6 and D8 loci, a contradictory interpretation could be that the increased ratio between naked DNA and the 30 minutes digestion sample could be largely due to background hybridisation. This fact does not detract from the primary conclusion that no specific DNase I hypersensitive site was found within the boundaries of the *Nco I* sites of either locus.

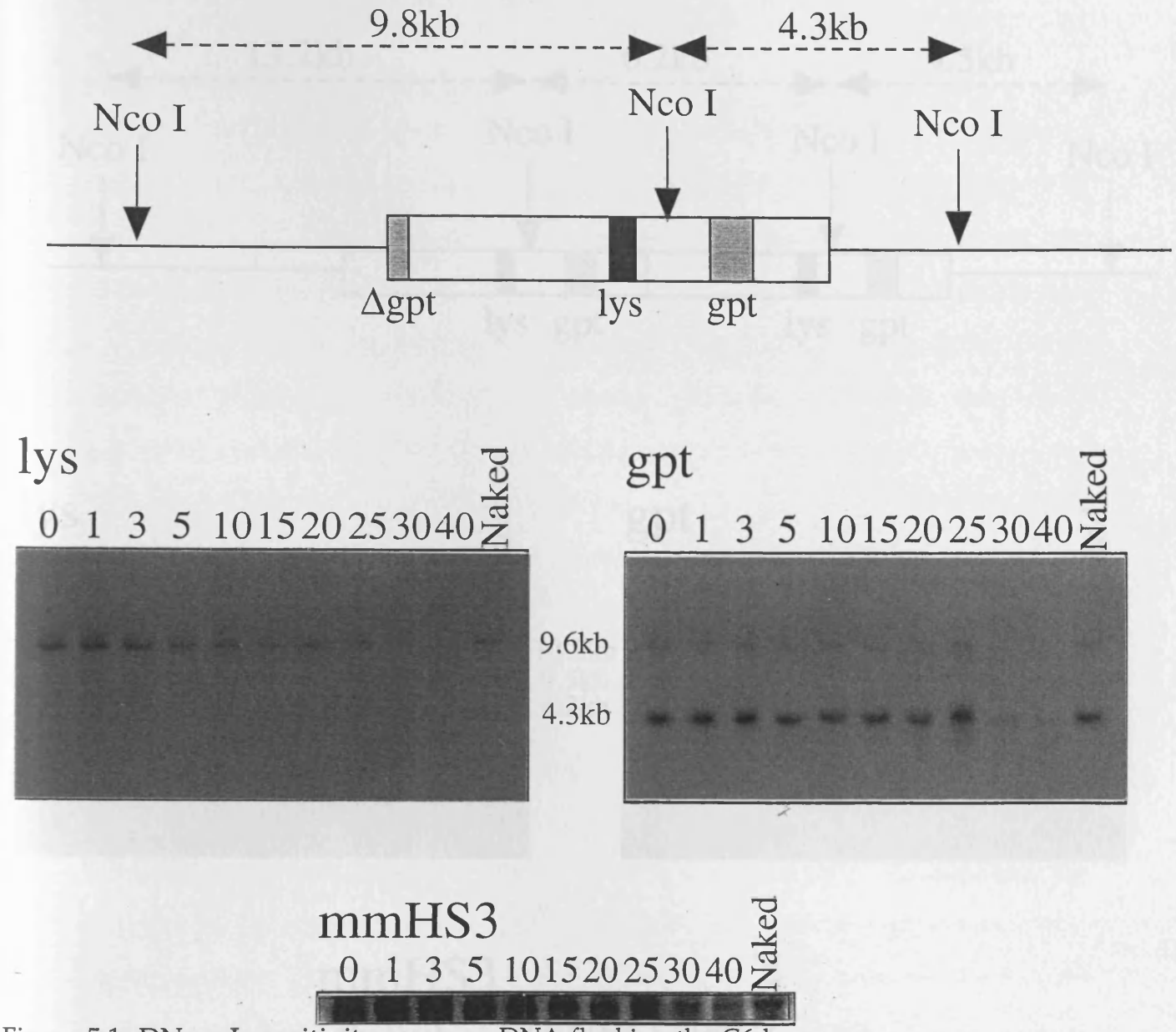


Figure 5.1 DNase I sensitivity assay on DNA flanking the C6 locus

Nuclei were isolated by cell lysis and incubated with 5μg/ml DNase I in the presence of 5mM $CaCl_2$ and incubated at 37°C for the times stated (minutes). DNase I was inactived by immediate addition of excess EDTA and storage on ice. DNA was isolated as described (Materials and Methods), digested to completion with *Nco I*, separated on a 1% agarose gel and transfered to Hybond N^+ nylon membrane. Southern hybridisation was carried out using probes specific to lysozyme and gpt generated by random hexanucleotide labelling of PCR products from primers lys4:lys3 and gpt3:gpt7 respectively. Filters were subsequently stripped and re-probed with a probe specific to Hypersensitive site 3 of the murine β-globin LCR (this site is known to be DNase I resistant in B cell lines). Naked=5μg naked DNA, not treated with DNase I, present as a reference to the expected size of Nco I restriction fragments.

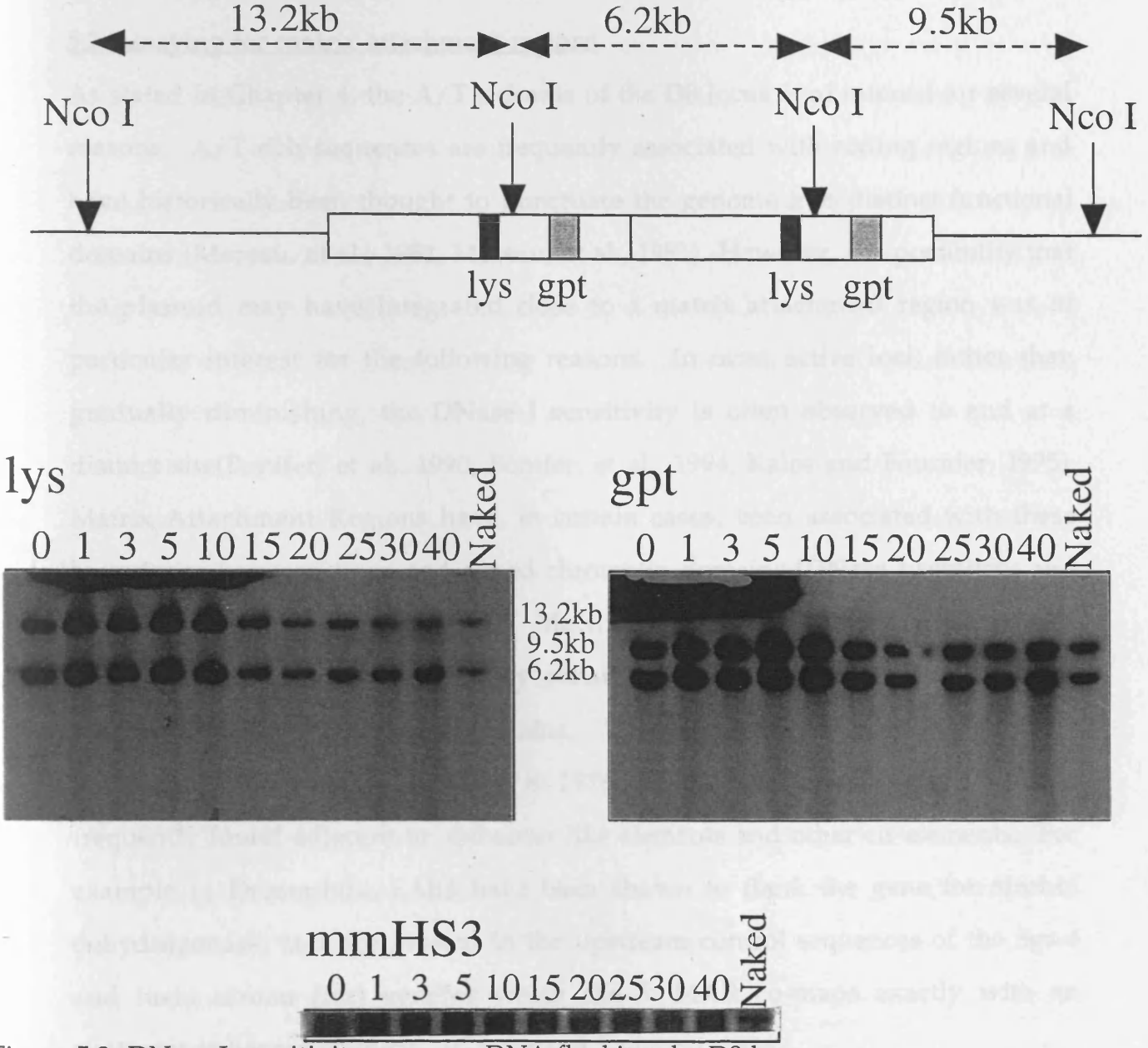


Figure 5.2 DNase I sensitivity assay on DNA flanking the D8 locus

Nuclear isolation, and subsequent assay were carried out as described for the C6 locus (figure 5.1). The 6.2kb fragment corresponds to the internal plasmid sequence with the tandem repeat.

5.2 Assaying for matrix attachment regions

As stated in Chapter 4, the A/T richness of the D8 locus is of interest for several reasons. A/T rich sequences are frequently associated with coding regions and have historically been thought to punctuate the genome into distinct functional domains (Moreau, et al., 1982, Moreau, et al., 1981) However, the possibility that the plasmid may have integrated close to a matrix attachment region was of particular interest for the following reasons. In most active loci, rather than gradually diminishing, the DNase I sensitivity is often observed to end at a distinct site(Bonifer, et al., 1990, Bonifer, et al., 1994, Kalos and Fournier, 1995). Matrix Attachment Regions have, in certain cases, been associated with these boundaries between open and closed chromatin domains (DNase I sensitive and DNase I insensitive respectively). Matrix attachment regions are thought to facilitate this 'boundary effect' by constraining the chromatin into discrete topological and/or torsional domains. This looping of chromatin was first described by Benyajati and Worcel in 1976. Matrix attachment regions are also frequently found adjacent to enhancer like elements and other cis-elements. For example in Drosophila, SARs have been shown to flank the gene for alcohol dehydrogenase, and are present in the upstream control sequences of the Sgs-4 and fushi tarazu (ftz) genefor which the 5' MAR co-maps exactly with an upstream enhancer element. (Gasser and Laemmli, 1986) .

The role of matrix attachment regions in gene expression control has been best demonstrated in the chicken lysozyme gene. Matrix attachment regions (referred to in the chicken lysozyme gene as 'A' elements) have been shown to co-map exactly to boundaries of the active chromatin domain (Phivan and Stratling, 1988). Also, transfection experiments in Drosophila (Nabirochkin, et al., 1998),

plants (Miynarova, et al., 1994) and mammalian cells (Phi-Van, et al., 1990) ((Phivan and Stratling, 1996)) have shown that the 'A' elements facilitate high level, position-independent reporter gene expression to constructs carrying appropriate promoter and enhancer sequences. Thus, the chicken lysozyme MARs are neither tissue-specific nor species-specific. This is also the case for most other matrix attachment regions described to date. This undoubtedly limits their potential as controlling elements, as it is difficult to envisage how constitutively active elements *per se* can mediate a temporal or tissue-specific gene activation. However, all MARs still have a potential role in conjunction with other control elements such as enhancers and locus control regions particularly in assisting in loop formation between cis-acting elements and thereby defining distinct chromatin domains.

Of particular interest is the fact that Matrix Attachment Regions are abundant in the immunoglobulin heavy chain locus (Cockerill, 1990), and indeed flank the wild type immunoglobulin heavy chain enhancer. (Cockerill, et al., 1987) The enhancer is truncated in the plasmids pLysSV40 and pLys$^+$ such that both MARs mapped by Cockerill are removed. This truncation had no rationale behind it other than to facilitate manipulation of the plasmid. The removal of the flanking matrix attachment region appeared to have no effect on the efficiency of the enhancer and appeared to broaden the cell specificity on the enhancer's activity (observations from this laboratory). The Matrix Attachment Regions within the IgH locus have been implicated in both switch recombination and activation of the heavy chain genes (via enhancer activation). This aside, there is little data available as to the role of the matrix attachment regions around the endogenous immunoglobulin heavy chain locus.

Given the lack of DNase I hypersensitive sites around the D8 or C6 loci, a possible interpretation of the high expression is that some cis-acting element other than an enhancer or a LCR close to the integration site is causing the IgH enhancer to function at, or close to the same activity as in the endogenous immunoglobulin heavy chain gene. It is possible that integration of the plasmid within appropriate *cis*-elements including any putative MARs could be mimicking the biochemical environment in which the IgH enhancer normally functions. As mentioned above there is a strong precedent for this model from studies in Drosophila (Phi-van, et al., 1990) in which 'A' elements from the chicken lysozyme gene mediated greatly enhanced expression of a heterologous gene when they flanked both sides of the construct, and mildly enhanced expression when flanking only one side if the construct. It is possible that the plasmid integrated into the D8 or C6 cell line in such a way that MARs on one or both sides of the plasmid serve to protect the plasmid from external chromatin effects and thus allow the IgH enhancer to function as it would in its endogenous locus.

5.2.1 Testing the ability of long PCR products from the D8 clones to bind isolated nuclear matrix *in vitro.*

This assay tests the ability of 5' end labelled DNA fragments to bind isolated (DNA free) nuclear matrix in vitro. Non-specific DNA-protein binding is blocked by the inclusion in the incubation of non-labelled *E.coli* DNA as a competitor. The method is summarised in figure 5.3

Nuclear matrix was isolated from J558L cells as described previously [Materials and Methods; (Cockerill and Garrard, 1986)]. Briefly, cells were lysed by mechanical disruption in isotonic buffer and nuclei separated from cell debris by

centrifugation through a cushion of 2M sucrose. DNA is digested with DNase I and then DNA fragments and other nuclear debris are then extracted from nuclear matrix with high salt buffer (2M NaCl).

Test DNA fragments were designed such that both sides of the integrated D8 plasmid could be tested for binding to the isolated nuclear matrix. The left hand side of the D8 locus (the 8kb *Sal-Sal* fragment; figure 4.6) was tested using a PCR fragment obtained by Long PCR using M13 universal primers, specific to pBluescript . A test fragment from the other side of the locus was obtained by long PCR out of D8 λ-clone 13 (Chapter 4) using primers lys3 (lysozyme gene specific) and T7 (λ-DASH specific; Figures 4.6 and 3.5).

Fragments to be tested were end-labelled by incubation with polynucleotide kinase in the presence of $\gamma[^{32}P]$-ATP (Materials and Methods). Labelled fragments were then incubated with isolated nuclear matrix in the presence of 500μg/ml unlabelled *E.coli* DNA. The unlabelled competitor DNA is added to prevent non-specific interactions between labelled fragments and nuclear matrix. After incubation, nuclear matrix was sedimented by centrifugation as described (Materials and Methods) and equal fractions were separated by agarose gel electrophoresis.

Figure 5.4 shows that the PCR product from the Sal I-Sal I subclone (left hand side of the diagram) appears to co-sediment with isolated nuclear matrix, whereas the right hand fragment failed to bind. Approximately 5% of total count were present in the pellet fraction. This level of binding is comparable to binding of MAR-containing fragments obtained by other workers in the presence of 500μg/ml of competitor DNA (Cockerill and Garrard, 1986).(Miynarova, et al.,

1994) No binding of the right hand fragment was detected even after over-exposure of the autoradiograph. This suggests that the binding of the left hand fragment is both real and specific, and obviates the need for a *bone fide* negative control DNA fragment such as radio-labelled pBR322.

Figure 5.3 Schematic summaries of the two MAR assays employed in this study.

A. To test the binding of DNA to nuclear matrix in isolated nuclear halos. Cells are lysed by Dounce homogenisation and washed in isolation buffer (Materials and Methods). DNA-MAR interactions are stabilised by incubation at 37°C for 20 minutes. Histones are removed by the gradual addition of extraction buffer and the resulting 'nuclear halos' resuspended in digestion buffer. After digestion the matrix bound and unbound DNA fragments are separated by centrifugation (Materials and Methods). DNA is subsequently isolated from each sample by phenol-chloroform extraction, electrophoresed on a 1% agarose gel, and Southern blotted as described (Materials and Methods)

B. Cells are washed thoroughly in PBS and homogenised in a Dounce homogeniser. Nuclei are initially isolated by standard centrifugation and are purified further by centrifugation through a 2M sucrose cushion. DNA is digested with DNAse I and nuclear matrix extracted with high salt (2M NaCl) buffer. Fragments to be tested are end labelled with ^{32}P (Materials and Methods) and incubated with nuclear matrix in the presence of *E.coli* DNA (500µg/ml) as a competitor. Nuclear matrix and bound DNA is then pelleted by centrifugation, DNA is isolated from both samples by phenol extraction, electrophoresed through an agarose gel, transfered to Hybond N^+ membrane and detected by Phosphorimager analysis.

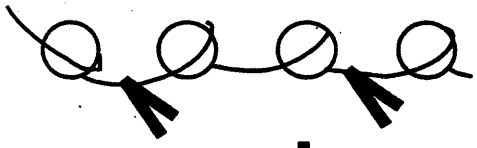
A

Wash cells and resuspend in Isolation buffer + digitonin -EDTA

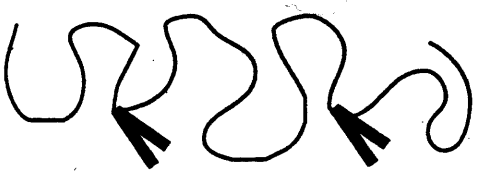
Lyse cells (20 cycles in Dounce homogenizer)

Pellet nuclei. Wash once with isolation buffer

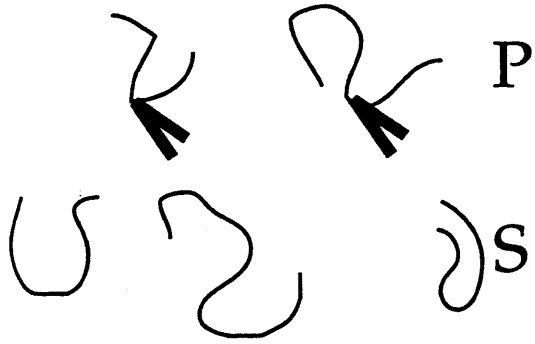
Incubate 37°C 20 minutes to stabilise nuclear matrix



Extract histones by gradual addition of extraction buffer



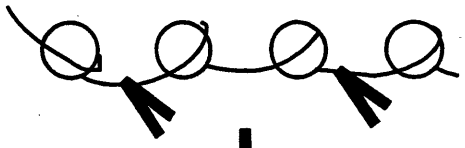
Resuspend in digestion buffer. Digest with restriction enzyme. Centrifuge



Purify DNA and Southern blot pellet (matrix attached) and supernatant (non-matrix attached) fractions

B

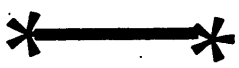
Wash cells in PBS. Resuspend in RSB homogenize with dounce homozenizer. Isolate nuclei by centrifugation. Wash. Purify nuclei further by centrifugation through 2M sucrose cushion.



Digest DNA to completion with DNase I

Extract nuclear matrix with high salt buffer

End label fragments to be assayed for binding to nuclear matrix



Incubate labelled fragments with isolated nuclear matrix in the presence of E.coli competitor DNA.

Isolate nuclear matrix and matrix bound fragments by centrifugation.

Purify DNA separately from supernatant and pellet fractions

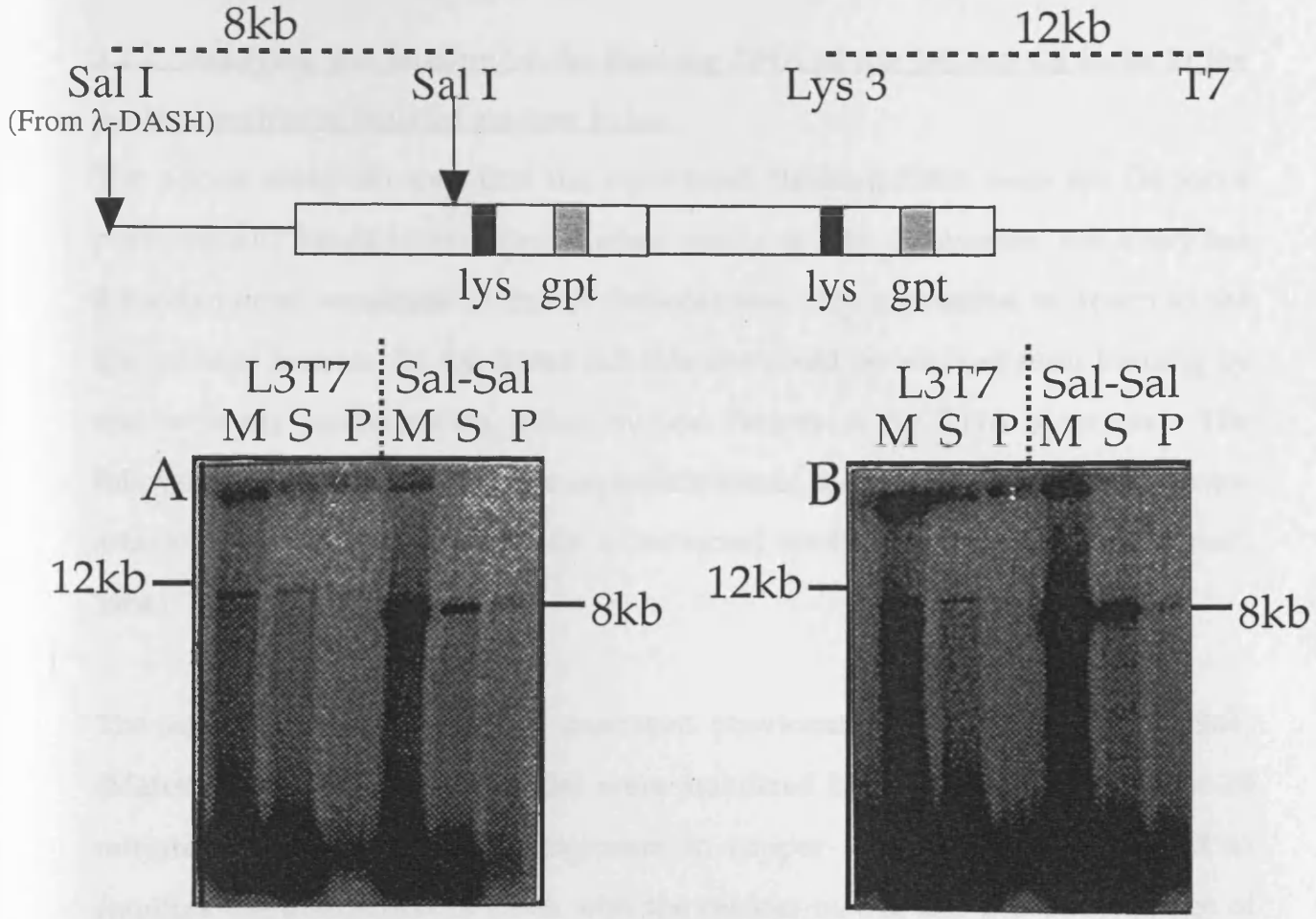


Figure 5.4 Assay of the ability of PCR fragments from the D8 locus to bind isolated nuclear matrix in vitro

Nuclear Matrix was isolated as described (Cockerill and Garrard, 1986); Materials and Methods. The L3T7 (right hand) fragment was generated by Long PCR from D8 clone 13 and the *Sal-Sal* (left hand) fragment was isolated by Long PCR from the pBluescript subclone using M13 universal primers. 2ng of each fragment was 5′ end labelled (Materials and Methods). Labelled fragments were incubated with nuclear matrix from $1.6x10^7$ cells in a total volume of 33µl with gentle shaking at 25°C for 2 hours in the presence of 500µg/ml E.coli competitor DNA.

Nuclear matrix was subsequently pelleted by centrifugation in a benchtop microfuge (13,000rpm for 30 mins). The pellet was washed once with RSB buffer without competitor DNA. DNA was extracted as described (Materials and Methods) and run on a 1% agarose gel which was then vaccuum dried onto Hybond N+ nylon membrane. Fragments were visualised by Phosphorimage analysis (A=24 hours exposure, B=72 hour exposure). M=marker, S=supernatant (non-bound DNA) P=pellet (bound DNA). Data shown are representitive of two independent experiments.

5.2.2 Assaying the binding of the flanking DNA of the D8 and C6 locus to the nuclear matrix of isolated nuclear halos.

The above assay showed that the right hand flanking DNA from the D8 locus preferentially binds to extracted nuclear matrix *in vitro*. However, this assay has a fundamental weakness in that it demonstrates only a potential to attach to the the nuclear matrix. In the intact cell this site could be blocked from binding by one or more nucleosomes, other nuclear factors or by DNA structure. The following assay, arguably more representitive of the *in vivo* state, is to examine attachment to the nuclear matrix in extracted nuclear halos (Mirkovitch, et al., 1984)

The assay was performed as described previously (Mirkovitch, et al., 1984) (Material and Methods). Nuclei were stabilized by incubation at 37°C for 20 minutes. Both heating and exposure to copper ions have been reported to stabilize the interaction of DNA with the nuclear matrix and prevent slippage of the binding site on exposure to buffer containing high salt.(Mirkovitch, et al., 1984) Nuclear membranes were then partially disrupted by incubation with buffer containing the mild detergent digitonin. Histones were extracted by incubation in an "extraction buffer" containing the mild and, more importantly, removable detergent lithium di-iodo salicylic acid (LIS). The removal of histones causes formation of the large DNA loops, characteristic of nuclear halos, which were partially visible under phase contrast microscopy. LIS is then totally removed by several washes in digestion buffer and the nuclei were digested with appropriate restriction enzymes.

Initial experiments whereby treated nuclei were incubated with restriction enzyme for 3 hours (Cockerill and Garrard, 1986) proved unsuccessful. This was due to unpredictability of the restriction enzyme cleavage and on other occasions due to total breakdown of the nuclei without accompanying restriction enzyme cleavage. In fact matrix degradation was occuring independently of any DNA breakdown (as determined by pulsed field gel electrophoresis; data not shown) and was therefore assumed to be due to protease activity.

These problems were circumvented with the following adaptations:
Firstly, the concentration of the protease inhibitors PMSF and trasylol in the extraction and digestion buffers were increased twofold. Secondly, a time-course restriction enzyme digestion was carried out. A time-course digestion should allow a point to be found where there is a compromise between optimal enzyme digestion and degradation of the nuclear matrix. However the time-course assay was slightly compromised by clumping of histone-extracted nuclei in the digestion buffer. This clumping is normally observed (U Laemmli; personal communication) but made equal aliquoting of DNA difficult. At the time points indicated (figures 5.5 and 5.6), an aliquot was removed from the digestion and sedimented by centrifugation as described (Materials and Methods). DNA was isolated from pellet and supernatant fractions, digested to completion with *Nco I*, and 5µg total DNA separated by agarose gel electrophoresis. DNA was subsequently transferred onto nylon membranes and Southern blotted (Materials and Methods).

Similar problems associated with hybridising with flanking sequence (inability to generate probes sufficiently low in background without loss of signal.) were encountered as for the DNase I sensitivity assays. Therefore, probes specific to

gpt and lysozyme (as described previously) were used to hybridize to flanking DNA on both sides of the C6 and D8 plasmids.

Flanking sequence to the integrated plasmid in the C6 cell line showed no binding to the nuclear matrix (Figure 5.5). As would be expected, before digestion (time=0), all DNA is found in the pellet. Within 90 minutes of digestion DNA from both sides of the plasmid is found in the supernatant. This demonstrates that there is no matrix attachment region surrounding the C6 plasmid (as far the *Nco I* sites) and also confirms that the plasmid itself has no affinity for the nuclear matrix. Known matrix attachment regions around the IgH enhancer and the chicken lysozyme gene are not present in the plasmid.

This provides a useful internal control for interpretation of the data obtained from the D8 cell line (Figure 5.6), in that the internal plasmid fragment (6.2kb) from the tandem repeat would not be expected to bind to the nuclear matrix. Again, as expected, all material was present in the pellet before digestion. Although digestion of the D8 nuclei is clearly less efficient than for the C6 nuclei, some plasmid and flanking DNA are released into the supernatant after 30 minutes. The fact that neither probe (gpt or lysozyme) shows a change in the ratio of flanking fragment to internal plasmid fragment as the time-course proceeds, and the fact that plasmid does not bind nuclear matrix (from the literature and the C6 experiment) show that neither side of the D8 plasmid is bound preferentially to the nuclear matrix in the intact nucleus. If the regions flanking the D8 locus did bind to the nuclear matrix then the ratio between the band corresponding to flanking sequence and the band corresponding to plasmid sequence would be expected to increase.

These data contradict those obtained from the binding assays on isolated nuclear matrix in which right hand fragment (corresponding to the gpt probe in figure 5.6) appeared to show some binding to nuclear matrix *in vitro*. The 13kb fragment hybridised in figure 4.6 spans the entire PCR product shown to bind nuclear matrix *in vitro*, hence this anomaly cannot be caused by the putative MAR being missing from the hybridised fragment. It is more likely, as alluded to earlier, that the potential matrix attachment region demonstrated in the *in vitro* binding experiment is simply blocked by other bound proteins in the intact nucleus.

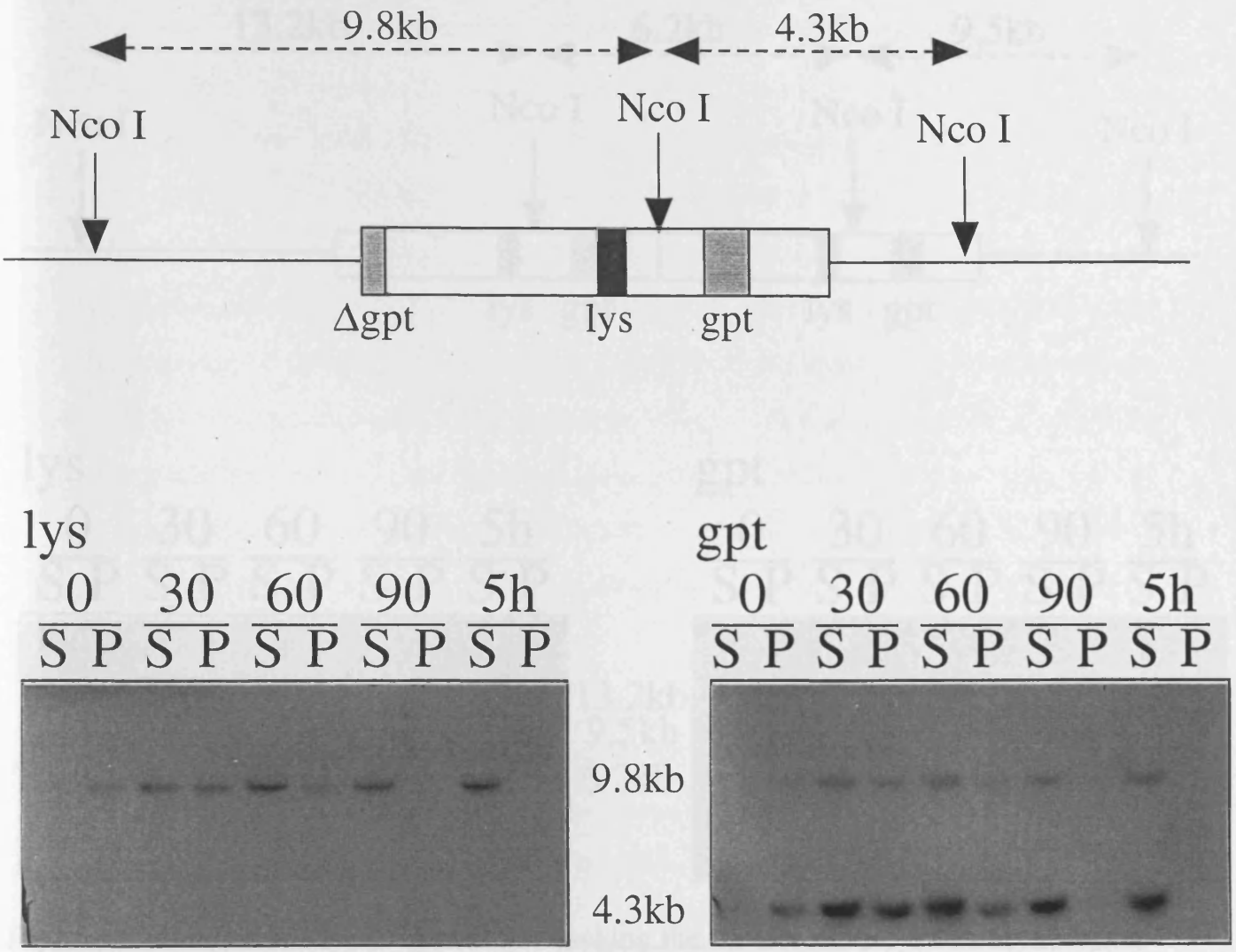


Figure 5.5 Assay of the binding of DNA flanking the C6 locus in isolated nuclear halos.
Nuclear halos were generated by LIS extraction of $1x10^8$ isolated nuclei as described (Mirkovitch, et al., 1984); Materials and Methods. *Nco I* was added at 500U/ml and aliquots of suspension were removed at the time intervals stated (minutes). Nuclei were pelleted by centrifugation (13,000rpm, 20 mins, room temp) and DNA was extracted from pellet and supernatant as described (Materials and Methods), digested to completion with Nco I, separated on a 1% agarose gel and Southern blotted. gpt and lysozyme specific probes were generated by PCR as previously described. P=pellet (scaffold bound); S=supernatant (non-bound).

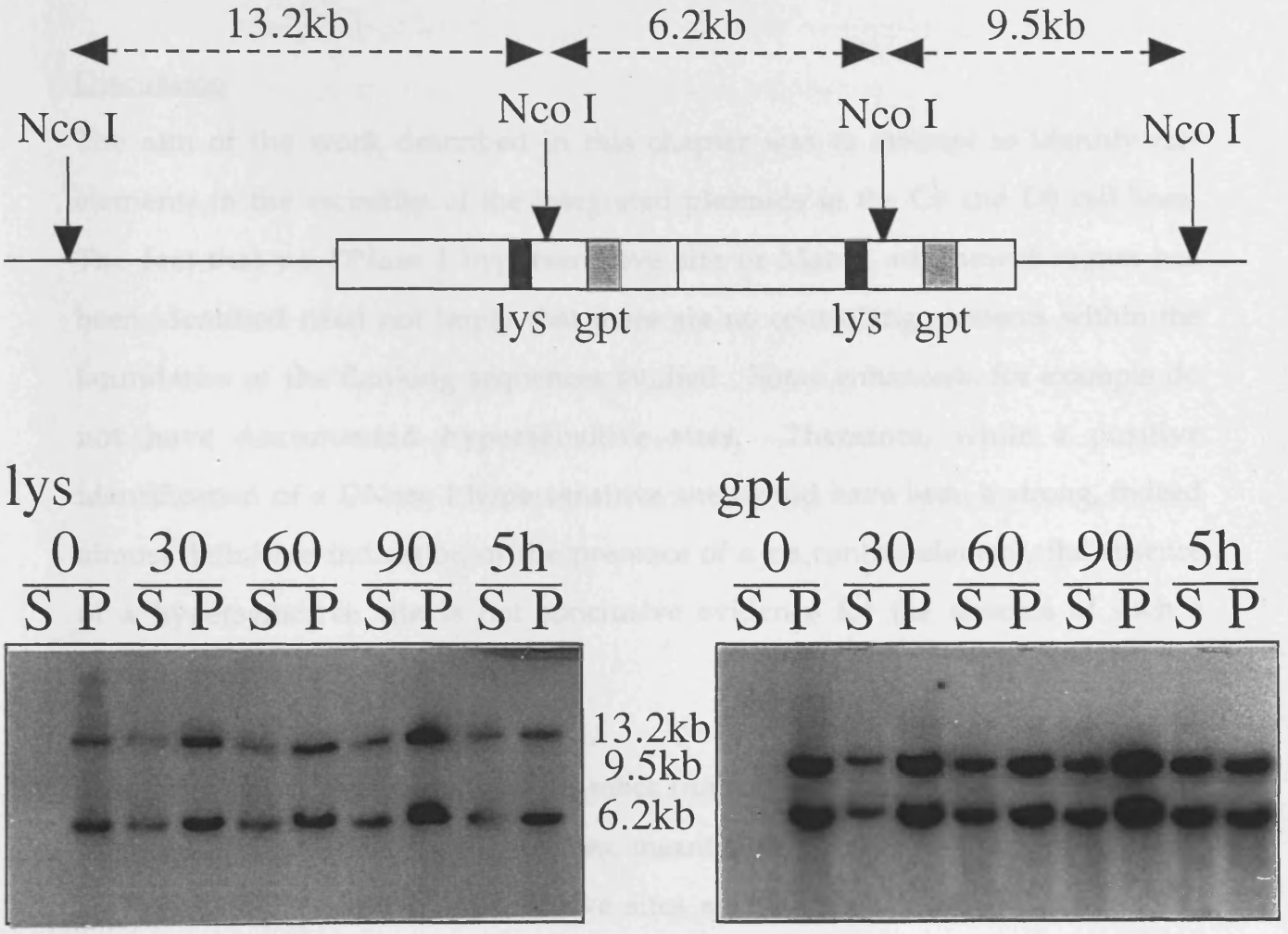


Figure 4.6 Assay of the binding of DNA flanking the D8 locus in isolated nuclear halos.
See Figure 4.5 for detailed legend.

Discussion

The aim of the work described in this chapter was to attempt to identify cis-elements in the vicinitity of the integrated plasmids in the C6 and D8 cell lines. The fact that no DNase I hypersensitive site or Matrix attachment region has been identified need not imply that there are no controlling elements within the boundaries of the flanking sequences studied. Some enhancers, for example do not have documented hypersensitive sites. Therefore, while a positive identification of a DNase I hypersensitive site would have been a strong, indeed almost definitive indication of the presence of a cis control element, the absence of a hypersensitive site is not conclusive evidence for the absence of such a control element.

The difficulty in generating useful probes from cloned flanking sequence, leading to the use of plasmid specific probes, meant that the area of flanking sequence screened for DNase I hypersensitive sites and matrix attachment regions was somewhat smaller than may have been desirable.

While it is possible that the studies have proved unfortunate in that the cis-elements lie just outside the region studied, it is equally possible that the elements influencing the high level of expression are some distance away. If the active loci in to which the plasmids have integrated was the same size as that for a typical LCR element then it could be argued that the study has proved extremely unlucky. However, the size of many active loci such as the immunoglobulins and the β-globin locus means that cis-elements could potentially be several kilobases away from the plasmid integration site.

It has also been reported that in gene 'knockout' experiments using contructs containing the strong promoters, the integrated vector can interact with LCR-like elements several megabases away and even with control elements on different chromosomes. In knockout experiments this leads to abrogation of gene expression at the endogenous locus and hence to an unrepresentative null phenotype(Pham, et al., 1996). For the purposes of this study, if a similar effect has occurred in the C6 and D8 transfectants then this would render the whole cloning approach of constructing libraries and probing for plasmid sequence a waste of time, as the high level of expression obtained is completely independent of the flanking DNA but, paradoxically, is still entirely position dependent.

Future Work.

This study has shown that there is no DNase I hypersensitive site or matrix attachment region within the *Nco I* fragments surrounding the integrated plasmids in the C6 and D8 cell lines. It would be wise, however, to assess the true effect of the λ clones obtained from these cell lines before they are discarded. One suitable approach would be to place the cloned fragments into a reporter vector and to retransfect them back into the J558L cell line. A vector for these transfections has been constructed and flanking fragments are currently being placed into it. The rationale for the design of the vector was that the effect of the fragments on both general transfection frequency and reporter expression levels within individual clones should be easily discernable from background levels. The absence of DNase hypersensitive sites within the cloned fragments may imply that the effect of the fragments on expression levels would be minimal (the clones would certainly not be expected to have LCR activity). Background expression is therefore to be minimised in preliminary experiments by the

absence of the IgH enhancer and also by the presence of only a truncated thymidine kinase promoter, as is present on the C6 and D8 plasmids.

Any further attempt to identify DNA sequences flanking the C6 and D8 plasmids, the most logical of which would be to chromosome walk through the D8 library already constructed using C_0t suppressed probes from C6 and D8 flanking DNA obtained in this study, would be greatly simplified by the use of stripped Southern blots from the DNase I and matrix attachment assays performed so far. DNA samples from each time point of the assays are also available. Therefore positive clones from the chromosome walks could be assessed directly for DNase I hypersensitivity and/or matrix attachment and the walking could be continued until such an element is reached. Thus, while no direct evidence of any interesting cis-elements around the plasmid has been obtained from this study, considerable advances towards the future identification of such an element in the future have been made.

Chapter 6.
Attempts to use the *Saccharomyces cerevisae* site specific recombinase, FLP, to place selected DNA sequences at the active C6 locus.

6.1 INTRODUCTION.

An attractive potential use of the C6 and D8 cell lines is to utilize the active loci to express cDNA sequences at high levels in J558 myeloma cells. Previous chapters aimed to clone the DNA flanking the plasmids and assay their effect on reporter gene expression. Another approach, running concomitantly in our laboratory, aims towards utilizing the loci by placing chosen DNA sequences at the active loci by homologous recombination mediated by the *Saccharomyces cerevisiae* enzyme, FLP ("flip") recombinase.

6.1.1 Flp recombinase is a member of the integrase family of site specific recombinases

Flp recombinase is one of a number of recombinases of the lambda integrase family. These include Cre recombinase, R recombinase and λ integrase itself. Cre (causes recombinaton) recombinase is a 38kDa protein of bacteriophage P1 which recognises a 34bp binding site, the lox P site (locus of recombination). The physiological function of cre recombinase is to maintain the plasmid form of P1 phage by resolving DNA dimers into monomers. R recombinase is encoded by the pSR1 plasmid of the budding yeast *Zygosaccharomyces rouxii* and performs a role identical to that of Flp recombinase.

Flp recombinase is encoded by the yeast 2μ plasmid. Its sole function in yeast is in the amplification of plasmid copy number prior to cell division. Like the other integrase family members, Flp binds in a sequence specific manner to a 34bp element known as the FRT (Flp Recognition Target site). The minimal FRT site consists of two 13bp inverted repeats flanking an 8bp core (figure 6.1). In the wild-type 2μ circle the FRT consists of the minimal recognition site plus an extra 13bp repeat on one side of the FRT. The extra repeat is essential only for integrative recombination. Excision and inversion occur with the minimal FRT site of two inverted repeats flanking the 8bp core with no loss of activity. (Lyznik, et al., 1993) Two Flp molecules bind to each FRT site and thus in the recombination reaction the enzyme binds to two duplex DNA molecules as a tetramer.

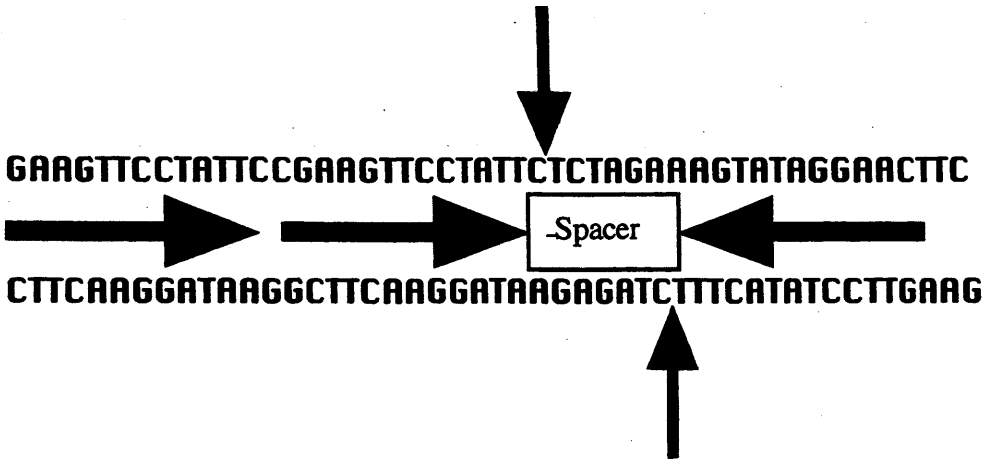


Figure 6.1. The FRT site
The FRT site consists of three 13bp symmetry elements (depicted by black arrows) and an 8bp core (spacer) region. Flp cleavage site are marked on the diagram with vertical arrows. Only the repeats flanking the spacer sequence appear to be essential for Flp activity. The function of the additional repeat is not determined (Sadowski, 1995)

6.1.2 The mechanism of FLP recombinase

All Integrase family members share an invariant tetrad of amino acids. These are an Arg^{191} His^{305} Arg^{308} triad and Tyr^{343} (numbers correspond to the amino acid sequence of Flp recombinase). Complementation studies have revealed that mutations in the RHR sequence could not complement each other; neither could Tyr^{343} mutants. However RHR mutants can be complemented by Tyr^{343} mutants and RHR triad mutants can be complemented by Tyr^{343} mutants (Pan, et al., 1993). These data suggest that the mechanism of FLP involves a composite active site, so that when two FLP monomers are bound, in one monomer the RHR triad is required and in the other monomer Tyr^{343} is used. Tyr^{343} catalyses cleavage of the DNA strand by forming a covalent bond with the 3′ phosphate at the cleavage site and acts as the leaving groups in the subsequent ligation to the recombining strand. Ligation requires the RHR triad, occurs independently of strand cleavage (Pan and Sadowski, 1992) and is mediated by nucleophilic attack by the 5′ hydroxyl group from the nicked strand on the 3′ phosphotyrosyl bond of the partner substrate.

6.1.3 FLP recombinase is active in higher eukaryotic cells.

Flp was first shown to be active in mammalian cells by O'Gorman et al in 1991. Flp was shown to activate a transiently transfected β-Galactosidase expression vector by deletion of a G418 resistance gene flanked by two FRT sites from the 5'UTR of the β-gal gene (figure 6.2). β-galactosidase activity can then be assayed either histochemically using the chromogenic substrate X-gal, or total β-gal activity can be assayed using standard assays. The neo-FRT system is now commercially available (Stratagene) and these constructs form the basis of FLP activity assays in this and many other studies. Since the first demonstration of

FLP activity in CV1 cells, FLP recombinase has been shown to be active in many cell types including plant protoplasts (Lyznik, et al., 1993), and embryonic stem cells (Dymecki, 1996), making the enzyme a potentially extremely powerful tool in the manipulation of whole animal and plant genomes.

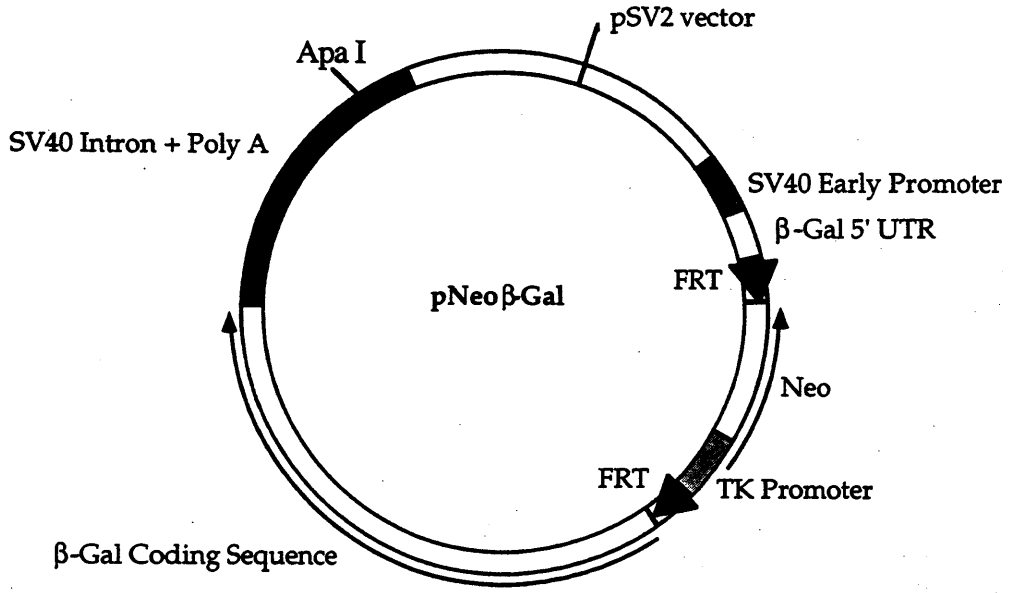


Figure 6.2 pNeoβ–gal. The plasmid stably transfected into the CV-1 cell line to generate the CV-1#5 cell line as a reporter of Flp activity.
The plasmid contains the β-galactosidase gene driven by the SV40 early promoter. A neomycin selectable marker driven by a thymidine kinase promoter and flanked by two FRT sites is inserted into the 5′ UTR of the β-gal gene which generates a frame shift in β-gal coding sequence. FLP activity excises the neomycin construct, restoring the reading frame of the β-galactosidase transcript. The single remaining FRT site does not disrupt the reading frame or the translation efficiency of the β-gal transcript

6.1.4 Placement of an FRT site at the active C6 locus

Previous work by Sarah Munson in our laboratory has led to the integration of a plasmid construct containing a single FRT site and a G418 resistance gene at the C6 locus. This was mediated by classical homologous recombination between the integrated plasmid pLysSV40 and pFRTNeoSV40 (Fig 6.3). This plasmid was designed to knock-out the gpt selectable marker already present at the locus within pLysSV40 and replace it with the Neomycin selectable marker. Thus, after electroporation with the plasmid pFRTNeoSV40, homologous recombinants were selected against the high background of random integrants by selecting for colonies in growth media containing G418 and 6-thioxanthine. 6-thioxanthine is a sulphoxyl analogue of xanthine which is converted by the action of XGPRT activity to 6-thioguanine which, when incorporated into DNA rapidly becomes toxic due to inhibition of transcription and DNA replication. Colonies which survived this selection for the $XGPRT^-$ Neo^+ phenotype were further tested for homologous recombination by Southern blotting.

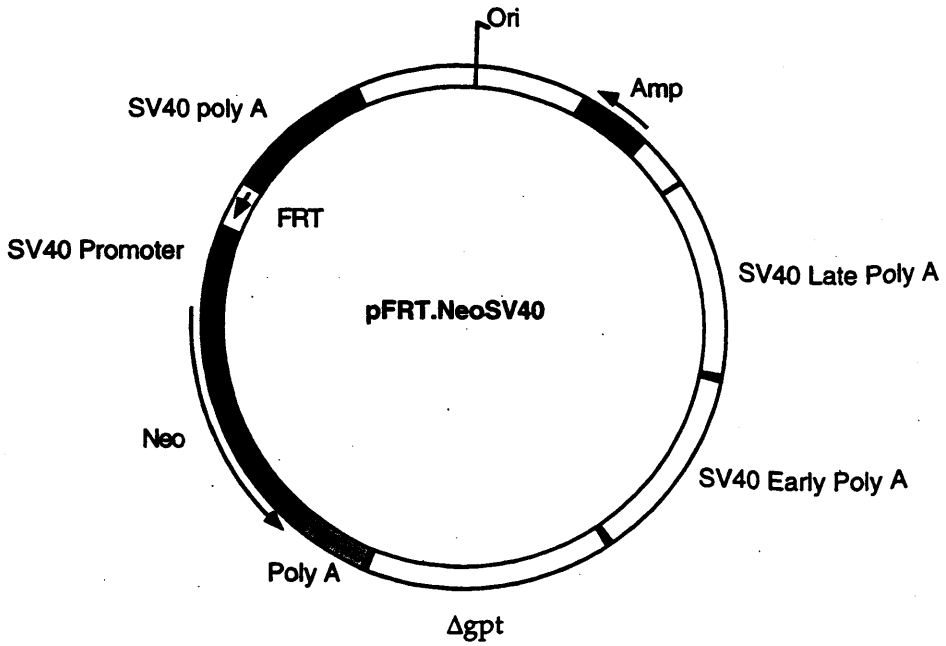


Figure 6.3 pFRT.NeoSV40.

The plasmid is based on pLysSV40 (the plasmid present in the C6 line line. There is a single FRT site between the SV40 polyadenylation site and the SV40 promoter. The lysozyme coding sequence is replaced with neomycin selectable marker. The plasmid is designed to "knockout" the gpt selection marker already present at the C6 locus so that homologous recombinants (cells in which the FRT site has been placed at the C6 locus) can be isolated by selecting for the Neo^+ gpt^- phenotype (see text for details)

6.1.5 Flp activity was undetectable in J558L cell lines and extremely low in CV1 cell lines

A major problem in utilising FLP in our cell lines which became evident shortly after the successful incorporation of an FRT site at the C6 locus was that no FLP activity could be obtained in transient transfections of J558 cells. and only very little activity could be obtained in CV1 cells, one of the cell lines in which FLP activity had been first demonstrated in a range of mammalian cells (O'Gorman, et al., 1991) In both cases FLP coding sequence was expressed from the plasmid pOG44 (Stratagene). This plasmid contains the cytomegalovirus promoter and the SV40 late polyadenylation signal. The aim of this chapter therefore was to attempt to optimise FLP expression to facilitate its use in homologous recombination in the mammalian cells used in our laboratory.

6.1.6 Epitope tagging of proteins as a means of separating protein expression from activity

This work began with studies on FLP expression in CV1 cells. There are many justifications for this. Firstly, expression and activity of FLP recombinase in this cell line is already documented (Ogorman, et al., 1991). Secondly, there is, readily available, a CV1 derived cell line expressing the SV40 large T antigen; the COS 7 cell line (Glutzman, 1981). Thus plasmid contructs, transfection techniques, optimal post-transfection incubation times can be tested in a background where at least some FLP activity can be expected. Thirdly, the CV1 and COS cell lines are adherant, and grow in monolayers. This makes transfection of DNA by DEAE-Dextran or Calcium Phosphate considerably more straightforward and efficient whereas transient transfection techniques in myeloma are notoriously inefficient. Adherance also renders the cells more amenable to histochemical β-

galactosidase assays and to immunocytochemial techniques without the need to immobilize cells by centrifugation onto slides or by culturing cells on poly-lysine coated cover slips.

The lack of detectable FLP activity in J558L cells and CV1 cells was evident from work by Sarah Munson in clones of these cells where an integrated pNeoβGal plasmid was the substrate for FLP (Fig 6.2). FLP mediated recombination excises the Neo gene from the 5'UTR of FLP, thus activating β-gal activity. Activity was assayed using the X-Gal based histochemical activity stain and by total β-galactosidase assays on cell lysates using a luminescence based assay (Materials and Methods; also see above).

The poor activity observed could be due to poor transfection, inefficient trancription or translation or aberrant cellular localisation of the FLP protein. There was however no means of separating protein production from enzyme activity since the only means of testing all of the above variables was to test the 'final stage'of β-galactosidase activity. Thus the first approach undertaken was to tag the FLP with a Myc 9E10 epitope tag (Evan *et al* .,1988). Such an approach was expected to facilitate the separation of FLP protein detection and FLP activity and is a logical first step towards explaining the low FLP activity in CV1 cells and, more importantly, the J558L cell lines.

Myc-9E10 is one of many epitopes routinely utilized in protein tagging. The more commonly used alternatives are described below:
The addition of polyhistidine tags to the N and C termini of proteins can facilitate purification on a nickel column as well as immunodetection using antibodies to poly histidine. Similarly, gutathione-S-transferase (GST) based

fusion proteins can be detected in western blotting and immunofluorecsence using antibodies to GST, and can also be affinity purified directly from lysed *E.coli* cells using glutathione-coated agarose beads. Recently, Green Fluorescent Protein (GFP) from the jelly fish *Aequorea victoria* has been developed as a protein reporter system. As well as allowing detection in western blotting and immuneoprecipitation using antibodies to GFP, GFP has an extra advantages of being fluorescent when stimulated with UV light. This fluorescence occurs in the absence of any co-factors and can be used to monitor the expression and localization of tagged proteins in real time without the need to fix cells prior to visualization. GFP can also be used to select for cells expressing the tagged protein using FACS analysis (fluorescence activated cell sorting). However all of these tagging systems utilize long and often unstable peptide sequences or protein domains. Given that at the strand exchange stage of the recombination reaction, the FLP is effectively bound to the two duplex DNA strands as a tetramer, in an initial study it was considered important that the tag was as small as possible, and that it would cause as little disruption as possible to the secondary structure of the flp monomers and the ability of FLP to form active tetramers. Although in theory any defined epitope for which a monoclonal antibody was available would be suitable, the use of myc-9E10 system is well documented and the epitope itself is aliphatic and so, it was reasoned should cause minimal disruption to the secondary structure of the FLP monomers or the quarternary structure of the recombination complex.

Secondary structure predictions by both the Chou-Fasman and Garnier-Osguthorpe-Robson algorithms suggest that the 9E10 peptide sequence favours alpha-helix formation. The N terminus of Flp was predicted to form coils and the C-terminal gave no definitive secondary structure prediction. Although

these predictions are notoriously unreliable, at face value, they were taken to suggest that the least disturbance to the secondary structure of Flp would be caused by placing the epitope tag at the carboxy end of the Flp coding sequence. Also, the C terminus of FLP was favoured over the N terminus because it was reasoned that a short peptide (as opposed to distinct domain such as GFP or GST) was more likely to affect protein folding during translation as the nascent peptide emerged from ribosome than if the epitope is placed at the C terminus where it should have less effect on folding of the nascent peptide chain. In fact, it has been demonstrated that there are many sites, including the C terminus, where residues can be deleted or inserted into the FLP protein without loss of activity. Interestingly, insertion of extra residues into the primary structure of Flp recombinase occasionally increased Flp activity *in vitro* (Amin and Sadowski 1994).

6.2 RESULTS

6.2.1 Placing the Myc-9E10 epitope coding sequence at the 3′ end of Flp coding sequence

A PCR-based approach was used to place the 9E10 tag at the C terminus of Flp. The flp coding sequence was amplified from the plasmid pOG44 (Stratagene) by PCR with primers FlpNBam and FlpCXho (appendix) containing *Bam H1*and *Xho 1* sites on the N and C terminal primers respectively, using a 12:1 mix of Taq (Advanced Biotechnology Limited) and PFU (Stratagene) polymerases. PFU polymerse is one of many thermostable polymerase with 3′-5′ exonuclease activity. While, in practice, this activity decreases overall reaction yield considerably, the rate of base mis-incorporation is reduced from 1.25×10^{-5} to 7.6×10^{-7} (Stratagene). The PCR product was digested with *Bam H1* and *Xho1* for 12 hours and cloned into *Bam H1* and *Xho1* digested "T7 plasmid tag C" plasmid

(figure 6.4). This vector is based on the mammalian expression vector pGEM2 (Promega) and was constructed in the laboratory of Richard Treisman (ICRF, Lincolns Inn Fields, London). The flp-9E10 construct was further subcloned into the plasmid pcDL-SRa296 (pSRα) (Takebe et al, 1988), which contains the SRα promoter (comprising the SV40 early promoter, the R segment and a truncated U5 sequence of the HTLV-1 LTR), the SV40 late gene splice junction, and the SV40 late gene polyadenylation signal. The SRα promoter is 10-30 fold more active than the SV40 early promoter alone in a wide range of cell types, including the B-cell line-Raji (Takebe et al 1988) and is the most active promoter for a range of mammalian cell types used in our laboratory (unpublished observations). The cloning procedure is summarized in figure 6.4.

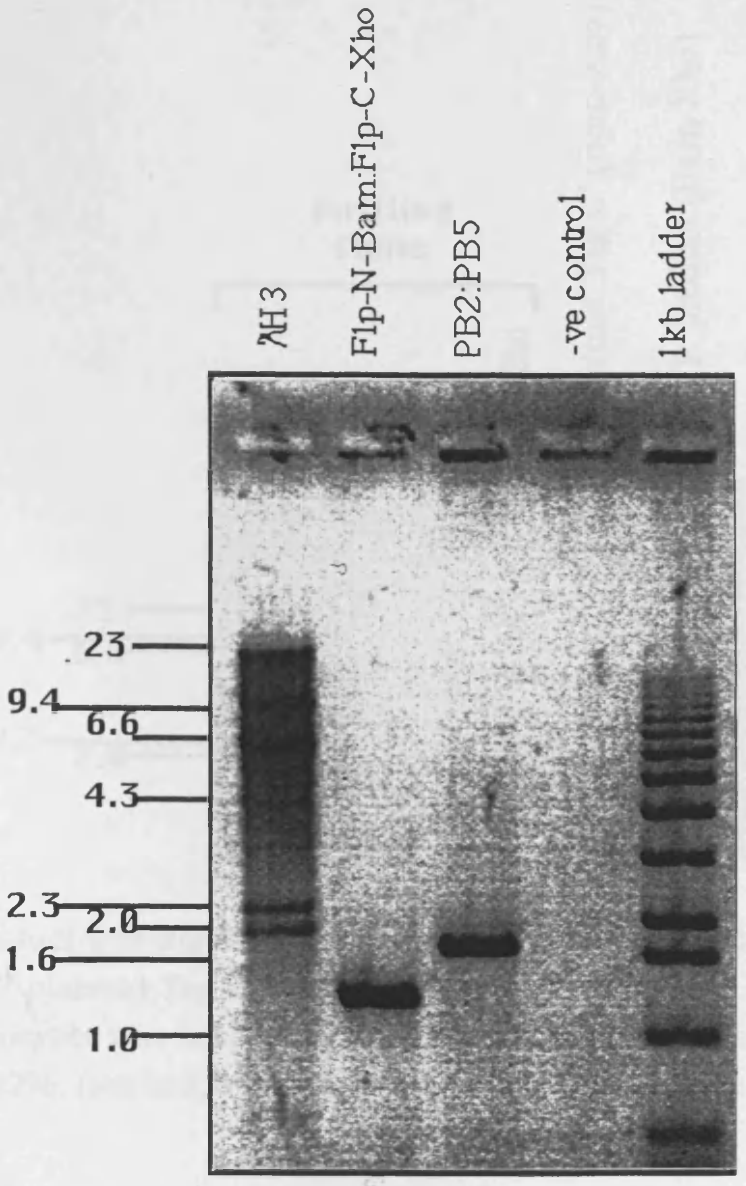


Fig 6.4

Primers Flp-N-Bam and Flp-C-Xho were used to PCR Flp coding sequence out of pTKFlp using a 12:1 mixture of Taq and Pfu polymerases. A separate PCR reaction from the same plasmid with PB2 and PB5 primers was used as a positive control. The Flp PCR product is the expected size of 1.3kb.

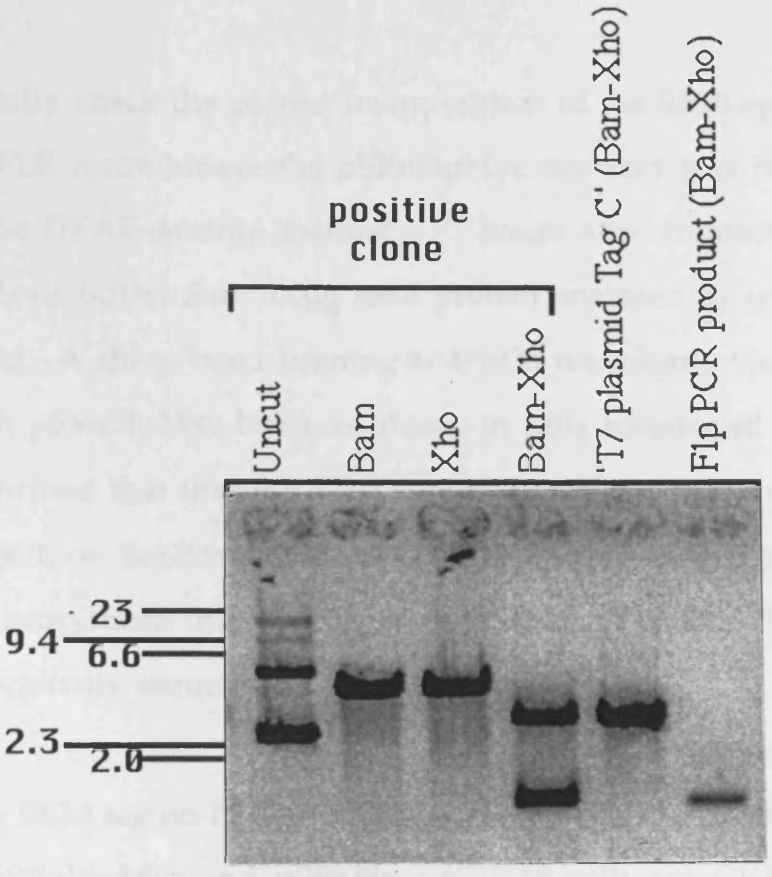


Fig 6.4 contd

The Flp PCR product was digested for 20h with BamH1 and Xho1 and ligated for 16 hours with 'T7 plasmid Tag C', also digested with BamH1 and Xho1.

The flp-9E10 construct was subsequently transfered by S. Uff into the expression vector pcDL-SRα296, (see text) via a BamH1-Xba1 digest.

In order to initially check the correct incorporation of the 9E10 epitope onto the C-terminus of FLP recombinase the pSRaFlpMyc contruct was transfected into COS cells by the DEAE-dextran method. 72 hours after tranfection cells were harvested into lysis buffer and 100μg total protein analysed by western blotting (data not shown). A sharp band running at 49kDa was clearly visable from cells transfected with pSRaFlpMyc but was absent in cells transfected with pSRaFlp. These data confirmed that that the 9E10 epitope is present in the correct reading frame and suggest, on the basis of the size of the epitope tagged protein, that no deletions were introduced into the Flp coding sequence during the cloning (the clone was subsequently sequenced by Sarah Munson).

The effect of the 9E10 tag on FLP activity was assayed by transfecting CV1#5 cells with 10μg of pSRaFlpMyc and pSRaFlp. (CV1#5 cells are CV1 cell (monkey kidney epithelial cells) into which the pNeoβ-gal plasmid (Stratagene; Fig 6.2) has been integrated, and which expresses the Neo gene at a high level (Sarah Munson; unpublished data). Activation of β-galactosidase by FLP-mediated recombination was assayed by both an X-gal based histochemical stain and by a total β-galactosidase assay on lysed cells.

A sub-confluent 9cm plate of CV1#5 cells were transfected with 5μg pSRαFlpMyc or pSRαFlp or pSRα plasmid only. After 72 hours cells were fixed and then incubated for 2 hours at 37°C under β-galactosidase activity stain solution (Materials and Methods; figure 6.5) The cells were then visualized directly by phase contrast microscopy. No difference in the number of blue (β-gal positive) cells between pSRαFlpMyc or pSRαFlp plasmids was observed (fig 6.5). No β-gal positive cells were observed with pSRα plasmid alone (data not shown).

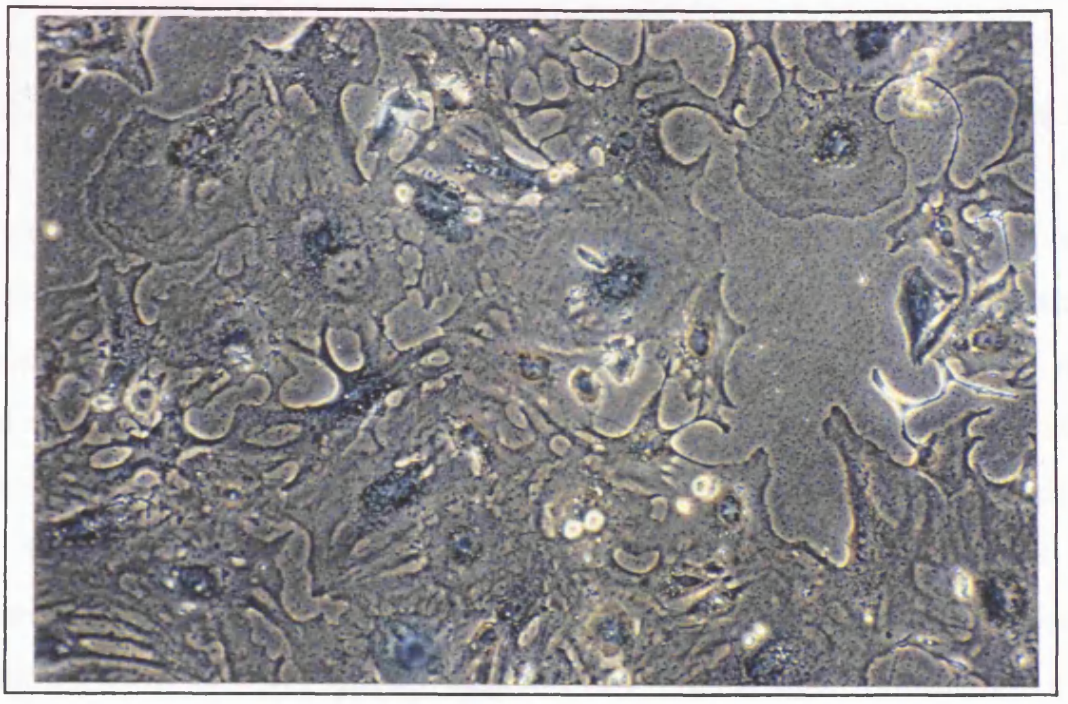
Transfected cells were also harvested in lysis buffer and assayed for total β-gal activity per unit protein. No significant differences between pSRαFLP and pSRαFLPMyc were observed (data not shown). Again no β-galactosidase activity was detected in cells transfected with pSRα plasmid.

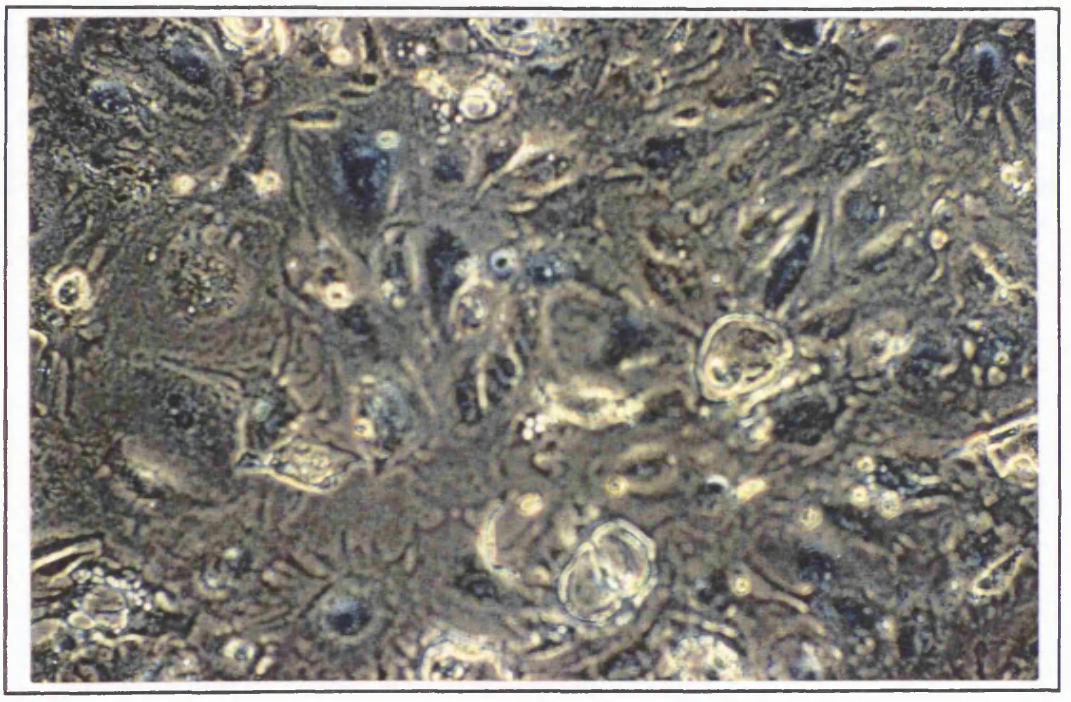
Having determined that the presence of the 9E10 epitope at the C terminus of the FLP protein has no detectable effect on the recombinase activity of FLP it was possible to begin to use the construct to determine the reason for the low efficiency of FLP mediated recombination. Transfection of pSRαFlpMyc into COS cells routinely yeilded detectable levels of Flp-Myc expression after 24,48,72 and 96 hours, with the peak in Flp levels being at 72 hours (data not shown). Similarly, transfection into CV1#5 cells yielded detectable β-gal activity after 24 hours and again the peak of β-galactosidase activity was at 72 hours. No change in the actual number of β-gal positive cells was evident between 24 and 72 hours, suggesting that the increase in total β-galactosidase activity between 24 and 72 hours is due entirely to accumulation of β-galactosidase within activated cells.

Figure 6.5

The presence of a Myc epitope tag at the C-terminus of FLP recombinase has no effect on excisive recombination activity.

A 9cm plate of subconfluent CV1#5 cells (harbouring a single integrated copy of the plasmid pNeob-gal) were transfected with 5µg pSRaFlp (top) and 5µg pSRaFlpMyc by the DEAE dextran method. After 72 hours the cells were fixed in situ by incubation for 5 minutes with 5% gluteraldehyde in PBS and then incubated at 37°c for 2 hours under a solution of b-galactosidase histochemical stain (Materials and Methods). Cells were then visualized directly under phase contrast light microscope. Photographs show random fields of cells and are representative of three experiments.





6.2.2 Flp recombinase is localised to the nucleus of transiently transfected COS cells.

In order to be effective in any manipulation of the genome Flp would be expected to localise to, or at least be present in the nucleus. Entry of proteins to the nucleus is tightly controlled by the nuclear pores. Small proteins of up to 40kDa can diffuse into the nucleus whereas large proteins of greater than 90kDa are excluded unless they contain a nuclear localisation signal. For example, the SV40 T antigen contains the nuclear localisation signal -P-K-K-K-R-K-V- (Richardson, et al., 1986). Nuclear localisation signal sequences are not cleaved, this allows them to move quickly back into the nucleus upon reformation of the nuclear membrane after cell division. There is no consensus nuclear localisation signal, however they have characteristic single or bipartite clusters of basic amino acids (Dingwall and Laskey 1991). There are therefore several candidate NLSs in the primary sequence of Flp recombinase (Figure 6.6) the most likely of which on the basis of basicity is underlined in the figure.

```
  1 MPQFGILCKT PPKVLVRQFV ERFERPSGEK IALCAAELTY LCWMITHNGT
 51 AIKRATFMSY NTIISNSLSF DIVNKSLQFK YKTQKATILE ASLKKLIPAW
101 EFTIIPYYGQ KHQSDITDIV SSLQLQFESS EEADKGNSHS KKMLKALLSE
151 GESIWEITEK ILNSFEYTSR FTKTKTLYQF LFLATFINCG RFSDIKNVDP
201 KSFKLVQNKY LGVIIQCLVT ETKTSVSRHI YFFSARGRID PLVYLDEFLR
251 NSEPVLKRVN RTGNSSSNKQ EYQLLKDNLV RSYNKALKKN APYSIFAIKN
301 GPKSHIGRHL MTSFLSMKGL TELTNVVGNW SDKRASAVAR TTYTHQITAI
351 PDHYFALVSR YYAYDPISKE MIALKDETNP IEEWQHIEQL KGSAEGSIRY
401 PAWNGIISQE VLDYLSSYIN RRI[LEQKLISEEDLD]
```

Figure 6.6
Amino acid sequence of S.cerevisiae Flp recombinase. A clusters of basic amino acids, comprising a potential nuclear localisation signals are indicated by an underscore. The Myc epoitope is shown at the C terminus in brackets.

The localisation of FlpMyc in transiently transfected COS cells was tested by immunofluorescence microscopy (Materials and Methods). Western blots showed some background binding of 9E10 antibody to COS cell lysates, which, although minor, could none the less be potentially detrimental to immunofluorecsence experiments (the background bands are also visible in figure 6.8). It was feared that this background binding may give spurious results in immunofluorescence. Cells were therefore transfected with either FlpMyc or RacMyc (in plasmid pDNA3RacMyc; generous gift of K Siddel; see below) as an epitope control. Rac is a small GTP binding signalling protein. It has previously been shown to localise to membrane ruffles (Tapon and Hall 1997). The fact that transfected Racmyc is shown in figure 6.7 to localise to the membrane ruffles using the 9E10 epitope adds considerable weight to the finding that that Flpmyc appears to be localised extensively to the nucleus (Figure 6.7).

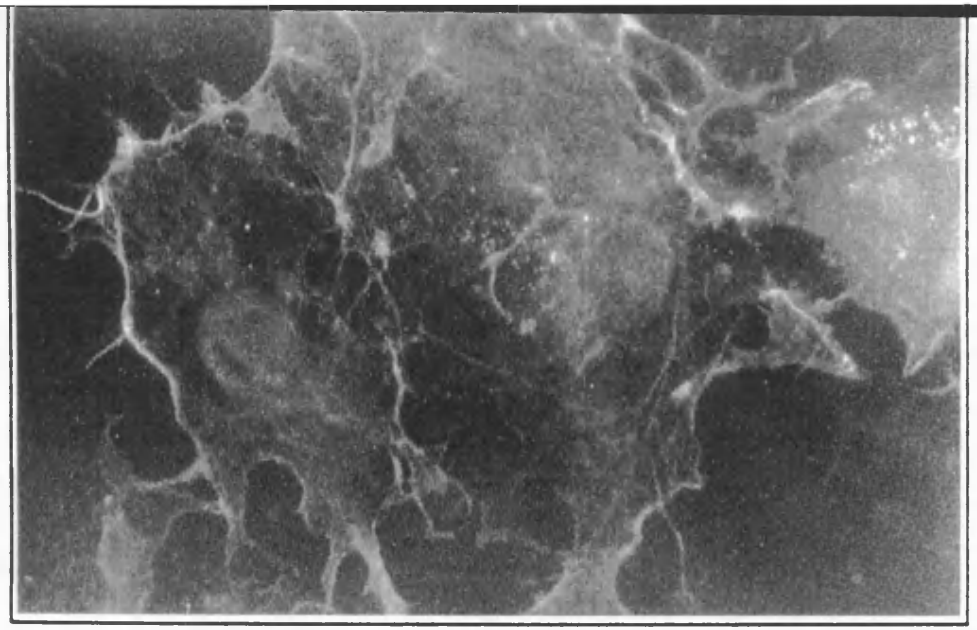
Figure 6.7

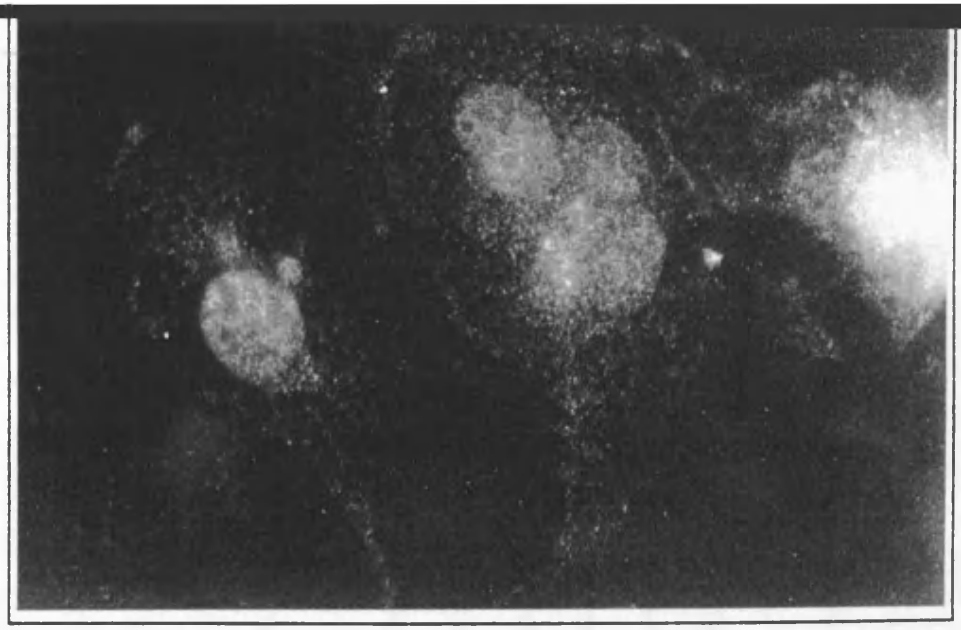
Immunofluorecsent localization of Flp recombinase to nucleus of transiently transfected COS cells.

Subconfluent COS cells, plated on Poly-L-lysine coated cover slips, were transfected with 5µg pSRaFlpMyc or pDNA3RacMyc using the DEAE-dextran method. 72 hours after transfection the cells were stained with 9E10 antibody (1/50), texas red conjugated secondary antibody, and counter stained with FITC conjugated phalloidin (Materials and Methods). Cells were subsequently visualised by fluorescence microscopy and photographed as described (Materials and Methods).

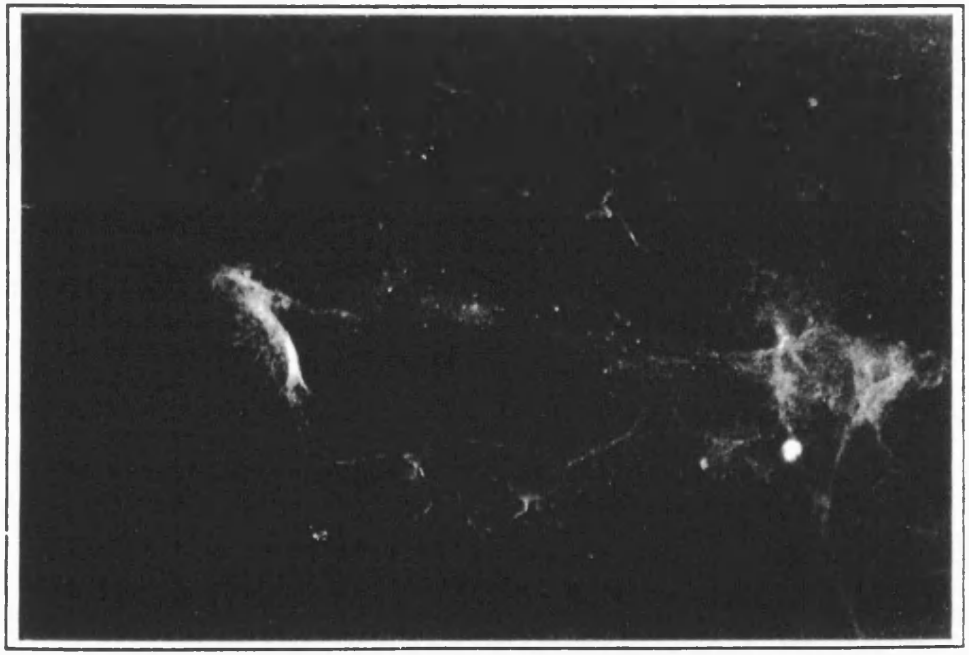
BEST COPY NOTE

THE FOLLOWING PAGES ARE STUCK IN SUCH A MANNER THAT FILMING IS IMPEDED





Rac-9E10/Phalloidin



Rac-9E10/Texas Red

6.2.3 Co-transfection with Flp recombinase markedly reducuces the expresion level of two second reporter genes.

Although Flp protein was readily detectable in COS cells transfected with pSRaFlpMyc plasmid, the detection of the protein on a western blot does not in itself give any indication of the quantity epitope present. A major advantage of epitope tagging over using antibodies raised to the specific protein is that it is possible to compare expression levels directly between different proteins tagged with the same epitope on the same western blot. This is not possible with two different antibodies because they will have differing avidities. If the same antibody to the same epitope is used, even when comparing different proteins, the intensity of the band obtained on the western blot can be assumed to be directly proportional to the quantity of protein present. In order to assess the level of expression of FlpMyc in relation to other transfected genes a control vector containing an N terminal tagged mutant Rac protein was transfected into CV1#5 and COS cells. After 72 hours the cells were lysed as previously described and 100µg total protein western blotted with 9E10 antibody as described previously. Comparison of COS cells transfected with RacMyc appeared to contain far more 9E10 epitope than those co-transfected with FlpMyc. Since the Flp and Rac protein are in different expression plasmids no conclusion can be drawn, or is necessary, concerning the relative efficiencies of the two expression vectors. pcDL-SRα296 has been used previously in our laboratory with CV1 and COS cells to good effect. Results in these studies supported the data of Takebee et al discussed previously, that pcDL-SRα296 plasmid is extremely effective expression vector routinely yielding a high level of expression in transiently transfected cells.

In order to properly assess the efficiency of FlpMyc and RacMyc expression they were co-transfected with each other and each was co-transfected separately with the luciferase expression vector pGL2 (Promega). Two different cell lines were used in the assay (COS and CV1#5).

CV1#5 cells have an added reporter function in β-galactosidase which should be activated in all cells which have expressed active Flp. In two separate co-transfection assays the total β-galactosidase for co-transfections of Flpmyc with Racmyc and Flpmyc with luciferase were approximately equal suggesting that the efficiency of Flp expression is independent of the second reporter gene. This is also evident from western blotting of COS cell lysates, in which the level of FlpMyc expression is the same in co-transfections with both RacMyc and luciferase (figure 6.9). In both CV1#5 and COS cells however, the level co-transfected reporter expression is considerably reduced in the presence of Flpmyc (Figures 6.8 and 6.9) The ratios between expression levels of FlpMyc and the second reporter (luciferase or RacMyc) were constant in three indepedent experiments in both CV1#5 and COS cells although the absolute levels varied due to slight differences in transfection efficiency. Therefore, for each cell transfected (in the case of DEAE-dextran mediated transfection transfected cells take up several thousand copies of plasmid) the negative effect of Flp on second reporter expression (Rac or Luciferase) is constant. This apparent negative effect of Flp expression is difficult to interpret.

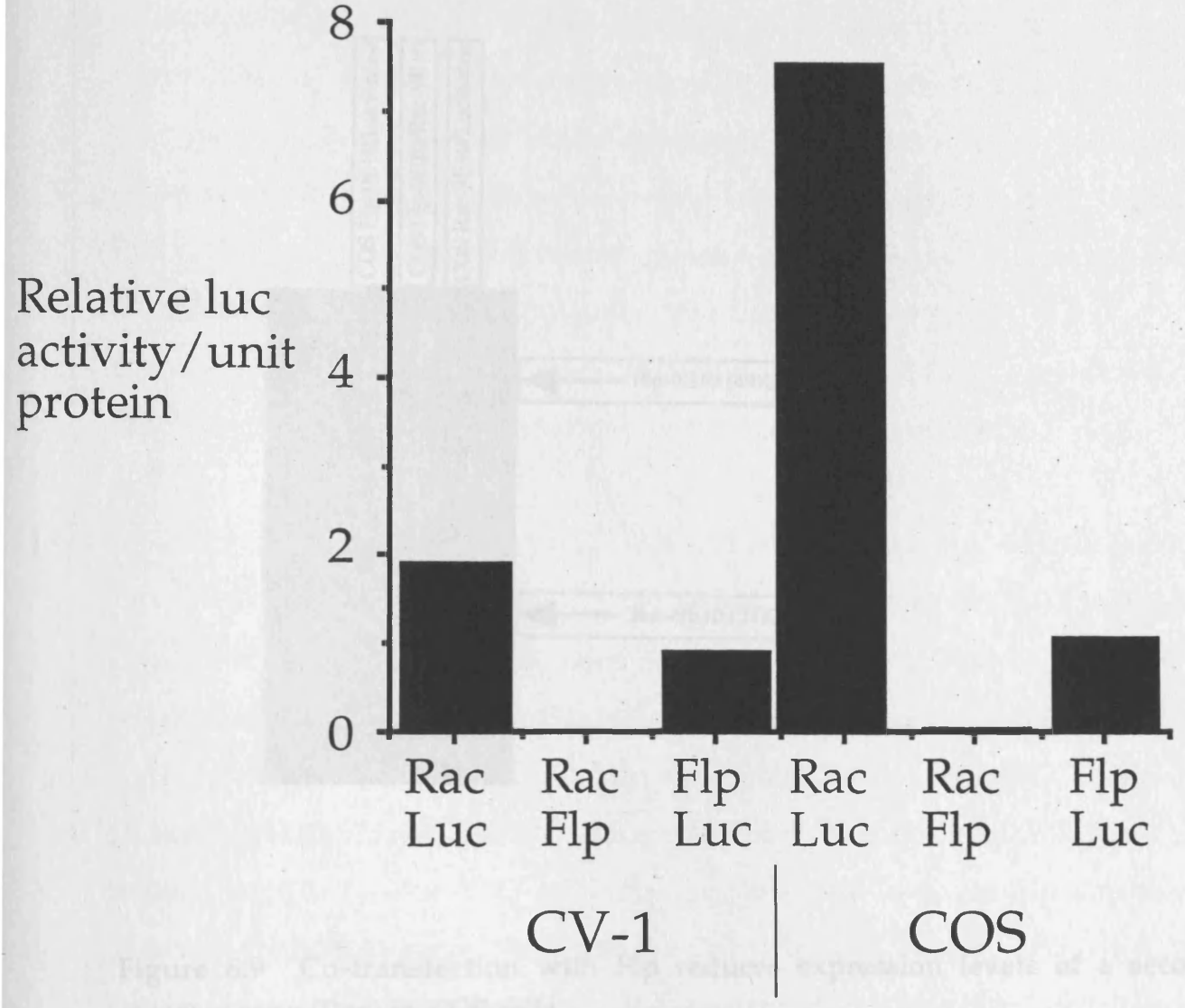


Figure 6.8 Co-transfection with Flp reduces luciferase expression from cotransfected pGL2 plasmid in both COS cells and CV-1 cells. Duplicate plates of cells were transfected with combinations of 5µg each of pSRαFlpMyc, pDNA3RacMyc and pGL2 (luciferase) as described (Materials and Methods). After 72 hours cells were lysed into 200µl lysis buffer. 0.1µg total protein was assayed for luciferse activity using the luminecsence based Promega kit. Assays were performed in triplicate and mean values are shown. The data is representitive of three independent experiments.

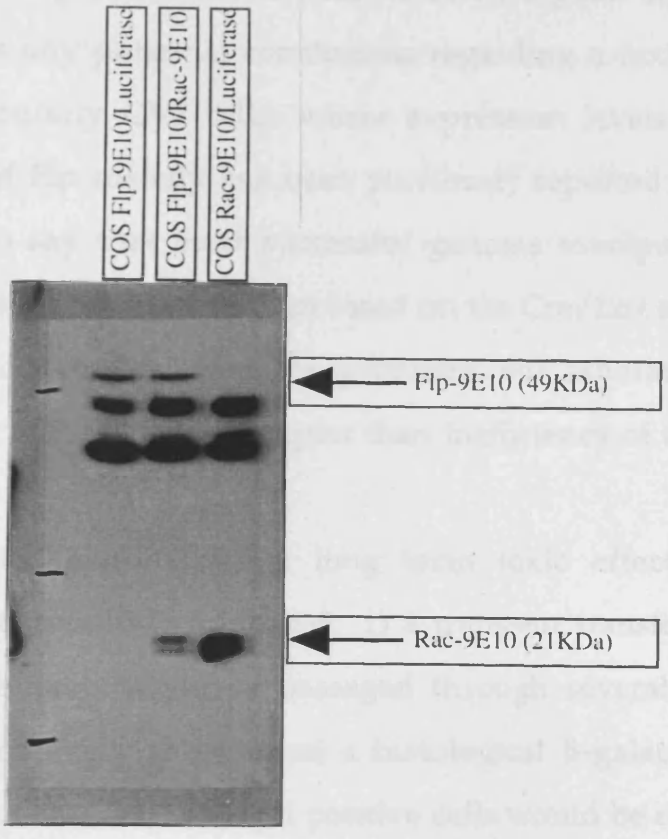


Figure 6.9 Co-transfection with Flp reduces expression levels of a second reporter gene (Rac) in COS cells.
Duplicate plates of cells were transfected with combinations of 5µg each of pSRαFlpMyc, pDNA3RacMyc and pGL2 (luciferase) as described (Materials and Methods). After 72 hours cells were lysed into 200µl lysis buffer.
50µg of total protein was separated by SDS PAGE, transfered to nylon membranes and western blotted with 9E10 antibody (Materials and Methods).

It would be absurd to suggest from this data alone that this observed 'negative effect' of co-transfection with Flp is toxic. The fact that transfected COS cells (which express constructs containing the SV40 enhancer at several orders of magnitude higher than CV1 cells) showed a peak of Flp expression at 3 days contradicts any potential conclusions regarding a toxic effect of Flp in any cell line, particularly CV1 cells, where expression levels are lower. Similarly no evidence of Flp toxicity has been previously reported in the literature, although it is fair to say that most successful genome manipulations using site specific recombinases to date have been based on the Cre/Lox system.
The data do confirm that the problems our laboratory have encountered in expressing FLP are more complex than inefficiency of transfection.

In order to demonstrate a long term toxic effect of Flp expression two experiments would be required: 1) a transient transfection of CV1#5 cells with Flp recombinase would be passaged through several cell cycles. At specified intervals (ca every three days) a histological β-galactosidase assay would be carried out. The ratio of β-gal positive cells would be expected to fall if the initial transient expression of Flp recombinase had any long term toxic effects. 2) a parallel approach would be to stably transfect cells with an Flp expression construct containing a selectable marker. The long-term viability of selected cells (against a suitable control construct not containing Flp) would indicate whether Flp expression was genuinely detrimental to the cell.

Some progress has been made in our laboratory on a variation of the latter of these two experiments Flp coding sequence was fused to Green Fluorescent Protein (discussed above) in the expression vector pGFP-N1 (Clontech)

containing a neomycin selection marker. This construct was stably transfected into CV1 cells. All of the neomycin-positive clones picked showed Flp activity, as shown by transient transfection of the plasmid pFRTNeoβ-gal followed by histochemical staining, but GFP was undetectable in any of the clones by fluorescence microscopy. It should be pointed out that GFP contains no localisation signal and when transfected alone is detectable in both cytoplasm and nuclei. It would be expected therefore from the localisation experiments described above (figure 6.7), that the Flp-GFP chimera would be detected to the nucleus. The absence of GFP in the nucleus would suggest that cells expressing the Flp-GFP fusion at sufficiently high level for GFP to be detectable were not surviving the cloning process due to a toxic effect of the high concentration of Flp.

6.3 Discussion

The data presented in this chapter has made a small but positive contribution towards defining the reasons for the diffulties experienced in our laboratory in using Flp recombinase in gene manipulation. The generation and characterisation of the 9E10 epitope tagged Flp will allow future workers to further investigate the problems associated with the efficient use of Flp in J558L derived cell lines. The immunolocalisation data showing that Flp can enter the mammalian nucleus eliminates the need in this or other laboratories to place an artificial nuclear localisation signal into the primary sequence of Flp. Although it clearly needs to be refined, the co-transfection data clearly show that the Flp expression construct currently used in our laboratory has a detrimental effect on the fate of co-transfected plasmids. Since the entire strategy for the Flp based utilisation of the C6 locus depends on cotransfection of Flp and constructs

containing FRT sites for homologous recombination, this problem needs to be further resolved. However, a more pressing problem has also been identified.

After the C6/FRT cell line was constructed, a major problem with the design of the FRT construct was realised. This was that even if sufficient Flp activity was available to facilitate integrative recombination between plasmids harboring an FRT site and the C6 locus, after such a recombination, in an environment of such high Flp activity it seemed extremely likely that a further recombination step between the FRT sites would lead inevitably to excisive recombination (figure 6.10). Given that after integration the FRT sites would be a relatively short distance apart (dictated by the size of the plasmid to be integrated), and the free plasmid would be free to diffuse away, the equilibrium between integration and excision would be heavily weighted towards excision.

This problem has been addressed by generating point mutations in the 8bp core region of the FRT (Schlake and Bode, 1994) Previous DNaseI footprinting studies defined the Flp binding site as a 12bp tract including 11bp of the inverted repeats and only the two outermost base pairs of the 8bp core. Absolute homology between the 8bp core is essential for DNA alignment during recombination, but as it is not involved in FLP protein binding *per se*, could be mutated freely without effect on FLP binding. Thus mutations in the core sites were generated such that they were able to recombine with identical mutant FRT sites but not with different mutants or the wild type FRT. Schlake et al also report such a mutant site and this allows the opportunity of producing constructs containing two FRT sites, one wild-type and one mutant, flanking the DNA sequence to be integrated. Two independent recombination events then occur by a single enzymic activity; the recombination between the two wild type FRT sites and the

recombination between the two mutant FRT sites. This reaction leads to one of two products in theoretically equal quantities depending on the direction taken in this double reciprocal crossover (Figure 6.11). In both cases the entire plasmid integrates at one side of the locus as shown in figure. Subsequently, a second recombination event excises one of the FRT pairs.

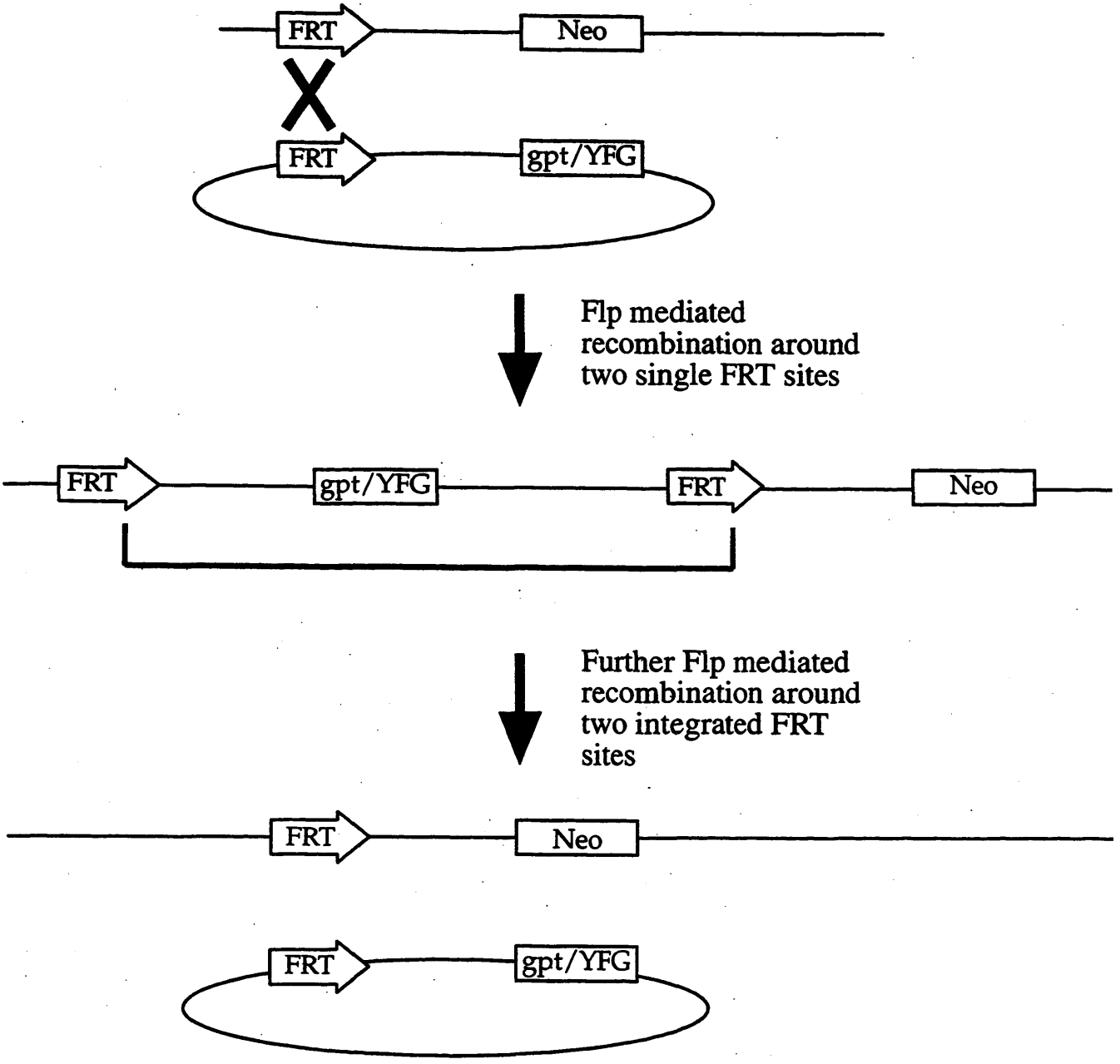
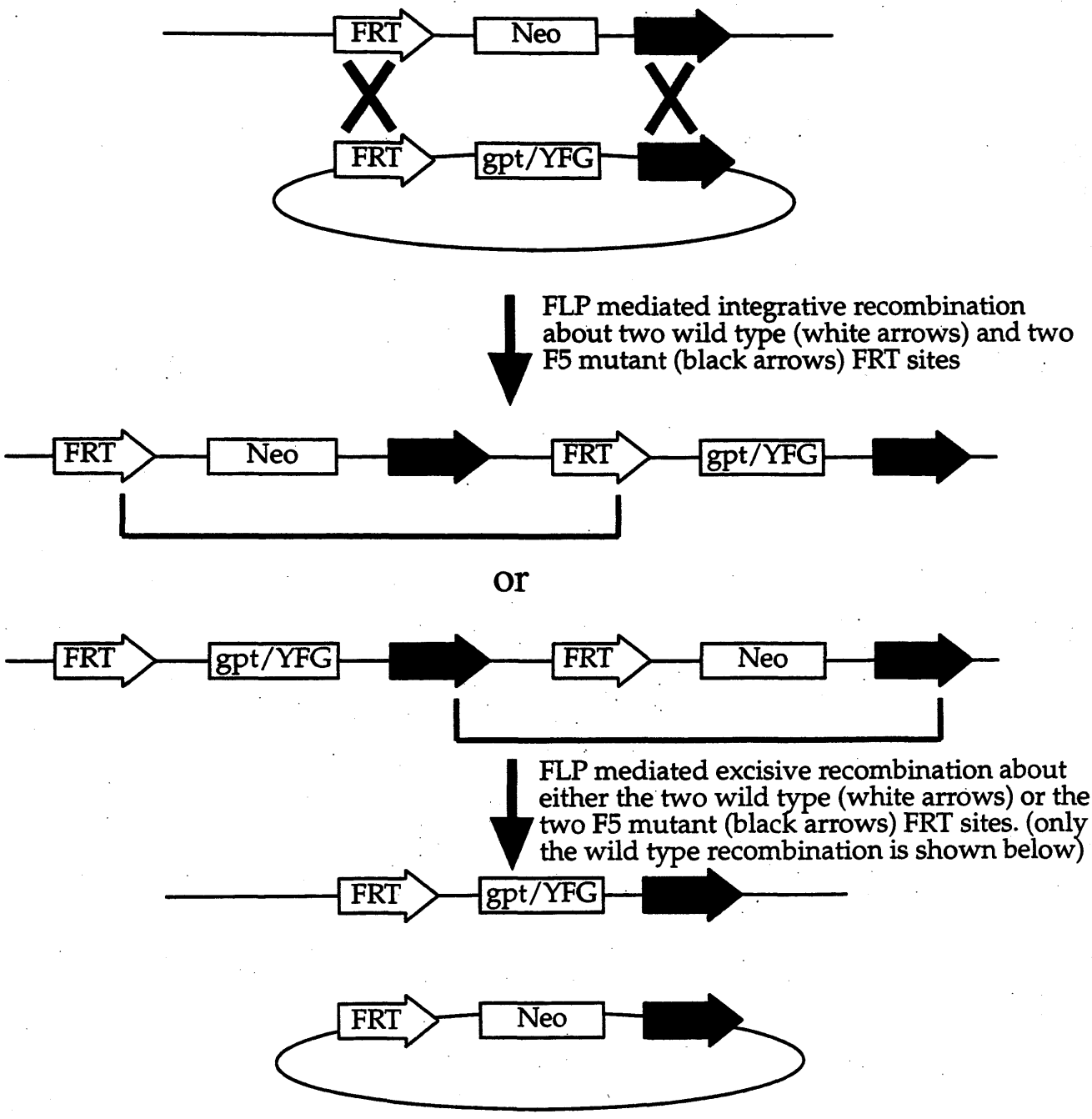


Figure 6.10. Schematic diagram showing the likely outcome of FLP mediated recombination in the present C6FRT locus.

As required recombination would occur between the FRT sites in the C6 locus and the plasmid containing gpt selectable marker and YFG (your favorite gene) placing both gpt and YFG at the C6 locus. However the integrated plasmid would be immediately excised by residual Flp activity to regenerate the starting products. Since after the initial integrative recombination the concentration of plasmid would be fairly low, the equilibrium of these recombination reactions will be weighted heavily towards excisive recombination.

Figure 6.11 Schematic diagram showing the idealised outcome of FLP mediated recombination at a modified C6FRT locus and a plasmid containing one wild type and one F5 mutant FRT site. The first round of recombination (integrative) leads to one of two species depending only on which side of the locus the plasmid integrates. A subsequent recombination reaction between either the two wild type or the two mutant FRT sites (each reaction would have equal probability) leaves one mutant and one wild type FRT site. The excised should diffuse away and recombination will be terminated. 50% of successfully transfected cells will contain gpt and YFG at the active locus.


FRT
Neo
FRT
gpt/YFG
FLP mediated integrative recombination about two wild type (white arrows) and two F5 mutant (black arrows) FRT sites
FRT
Neo
FRT
gpt/YFG
or
FRT
gpt/YFG
FRT
Neo
FLP mediated excisive recombination about either the two wild type (white arrows) or the two F5 mutant (black arrows) FRT sites. (only the wild type recombination is shown below)
FRT
gpt/YFG
FRT
Neo

SPECIAL NOTE

ITEM SCANNED AS SUPPLIED
PAGINATION IS AS SEEN

6.4 Future Work

A substantial effort will be required in order to replace the single FRT site at the C6 locus with a construct containing the two FRT sites. However far more is now known about the C6 locus than when the original C6/FRT cell line was generated. 9.4kb of flanking sequence has now been cloned which, in addition to the plasmid sequence, constitutes around 15kb of homologous sequence. A construct containing this sequence should have a dramatically increased efficiency of homologous recombination over the previous construct used which was based on plasmid sequence only. The presence of the B1 element in the construct, known to favour illegitimate recombination, ought to be of little consequence since non-homologous integrants can be eliminated extremely efficiently by selecting colonies in media containing thioxanthine. The main factor known to favour homolous recombination is the length and fidelity of the homologous DNA sequence (L Hemmings; personal communication).

The relative inefficiency and potential toxic effect of the FLP enzyme on CV1 and COS cells is still largely unresolved. The presence of any β-gal positive CV1#5 cells at all in the *in situ* β-gal assays (figure 6.6) would suggest that in certain cases, or more likely at lower doses, Flp recombinase has no lasting negative effect on the the cells. The experiment described above to determine the long term effect of Flp on the cell would be required to determine whether this is true for longer time periods than the three days in which transfected cells were routinely incubated after transfection. A logical next step in increasing the overall effectiveness of Flp, now being carried out in the laboratory, is to determine the effect of less pernicious transfection techniques such that only just sufficient Flp coding sequence is delivered to each cell. Electroporation, for

example delivers considerably less DNA to the cell (around 10 copies per transfected cell) whereas DEAE dextran delivers around 1000 copies per transfected cell. Despite the problems the laboratory is encountering in the use Flp to utilise the C6 locus, the concept of a system whereby proteins can be expressed at high levels in myeloma cells remains theoretically sound and is still being pursued by other workers.

Chapter 7
General Discussion

Future work

The data presented in chapter 5 of this thesis suggest that there is no DNase 1 hypersensitive site or matrix attachment region within the regions of DNA cloned. The absence of a hypersentive site would suggest that the cloned regions do not contain a control element such as an LCR or enhancer. A more direct assay of the effect of the cloned regions on gene expression is presently underway. An appropriate expression vector, containing the a luciferase reporter gene driven by the SV40 late promoter gpt selectable marker is being constructed. Cloned sequences will be cloned into the vector and their effect on transfection frequency and luciferase expression will be assayed. Initially, the IgH enhancer is not being included in the vector to minimise background expression in control transfections. This assay will provide a more definitive answer to whether the cloned sequences contain elements involved in transcription activation than the DNase 1 sensitivity assays or the MAR assays. It should be pointed out, however that since both of these assays proved negative (and since the FASTA searches on the sequence data showed little of interest in terms of expressed genes or known control elements), it seems unlikely that the cloned regions contain an element capable of facilitating high level trancription. It is more likely that in both cell lines, the element or elements of interest will be a considerable distance from the integration sites.

With this in mind, the main focus of work in the laboratory has shifted towards the use of FLP recombinase to 'flip' expression cassettes into the active locus of the C6 cell line. Such an approach requires no knowledge of the cause of the high level of transcription in the C6 and D8 cell line and such a system will

therefore be largely of biotechnological value and will give little insight into the reason for the high level of transcription in the two cell lines.

The drive for a system in which recombinant protein can be expressed in large scale tissue culture has now, to a certain extent, been superceded by transgenic animal technology. The commercial production of proteins in whole animals was limited by factors such as the position effect, position effect variagation and by the fact that a high proportion of animal expressing the gene were lost in breeding. The advent of technology whereby higher animals can be cloned now means that entire flocks of Dollys (sheep) and Dorises (cows) will be available for the production of proteins of commercial and medical value. In both the sheep and the cow the protein can be placed under the control of the β-lactoglobulin control elements and the protein secreted into the milk. Therefore, although the initial investment for anyone wishing to use this method will be initially very high, the long term large scale production of proteins by this method will be little more than the normal cost of farming a herd of sheep or cows.

Cell culture based systems will still have a use for the small scale production of proteins and for the production of proteins which are toxic or deleterious to the whole animal. The development of a mammalian system based on the J558L cell line using either Flp recombinase mediated insertion into the C6 locus or using cloned sequences from the D8 or C6 locus (if they are shown to effect high-level, position independent expression) remains a worthwhile and viable goal, for both commercial and academic reasons. This is still being pursued by other workers in the laboratory.

Chapter 8
References

Abe, M., Oshima, R. (1990) *Journal of Cell Biology. 111,* 1197-1206.

Aronow, B., Silbiger, R., Dusing, M., Stock, J., Lida, J., Pretorius, I.-M., Ayyub, H. (1992) Functional analysis of the human adenosine deaminase gene thymic regulatory region and its ability to generate position-independent transgene expression. *Molecular and Cellular Biology. 12,* 4170-4185.

Assendelft, G., O. Hanscombe, F. Grosveld and D. R. Greaves. (1989) The β–Globin Dominant Control Region Activates Homologous and Heterologous Promoters in a Tissue-Specific Manner. *Cell. 56,* 969-977.

Belgrader, P., R. Dey and R. Berezney. (1991) Molecular-Cloning Of Matrin-3 - a 125-Kilodalton Protein Of the Nuclear Matrix Contains an Extensive Acidic Domain. *Journal Of Biological Chemistry. 266,* 9893-9899.

Benezra, R., C. R. Cantor and R. Axel. (1986) Nucleosomes Are Phased Along the Mouse Beta-Major Globin Gene In Erythroid and Nonerythroid Cells. *Cell. 44,* 697-704.

Berek, C. and C. Milstein. (1987) Mutation Drift and Repertoire Shift In the Maturation Of the Immune- Response. *Immunological Reviews. 96,* 23-41.

Berrios, M., N. Osheroff and P. A. Fisher. (1985) Insitu Localization Of Dna Topoisomerase-Ii, a Major Polypeptide Component Of the Drosophila Nuclear

Matrix-Fraction. *Proceedings Of the National Academy Of Sciences Of the United States Of America. 82,* 4142-4146.

Betz, A. G., C. Milstein, A. Gonzalezfernandez, R. Pannell, T. Larson and M. S. Neuberger. (1994) Elements Regulating Somatic Hypermutation Of an Immunoglobulin-Kappa Gene - Critical Role For the Intron Enhancer Matrix Attachment Region. *Cell. 77,* 239-248.

Bonifer, C., M. Vidal, F. Grosveld and A. E. Sippel. (1990) Tissue specific and position independent expression of the complete domain for chicken lysozyme in transgenic mice. *Embo Journal. 9,* 2843-2848.

Bonifer, C., M. Vidal, F. Grosveld and A. E. Sippel. (1990) Tissue specific and position independent expression of the complete gene domain for chicken lysosyme in transgenic mice. *EMBO. 9,* 2848.

Bonifer, C., N. Yannoutsos, G. Kruger, F. Grosveld and A. E. Sippel. (1994) Dissection Of the Locus-Control Function Located On the Chicken Lysozyme Gene Domain In Transgenic Mice. *Nucleic Acids Research. 22,* 4202-4210.

Brake, A., Merryweather, J., Loit, D., Heberlei,U., Masriarz, F., Mullenbach, G., Urdea, M., Valenzuela, P., Barr, P. (1984) α-Factor directed synthesis and secretion of mature foriegn proteins in *Saccharomcyes cerevisiae. Proceedings of the National Academy of Sciences USA. 81,* 4642-4646.

Chamberlain, J. W., H. A. Vasavada, S. Ganguly and S. Weissman. (1991) Identification of cis Sequences Controlling Efficient Position-Independent Tissue-Specific Expression of Human Major Histocompatibility Complex Class 1 Genes in transgenic Mice. *Molecular and Cellular Biology. 11,* 3564-3572.

Chamberlain, J. W., H. A. Vasavada, S. Ganguly and S. M. Weissman. (1991) Identification of cis sequences controlling efficient position-independent tissue specific expression of the human major histocompatibility complex class 1 genes in transgenic mice. *Mol. Cell. Biol. 11,* 3564-3572.

Chang, N., Matteucci, M., Perry, J., Wulf,J., Chen, C., Hitzeman, R. (1986) Saccharomyces cerevisiae secretes and correctly processes human interferon hybrid proteins containing yeast invertase signal peptides. *Molecular and Cellular Biology. 6,* 1812-1819.

Cheyette, T. E., T. Ip, S. Faber, Y. Matsui and R. Chalkry. (1992) Characterisation of the factors binding to a PEPCK gene upstream hypersensitive site with LCR activity. *Nucleic acids research. 20,* 3427-3433.

Clarke, L., Carbon, J. (1976) A colony bank containing synthetic Col EI hybrid plasmids representative of the entire E. coli genome. *Cell. 9,* 91-97.

Cockerill, P., M. Yuen and W. Garrard. (1987) The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. *Journal of Biological Chemistry. 262,* 5394-5397.

Cockerill, P. N. (1990) Nuclear Matrix Attachment Occurs In Several Regions Of the Igh Locus. *Nucleic Acids Research. 18,* 2643-2648.

Cockerill, P. N. and W. T. Garrard. (1986) Chromosomal loop anchorage of the Kappa immunoglobulin gene occurs next to the enhancer in a region containing Topoisomarase II sites. *Cell. 44,* 273-282.

Dale, T., Krnacik, M., Schmidhauser, C., Yang, C., Bissell, M., Rosen, J. (1992) High level expression of the rat whey acidic protein gene is mediated by elements in the promoter and 3' untranslated region. *Molecular and Cellular Biology. 12,* 905-914.

Dickinson, L. A., T. Joh, Y. Kohwi and T. Kohwishigematsu. (1992) A Tissue-Specific Mar/Sar Dna-Binding Protein With Unusual Binding- Site Recognition. *Cell. 70,* 631-645.

Dickinson, L. A. and T. Kohwishigematsu. (1995) Nucleolin Is a Matrix Attachment Region Dna-Binding Protein That Specifically Recognizes a Region With High Base-Unpairing Potential. *Molecular and Cellular Biology. 15,* 456-465.

Dingwall, C., and R. A. Laskey (1991) Nuclear targetting sequences- A consenus. *Trends in Biochemical Sciences 16 Vol.12* 478-481

Dolan, M., B. J. Sugarman, J. B. Dodgson and J. D. Engel. (1981) Chromosomal Arrangement Of the Chicken Beta-Type Globin Genes. *Cell. 24,* 669-677.

Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker and J. S. Mattick. (1991) "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research. 19,* 4008.

Dworetzky, S. I., K. L. Wright, E. G. Fey, S. Penman, J. B. Lian, J. L. Stein and G. S. Stein. (1992) Sequence-Specific Dna-Binding Proteins Are Components Of a Nuclear Matrix-Attachment Site. *Proceedings Of the National Academy Of Sciences Of the United States Of America. 89,* 4178-4182.

Dymecki, S. M. (1996) A Modular Set Of Flp, Frt and Lacz Fusion Vectors For Manipulating Genes By Site-Specific Recombination. *Gene. 171,* 197-201.

Enquist, L. and A. Skalka. (1973) Replication of bacteriophage l DNA dependent on the function of host and viral genes. *Journal of Molecular Biology. 75,* 185.

Evan, G., I. G.K. Lewis, G. Ramsay, J.M. Bishop. (1985) Isolation of monoclonal antibodies specific for human c-myc protooncogene product. *Molecular and Cellular Biology 5* 3610-3616

Fulton, R. and B. Van Ness. (1993) Luminescent reporter gene assays for luciferase and b-galactosidase using a liquid scintilation counter. *BioTechniques. 14,* 762-763.

Gasser, S. and U. Laemmli. (1987) A glimpe at chromosomal order. *Trends in Genetics. 3,* 16-22.

Gasser, S. M. and U. K. Laemmli. (1986) The Organization Of Chromatin Loops - Characterization Of a Scaffold Attachment Site. *Embo Journal. 5,* 511-518.

Gething, M.-J. and J. Sambrook. (1981) Cell surface expression of influenza haemagglutinin from a cloned DNA copy of the RNA gene. *Nature. 293,* 620-5.

Glassford, S. (1993) Leicester University Thesis.

Glutzman, Y. (1981) SV40 -transformed simian cells support the replication of early SV40 mutants. *Cell. 23,* 175-82.

Greaves, D., F. Wilson, D. Lang and D. Kioussis. (1989) Human CD2 3'-Flanking Seqences Confer High Level, T Cell-Specific, Position Independent Gene Expression in Transgenic Mice. *Cell. 56,* 979-986.

Griffith, J., M. Bleyman, C. A. Rauch, P. A. Kitchin and P. T. Englund. (1986) Visualization Of the Bent Helix In Kinetoplast Dna By Electron- Microscopy. *Cell. 46,* 717-724.

Gross, D. S. and W. T. Garrard. (1988) Nuclease Hypersensitive Sites In Chromatin. *Annual Review Of Biochemistry. 57,* 159-197.

Grosschedl, R., K. Giese and J. Pagel. (1994) Hmg Domain Proteins - Architectural Elements In the Assembly Of Nucleoprotein Structures. *Trends In Genetics. 10,* 94-100.

Grosschedl, R. and S. Morf. (1981) Functional Dissection Of an H2a Histone Gene Promoter. *Experientia. 37,* 646-647.

Grosveld, F., G. Assendelft, D. R. Greaves and G. Kollias. (1987) Position-Independent, High-Level Expression of the Human B-Globin Gene in Transgenic Mice. *Cell. 51,* 975-985.

Gruss, P. and G. Khoury. (1981) Expression of simian virus 40-rat preproinsulin recombinants in monkey kidney cells: use in preproinsulin RNA processing signals. *Proceding of the National Academy of sciences USA. 78,* 133-7.

Hambor, J. E., J. Mennone, M. E. Coon, J. H. Hanke and P. Kavathas. (1993) Identification and Characterization Of an Alu-Containing, T-Cell- Specific Enhancer Located In the Last Intron Of the Human Cd8-Alpha Gene. *Molecular and Cellular Biology. 13,* 7056-7070.

Harrison. (1993) Construction of a Generally Usefull Mammalian Cell Host for High Level Production of Heterologous Proteins. *Grant Proposal.*

Harrison, T. M., K. Hudson, S. Munson, S. Uff and S. Glassford. (1994) Derivation and partial analysis of two highly active myeloma cell transfectants. *Biochem. et Biophys. Acta 1216* 147-156

Higgs, D. R., W. G. Wood, A. P. Jarman, J. Sharpe, J. Lida, I.-M. Pretorius and H. Ayyub. (1990) A major positive regulatory region located far upstream of the human α-globin gene locus. *Genes and Development. 4,* 1588-1601.

Higgs, D. R., W. G. Wood, A. P. Jarman, J. Sharpe, J. Lida, I. M. Pretorius and H. Ayyub. (1990) A major postive regulatory region located far upstream of the a-globin gene locus. *Genes and Development. 4,* 1588-1601.

Hudson, K. (1988) Leicester University Thesis 1988.

Jarman, A. P. and D. R. Higgs. (1988) Nuclear Scaffold Attachment Sites In the Human Globin Gene Complexes. *Embo Journal. 7,* 3337-3344.

Jerpseth, D., A. Greener, J. M. Short, J. Viola and P. L. Kretz. (1992) *Strategies (published and distributed by Strategene, Cambridge Science Park, Cambridge CB4 4GF). Issue 5* 81-83

Kalos, M. and R. E. K. Fournier. (1995) Position-Independent Transgene Expression Mediated By Boundary Elements From the Apolipoprotein-B Chromatin Domain. *Molecular and Cellular Biology. 15,* 198-207.

Karn, J., H. Matthes, M. Gait and S. Brenner. (1984) A new selective phage cloning vector with sites for Xba I, Bam HI, Hind III, Eco RI, Sst I, and Xho I. *Gene. 32,* 217-224.

Klehr, D., K. Maass and J. Bode. (1991) Scaffold-Attached Regions From the Human Interferon-Beta Domain Can Be Used to Enhance the Stable Expression Of Genes Under the Control Of Various Promoters. *Biochemistry. 30,* 1264-1270.

Klug, A., L.C. Lutter and D.Rhodes. (1982) Helical periodicity of DNA on and off the nucleosome as probed by nucleases. *Cold Spring Harbour symposia on quantitative biology. 47* 285-292

Lang, G., C. Mamalaki, D. Greenberg, N. Yannoutsos and D. Kioussis. (1991) Deletion Analysis Of the Human Cd2 Gene Locus-Control Region In Transgenic Mice. *Nucleic Acids Research. 19,* 5851-5856.

Lee, D. Y., J. J. Hayes, D. Pruss and A. P. Wolffe. (1993) A Positive Role For Histone Acetylation In Transcription Factor Access to Nucleosomal Dna. *Cell. 72,* 73-84.

Liang, S.-M., Allet, B., Rose, K., Hirschi, M., Liang, C-M., Thatcher, D. (1985) *Biochem J. 229,* 429-439.

Lorch, Y., J. W. Lapointe and R. D. Kornberg. (1987) Nucleosomes Inhibit the Initiation Of Transcription But Allow Chain Elongation With the Displacement Of Histones. *Cell. 49,* 203-210.

Losa, R. and D. D. Brown. (1987) A Bacteriophage RNA-Polymerase Transcribes Invitro Through a Nucleosome Core Without Displacing It. *Cell. 50,* 801-808.

Lyznik, L. A., J. C. Mitchell, L. Hirayama and T. K. Hodges. (1993) Activity Of Yeast Flp Recombinase In Maize and Rice Protoplasts. *Nucleic Acids Research. 21,* 969-975.

Miller, F. D., G. H. Dixon, J. B. Rattner and J. H. Vandesande. (1985) Assembly and Characterization Of Nucleosomal Cores On B-Form Vs Z- Form DNA. *Biochemistry. 24,* 102-109.

Mirkovitch, J., M. E. Mirault and U. K. Laemmli. (1984) Organization Of the Higher-Order Chromatin Loop - Specific DNA Attachment Sites On Nuclear Scaffold. *Cell. 39,* 223-232.

Miyamoto, C., G. Smith, J. Farrell-Towt, R. Chizzonite, M. Summers and G. Ju. (1985) Production of human c-myc protein in insect cells infected with bacculaovirus expression vector. *Mol. Cell. Biol.. 5,* 2860-5.

Miynarova, L., A. Loonen, J. Heldens, R. C. Jansen, P. Keizer, W. J. Stiekema and J. P. Nap. (1994) Reduced Position Effect In Mature Transgenic Plants Conferred By the Chicken Lysozyme Matrix-Associated Region. *Plant Cell. 6,* 417-426.

Moreau, J., L. Marcaud, F. Maschat, J. Kejzlarovalepesant, J. A. Lepesant and K. Scherrer. (1982) A&T-Rich Linkers Define Functional Domains In Eukaryotic Dna. *Nature. 295,* 260-262.

Moreau, J., L. Matyashsmirniaguina and K. Scherrer. (1981) Systematic Punctuation Of Eukaryotic DNA By A&T-Rich Sequences. *Proceedings Of the National Academy Of Sciences Of the United States Of America-Biological Sciences. 78,* 1341-1345.

Mulligan, R. and P. Berg. (1981) Selection for animal cells that express the E. coli gene coding for coding for xanthine-guanine phosphoribosyl transferase. *Procedings of the National academy of sciences. USA. 78,* 2072-6.

Nabirochkin, S., M. Ossokina and T. Heidmann. (1998) A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *Journal Of Biological Chemistry. 273,* 2473-2479.

Naylor, L. H. and E. M. Clark. (1990) D(Tg)N.D(Ca)N Sequences Upstream Of the Rat Prolactin Gene Form Z-DNA and Inhibit Gene-Transcription. *Nucleic Acids Research. 18,* 1595-1601.

Neuberger, M. (1985) Expression and regulation of immunoglobulin heavy chain gene transfected into lymphoid cells. *EMBO Journal. 2,* 1373-1378.

Ogorman, S., D. T. Fox and G. M. Wahl. (1991) Recombinase-Mediated Gene Activation and Site-Specific Integration In Mammalian-Cells. *Science. 251,* 1351-1355.

Old, R. W. and S. B. Primrose. 1989. *Principals of gene manipulation.* 3rd Ed

Orkin, S. H. (1990) Globin Gene Regulation and Switching: Circa 1990. *Cell. 63,* 665-672.

Owenhughes, T. and J. L. Workman. (1996) Remodeling the Chromatin Structure Of a Nucleosome Array By Transcription Factor-Targeted Trans-Displacement Of Histones. *Embo Journal. 15,* 4702-4712.

Palmiter, R., Sandgren, E., Koeller, D., Brinster, R. (1993) *Molecular and Cellular Biology. 13,* 5266-5275.

Pan, G. H., K. Luetke and P. D. Sadowski. (1993) Mechanism Of Cleavage and Ligation By Flp Recombinase - Classification Of Mutations In Flp Protein By Invitro Complementation Analysis. *Molecular and Cellular Biology. 13,* 3167-3175.

Pan, G. H. and P. D. Sadowski. (1992) Ligation Activity Of Flp Recombinase - the Strand Ligation Activity Of a Site-Specific Recombinase Using an Activated DNA Substrate. *Journal Of Biological Chemistry. 267,* 12397-12399.

Paranjape, S., Kamakaka, R., Kadonaga, J. (1994) Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annual Review of Biochemistry. 63,* 265-297.

Paulson, J. and U. Laemmli. (1977) The structure of histone depleted metaphase chromosomes. *Cell. 12,* 817-828.

Pearson, W. R. and D. J. Lipman. (1988) Improved Tools For Biological Sequence Comparison. *Proceedings Of the National Academy Of Sciences Of the United States Of America. 85,* 2444-2448.

Perry, M. and R. Chalkley. (1981) The Effect Of Histone Hyperacetylation On the Nuclease Sensitivity and the Solubility Of Chromatin. *Journal Of Biological Chemistry. 256,* 3313-3318.

Pfeifer, G. P., S. D. Steigerwald, P. R. Mueller, B. Wold and A. D. Riggs. (1989) Genomic Sequencing and Methylation Analysis By Ligation Mediated Pcr. *Science. 246,* 810-813.

Pham, C. T. N., D. M. Macivor, B. A. Hug, J. W. Heusel and T. J. Ley. (1996) Long-Range Disruption Of Gene-Expression By a Selectable Marker Cassette. *Proceedings Of the National Academy Of Sciences Of the United States Of America. 93,* 13090-13095.

Phi-Van, L., J. P. Kries, W. Ostertag and W. H. Stratling. (1990) The chicken lysozyme 5′ matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. *Mol.Cell.Biol. 10,* 2302-2307.

Phi-van, L., J. P. vonKries, W. Ostertag and W. H. Stratling. (1990) The chicken lysozyme 5′ matrix attachment region increases transcription from a heterologous promoter in hetewrologous cells and dampens position effects on the expression of transfected genes. *Mol. Cell Biol. 10,* 2302-2307.

Phivan, L. and W. H. Stratling. (1988) The Matrix Attachment Regions Of the Chicken Lysozyme Gene Co-Map With the Boundaries Of the Chromatin Domain. *Embo Journal. 7,* 655-664.

Phivan, L. and W. H. Stratling. (1996) Dissection Of the Ability Of the Chicken Lysozyme Gene 5′-Matrix Attachment Region to Stimulate Transgene Expression and to Dampen Position Effects. *Biochemistry. 35,* 10735-10742.

Phivan, L., J. P. Vonkries, W. Ostertag and W. H. Stratling. (1990) The Chicken Lysozyme-5′ Matrix Attachment Region Increases Transcription From a Heterologous Promoter In Heterologous Cells and Dampens Position Effects On the Expression Of Transfected Genes. *Molecular and Cellular Biology. 10,* 2302-2307.

Ptashne, M. (1986) Gene-Regulation By Proteins Acting Nearby and At a Distance. *Nature. 322,* 697-701.

Ptashne, M. (1988) How eukaryotic transcriptional activators work. *Nature. 335,* 683-689.

Reitman, M., E. Lee, H. Westphal and G. Felsenfeld. (1990) Site independent expression of the chicken bA-globin gene in transgenic mice. *Nature. 348,* 749-752.

Reyes, J. C., C. Muchardt and M. Yaniv. (1997) Components of the human SWI/SNF complex are enriched in active chromatin and are associated with the nuclear matrix. *Journal Of Cell Biology. 137,* 263-274.

Richardson, W., B. Roberts and A. Smith. (1986) Nuclear localisation signals in polyoma virus large T. *Cell. 44,*

Sadowski, P. D. (1995) The Flp recombinase of the 2-μm plasmid of Saccharomyces cerevisae. *51,* 53-93.

Saitoh, N., I. G. Goldberg, E. R. Wood and W. C. Earnshaw. (1994) Scii - an Abundant Chromosome Scaffold Protein Is a Member Of a Family Of Putative Atpases With an Unusual Predicted Tertiary Structure. *Journal Of Cell Biology. 127*, 303-318.

Sambrook, J., E. F. Fritsch and T. Maniatis. (1989) Molecular Cloning, A Laboratory Manual. *Second Edition. Cold Spring Harbor Laboratory Press.*,

Sander, M. and T. S. Hsieh. (1985) Drosophila Topoisomerase-Ii Double-Strand Dna Cleavage - Analysis Of DNA-Sequence Homology At the Cleavage Site. *Nucleic Acids Research. 13*, 1057-1072.

Schlake, T. and J. Bode. (1994) Use Of Mutated Flp Recognition Target (Frt) Sites For the Exchange Of Expression Cassettes At Defined Chromosomal Loci. *Biochemistry. 33*, 12746-12751.

Schmid, C. W. and W. R. Jelinek. (1982) The Alu Family Of Dispersed Repetitive Sequences. *Science. 216*, 1065-1070.

Selker, E. U. (1990) DNA methylation and chromatin structure: a view from below. *TIBS. 15*, 103-107.

Siebelist, U., L. Hennighausen, J. Battey and P. Leder. (1984) Chromatin Structure and Protein Binding in the Putative Regulatory Region of the c-myc Gene in Burkitt Lymphoma. *Cell. 37*, 381-391.

Slack, J. L., D. J. Liska and P. Bornstein. (1991) An Upstream Regulatory Region Mediates High-Level, Tissue-Specific Expression of the Human a1(1) Collagen Gene in Transgenic Mice. *Molecular and Cellular Biology. 11,* 2066-2074.

Smith, G., M. Summers and M. Fraser. (1983) Production of human β-interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol. 3,* 2156-65.

Takebe, Y., M. Seiki, J. I. Fujisawa, P. Hoy, K. Yokota, K. I. Arai, M. Yoshida and N. Arai (1988) SR-alpha promoter - an efficient and versatile mammalian cDNA expression system composed of the simian virus-40 early promoter and the R-U5 segment of human T-cell leukemia-virus type-1 long terminal repeat *Molecular and Cellular Biology, vol.8, no.1* 466-472

Talbot, D., P. Collis, M. Antoniou, M. Vidal, F. Grosveld and D. R. Greaves. (1989) A dominant control region from the human β-globin locus confering integration site-independent gene expression. *Nature. 338,* 352-355.

Talbot, D., P. Descombes and U. Scibler. (1994) The 5' flanking region of the rat LAP (C/EBPb) gene can direct high-level, position -independent expression in multiple tissues in transgenic mice. *Nucleic Acids Research. 22,* 756-766.

Tapon, N. and A.Hall. (1997) Rho Rac and cdc 42 GTPases regulate the organisation of the actin cytoskeleton. *Curr. opin. Cell Biol.* 9 86-92

Thompson, E. M., E. Christians, M. G. Stinnakre and J. P. Renard. (1994) Scaffold Attachment Regions Stimulate Hsp70.1 Expression In Mouse Preimplantation

Embryos But Not In Differentiated Tissues. *Molecular and Cellular Biology. 14,* 4694-4703.

Tiesman, J. and A. Rizzino. (1990) Nucleotide-Sequence Of the 5'-Flanking Region Of the Mouse K-FGF Oncogene Exhibits an Alternating Purine - Pyrimidine Motif With the Potential to Form Z-DNA. *Gene. 96,* 311-312.

Townes, T. M. and R. R. Behringer. (1990) Human globin locus activation region (LAR): role in temporal control. *Trends in Genetics. 6,* 219-223.

Tremethick, D. J. and P. L. Molloy. (1986) High Mobility Group Protein-1 and Protein-2 Stimulate Transcription Invitro By Rna Polymerase-II and Polymerase-III. *Journal Of Biological Chemistry. 261,* 6986-6992.

Turner, B. M. and L. P. Oneill. (1995) Histone Acetylation In Chromatin and Chromosomes. *Seminars In Cell Biology. 6,* 229-236.

Vanholde, K. E., D. E. Lohr and C. Robert. (1992) What Happens to Nucleosomes During Transcription. *Journal Of Biological Chemistry. 267,* 2837-2840.

Vanwijnen, A. J., J. P. Bidwell, E. G. Fey, S. Penman, J. B. Lian, J. L. Stein and G. S. Stein. (1993) Nuclear Matrix Association Of Multiple Sequence-Specific Dna-Binding Activities Related to Sp-1, Atf, CCAAT, C/Ebp, Oct-1, and Ap-1. *Biochemistry. 32,* 8397-8402.

Walters, M. C., S. Fiering, J. Eidemiller, W. Magis, M. Groudine and D. I. K. Martin. (1995) Enhancers Increase the Probability But Not the Level Of Gene-

Expression. *Proceedings Of the National Academy Of Sciences Of the United States Of America. 92,* 7125-7129.

Whitelaw, C. B. A., S. Harris, M. McClenaghan, J. P. Simons and A. J. Clark. (1992) Position-Indepentent expression of the Ovine β-lactoglobulin gene in transgenic mice. *Biochemical Journal. 286,* 31-39.

Williams, D., Van Frank, R., Muth, R., Burnett, J. (1982) *Science. 215,* 687-689.

Wilson, R., R. Ainscough, K. Anderson, C. Baynes, M. Berks, J. Burton, M. Connell, J. Bonfield, T. Copsey, J. Cooper, A. Coulson, M. Craxton, S. Dear, Z. Du, R. Durbin, A. Favello, A. Fraser, L. Fulton, A. Gardner, P. Green, T. Hawkins, L. Hillier, M. Jier, L. Johnston, M. Jones, J. Kershaw, J. Kirsten, N. Laisster, P. Latreille, C. Lloyd, B. Mortimore, M. Ocallaghan, J. Parsons, C. Percy, L. Rifken, A. Roopra, D. Saunders, R. Shownkeen, M. Sims, N. Smaldon, A. Smith, M. Smith, E. Sonnhammer, R. Staden, J. Sulston, J. Thierrymieg, K. Thomas, M. Vaudin, K. Vaughan, R. Waterston, A. Watson, L. Weinstock, J. Wilkinsonsproat and P. Wohldman. (1994) 2.2 Mb Of Contiguous Nucleotide-Sequence From Chromosome-III of C- Elegans. *Nature. 368,* 32-38.

Wolffe, A. P. (1994) Architectural Transcription Factors. *Science. 264,* 1100.

Wu, C., R. Holmgren, K. Livak, W. Yuk-Chor and E. SCR. (1978) Chromatin structure of specific genes. *Journal of Cell Biology. 79,* 113.

Zong, R. T. and R. H. Scheuermann. (1995) Mutually Exclusive Interaction Of a Novel Matrix Attachment Region Binding-Protein and the Nf-Mu-Nr Enhancer Repressor - Implications For Regulation Of Immunoglobulin Heavy-Chain Expression. *Journal Of Biological Chemistry. 270,* 24010-24018.

Appendix 1
Sequences of primers used in this study

PRIMER	Length (bases)	Sequence (5'-3')
Lys 3	24	AGCCGGCAGCCTCTGATCCACGCC
Lys 4	24	GGACGATGTGAGCTGGCAGCGGCT
gpt 3	24	TTGCCGCCACGGTAGGCAATCACC
gpt 4	24	GGTGATTGGCTACCGTGGCGGCAA
gpt 7	20	CGAACGTACCGAGGATGATT
M13rev	24	TCACACAGGAAACAGCTATGACCA
M13for	24	CGACGTTGTAAAACGACGGCCAGT
βlac-1	20	AACAGGAAGGCAAAATGCCG
CMV-1	20	GCATCACCATGGTAATAGCG
CMV nov	20	CCCGGAGGAAAGGAGATGCA
gpt nov	20	CATGCCTGACCCTGACTAGC
T3 (DASH)	30	GCCGCGGAATTAACCCTCACTAAAGGGAAC
T7 (DASH)	30	CGCGGCCGCGTAATACGACTCACTATAGGG
mHS3 for	20	GCAGAAGCCTTTAATCCCAG
mHS3 rev	21	TTGCTGTTTAAGCCTCCTTGG
FlpNBam	31	CGCGGATCCGCCACCATGCCACAATTTGATA
FlpCXho	33	TGATCACTCGAGTATGCGTCTATTTATGTAGGA

Appendix 2

Summaries of FASTA comparisons between C6 and D8 clones and known sequences. In each case a list of the most homologous sequences is presented together with printouts of the best 3-4 matches

Appendix 2A
FASTA search on sequence data obtained from DNA immediately flanking the C6 plasmid (figure 3.10). The B1 repeat element was excluded from the search. No significant homologies were obtained.

The best scores are:

		init1	initn	opt..
em_in:dm27p /rev	L49408 Drosophila melanogaster (clone: ...	65	242	65
em_ro:mmu58105	U58105 Mus musculus Btk locus, alpha-D-ga...	105	238	114
em_hum1:hsl139h8	Z54334 Human DNA sequence from cosmid L...	84	226	104
em_hum2:hsu162c4	Z80903 Human DNA sequence from cosmid U...	95	204	106
em_hum2:hsxpdg1 /rev	L47234 Homo sapiens ERCC2 gene, gen...	80	202	98
em_hum2:hsu66061 /rev	U66061 Human germline T-cell recep...	90	199	105
em_hum1:hsl165d7	Z68273 Human DNA sequence from cosmid L...	65	198	80
em_hum1:hs384d8	U62317 Chromosome 22q13 BAC Clone CIT987...	84	196	104
em_om:ocu60902 /rev	U60902 Otolemur crassicaudatus epsil...	70	192	79
em_hum2:hsu50871	U50871 Human familial Alzheimer's disea...	76	192	98
em_hum2:hsu66060	U66060 Human germline T-cell receptor b...	85	192	88
em_hum1:hs91k3a	Z69917 Human DNA sequence from cosmid 91...	70	190	86
em_hum2:hsu120e2 /rev	Z73496 Human DNA sequence from cos...	72	188	77
em_hum1:hscamf3x1	X96421 H.sapiens Y chromosome cosmid c...	65	188	86
em_in:dm31961 /rev	U31961 Drosophila melanogaster bithor...	65	187	73
em_hum1:hsevmhc /rev	X87344 H.sapiens DMA, DMB, HLA-Z1, ...	85	186	115
em_hum1:hsflng6pd /rev	L44140 Homo sapiens chromosome X ...	91	186	109
em_hum1:hs506g2a /rev	Z82901 Human DNA sequence from PAC...	76	185	84
em_hum1:hsl130g1 /rev	Z68870 Human DNA sequence from cosm...	77	183	84
em_hum2:hsv521f8	Z68331 Human DNA sequence from cosmid V...	82	181	82
em_hum1:hsl185f6	Z69922 Human DNA sequence from cosmid L...	71	180	83
em_hum1:hsf4g12	Z70289 Human DNA sequence from fosmid 4G...	78	178	89
em_in:cem163	Z79603 Caenorhabditis elegans cosmid M163. ...	63	176	76
em_hum2:hsu173h7 /rev	Z80774 Human DNA sequence from cos...	67	175	76
em_or:mimpcg	M68929 Marchantia polymorpha mitochondrion,...	73	175	74
em_hum1:hsbat2 /rev	Z15025 H.sapiens Bat2 gene. 2/93	82	174	100
em_hum1:hs130n4b /rev	Z75888 Human DNA sequence from cos...	77	174	104
em_hum1:hsdgcrcen	L77570 Homo sapiens DiGeorge syndrome ...	76	173	103
em_in:cet19c4 /rev	Z75549 Caenorhabditis elegans cosmid ...	59	173	64
em_hum1:hs5211210	U52112 Human Xq28 genomic DNA in the r...	79	173	89
em_hum1:hs293e14	Z82900 Human DNA sequence from PAC 293E...	83	172	102
em_hum1:hscos12	Z69706 Human DNA sequence from cosmid CO...	65	172	95
em_hum2:hsu66083	U66083 Human contig of two cosmids from...	63	172	107
em_hum2:hsu63312	U63312 Human cosmid LL12NC01-242E1, ETV...	67	172	98
em_hum1:hs67c13	Z80896 Human DNA sequence from PAC 67C13...	72	172	81
em_hum1:hsj13817a	Z72519 Human DNA sequence from cosmid ...	71	172	71
em_ro:mmbtk07	L29795 Mouse Bruton agammaglobulinemia tyr...	105	172	114
em_ro:rncryg	M19359 Rat gamma-crystallin gene cluster, e...	71	172	78
em_hum2:hsu85h7	Z73900 Human DNA sequence from cosmid U8...	86	171	91
em_in:ceuc37c3	U64857 Caenorhabditis elegans cosmid C37C...	60	170	68

csixconnov.dat /rev
em_in:dm27p

ID DM27P standard; DNA; INV; 83522 BP.
AC L49408;
NI g1103950
DT 07-DEC-1995 (Rel. 46, Created)
DT 09-DEC-1995 (Rel. 46, Last updated, Version 2) . . .

SCORES Init1: 65 Initn: 242 Opt: 65
94.4% identity in 18 bp overlap

```
        119       109        99        89        79        69
csixco TCCAGCCCTGCATCTCCTTTCCTCCGGGCTTCTTTTCTTTTGTTTGTTCTTTTCAACAGG
                                        |||||||||||| ||||||
dm27p  GTTTTATGCAAGCAACGGTTCCATTTCAGCTCTTTTCTTTTTTTTGTTTCACTTTTTTTA
         7890      7900      7910      7920      7930      7940

         59        49        39        29        19         9
csixco CTGAGTTTTGNTAAGGTCCGCACAGGAGCATGATAGCTTNCACCAGTGGCTGTGCACTGA

dm27p  ATTAAATTAAAATATTCATCATTTTATCAGATTTTGTGTATGGATATCACACAAATATGG
         7950      7960      7970      7980      7990      8000
```

csixconnov.dat
em_ro:mmu58105

ID MMU58105 standard; DNA; ROD; 88871 BP.
AC U58105;
NI g1666699
DT 13-NOV-1996 (Rel. 49, Created)
DT 13-NOV-1996 (Rel. 49, Last updated, Version 1) . . .

SCORES Init1: 105 Initn: 238 Opt: 114
85.7% identity in 35 bp overlap

```
          360       370       380       390       400       410
csixco GCATGNCAGCACATNNCTTTAGTCCCAGNAGNCAGAGGCAGAGNCAGAGGCAGAGGCAGA
                                          |:|||||  ||||||:||||||  |||||||||
mmu581 TTGGCAGAGGAGAGGGAGGGTTAGGGACGGGGCAGAGACAGAGGCAGAGGGAGAGGCAGA
    64630     64640     64650     64660     64670     64680

          420
csixco GNCAGAG
       |:|||
mmu581 GGCAGGTGGATCTTTGTGAGGACAATCTGATCTACAGAGTGAGTTCCAGGACAGCAAGAG
    64690     64700     64710     64720     64730     64740
```

csixconnov.dat
em_hum1:hsl139h8

ID HSL139H8 standard; DNA; HUM; 31929 BP.
AC Z54334;
NI g1015495
DT 06-OCT-1995 (Rel. 45, Created)
DT 06-OCT-1995 (Rel. 45, Last updated, Version 1) . . .

SCORES Init1: 84 Initn: 226 Opt: 104
70.4% identity in 54 bp overlap

```
     320       330       340       350       360       370
csixco ATTAACAGTATTTANNGATGGCTAGTCNGGGTCAGGCATGNCAGCACATNNCTTTAGTCC
                                          | ||||||||:  ||||||::|| || |||
hsl139 ACCCGTCTCTATACTAAAAATACAAAATTAGCCAGGCATGGGGGCACATGACTGTAATCC
   30110     30120     30130     30140     30150     30160

     380         390       400       410       420
csixco CAG-NAGNC-AGAGGCAGAGNCAGAGGCAGAGGCAGAGNCAGAG
       ||| :| :|  |||||  |||:|||
hsl139 CAGCTACTCGGGAGGCTGAGGCAGGATAATTGCTTGAACCCAGGAGGCAGAGGTTGTGGT
   30170     30180     30190     30200     30210     30220
```

Appendix 2B
FASTA search on sequence data obtained for the distil end of the 6.2kb fragment from the C6 cell line Figure 3.11. The longest stretches of homology are to sequences in a cosmid from *C.elegans* (see text).

The best scores are:

	init1	initn	opt..
em_hum1:hs07000 U07000 Human breakpoint cluster region (...	62	220	68
em_in:cef45e1 U28732 Caenorhabditis elegans cosmid F45E1...	90	215	102
em_fun:scorfsaa /rev L47993 Saccharomyces cerevisiae ORF...	64	213	74
em_hum1:hs473j10 Z81009 Human DNA sequence from PAC 473J...	70	205	70
em_hum1:hscg1160 /rev Z46936 Human cosmid cG1160 from Xq...	76	195	77
em_in:cek02e11 /rev Z77665 Caenorhabditis elegans cosmid...	83	190	88
em_hum1:hsmmda M63796 Human DNA from cosmid MMDA from ch...	68	187	71
em_hum2:hsu66083 U66083 Human contig of two cosmids from...	61	182	82
em_in:cer153 /rev U28729 Caenorhabditis elegans cosmid R...	71	182	84
em_in:cef35b12 Z74032 Caenorhabditis elegans cosmid F35B...	65	180	70
em_hum1:hsfbrg1 /rev M10014 Human fibrinogen gamma chain...	60	179	68
em_hum1:hsfbrgg /rev X02415 Human gene for fibrinogen ga...	60	179	68
em_hum1:hsn37f10 Z69714 Human DNA sequence from cosmid N...	62	178	71
em_in:cec05c9 /rev Z68000 Caenorhabditis elegans cosmid ...	62	174	78
em_hum1:hs232122 Z73986 Human DNA sequence from cosmid 2...	76	173	84
em_in:cec09g9 /rev Z54235 Caenorhabditis elegans cosmid ...	61	172	85
em_hum1:hs5211210 U52112 Human Xq28 genomic DNA in the r...	57	171	65
em_in:cezk1193 U41553 Caenorhabditis elegans cosmid ZK11...	65	170	85
em_in:cef13h8 /rev U23139 Caenorhabditis elegans cosmid ...	62	170	67
em_in:cef43c9 U40427 Caenorhabditis elegans cosmid F43C9...	69	170	71
em_fun:sc9346 /rev Z48784 S.cerevisiae chromosome IV cos...	92	170	115
em_hum1:hso19a Z81310 Human DNA sequence from cosmid 019...	62	169	70
em_in:cet13f2 /rev Z81122 Caenorhabditis elegans cosmid ...	62	169	69
em_in:cef33g12 U41278 Caenorhabditis elegans cosmid F33G...	60	169	86
em_ro:rncryg M19359 Rat gamma-crystallin gene cluster, e...	56	169	58
em_fun:sc5610 Z38060 S.cerevisiae chromosome IX sequence...	58	167	66
em_in:cec32d5 /rev U23511 Caenorhabditis elegans cosmid ...	58	166	64
em_ro:mmbgcxd /rev X14061 M.musculus beta-globin complex...	58	166	65
em_in:cec34b4 Z78059 Caenorhabditis elegans cosmid C34B4...	64	165	81
em_in:cef52h3 Z66512 Caenorhabditis elegans cosmid F52H3...	63	165	69
em_ph:br38906 U38906 Bacteriophage r1t integrase, repres...	56	165	56
em_in:cef30a10 Z81072 Caenorhabditis elegans cosmid F30A...	62	165	72
em_hum1:hs65o19 /rev Z73358 Human DNA sequence from cosm...	57	165	57
em_ba:ecapah02 D10483 E.coli K12 genome, 0-2.4min. regio...	60	164	60
em_ba:ecuw76 /rev U00039 E. coli chromosomal region from...	57	163	57
em_ba:ssd904 /rev D90904 Synechocystis sp. PCC6803 compl...	57	162	73
em_in:cet03f6 Z81113 Caenorhabditis elegans cosmid T03F6...	87	162	91
em_fun:scrpd3cos Z46259 S.cerevisiae FY1676 RPD3 gene. 4/96	54	162	64
em_in:dmu62892 /rev U62892 Drosophila melanogaster retin...	54	162	61
em_hum1:hsn119a7 /rev Z80901 Human DNA sequence from cos...	77	161	77

pBSE62for.dat
em_hum1:hs07000

ID HS07000 standard; DNA; HUM; 152141 BP.
AC U07000;
NI g487344
DT 02-JUN-1994 (Rel. 39, Created)
DT 31-JAN-1996 (Rel. 46, Last updated, Version 32) . . .

SCORES Init1: 62 Initn: 220 Opt: 68
60.4% identity in 53 bp overlap

```
          310       320       330       340       350       360
pBSE62 CAAGGTATCCTGGAGGNTCTAGGGAAGACCCTGATCTCTAGANCANCTGTCTACTA-TCC
                                    |||||||| || :||:|||||   || |||
hs0700 GACTCGGGTGGGCTATGCAGGAAGCCTGTGCTGATCTCAAGGCCAACTGTCCTGTACTCC
         115480    115490    115500    115510    115520    115530

             370       380
pBSE62 AATGTAGGNCTCAATCTACANAGGATG
       |     ||:  ||  || | :||
hs0700 AGGCCGGGGACCAGGCTGCTGAGCGCTTCCGAGAGGGACATCTCCTGGGCAGGGAGTGAC
         115540    115550    115560    115570    115580    115590
```

pBSE62for.dat
em_in:cef45e1

ID CEF45E1 standard; DNA; INV; 38122 BP.
AC U28732;
NI g1463022
DT 16-JUN-1995 (Rel. 44, Created)
DT 01-OCT-1996 (Rel. 49, Last updated, Version 64) . . .

SCORES Init1: 90 Initn: 215 Opt: 102
57.9% identity in 159 bp overlap

```
                      10        20        30        40        50
pBSE62          TTCCCTGCTTAGCTATTGGGGCATCATTAGACCTCTTCT-GGGGCTCAGCT
                                        ||||||  |   | |   | ||  | ||| ||
cef45e GCATATAATTTACATCTAGAGAAGACCACCCATCATCATCTCACAGGTAGGTCCACATCT
         33260     33270     33280     33290     33300     33310

              60        70        80        90       100
pBSE62 GTCATTATGCCAGATATC-TTTATTTTA-GAAATACTGTATGGACCTTTCTCAGACATT-
        || || ||      |  | | | |||||   |||  ||| ||  ||    | || ||| |
cef45e -TCCTTCTGTTGTTTCCCGTATGTTTTATTCAATTTTGT-TG--CCGAACACACACAATA
            33320     33330     33340     33350     33360

       110       120       130       140       150       160
pBSE62 -ATAGCACTCAGAGTGGGATAATCTTCCTTTCTCCCTTTGTTAGTATTCGATGGCAATGT
        | |||    |  |||     || |||||||| |||| ||| | || | ||||||   | |
cef45e GAGAGCCAACCTAGTCACTTATTCTTCCTCTCTCTCTTCGCTAATTTTCGATTCAATTCC
    33370     33380     33390     33400     33410     33420

       170       180       190       200       210       220
pBSE62 CACATTTCTGAGCATGTCACTACTACCTAACCATCAGTCCTCCCCTCGACTCTGGAAACT
       |  ||||||
cef45e CCTATTTCTCTTCGCCCTCGTTGCCCTCTAGCTTTGCGACCACACTGCATAATATTCATT
    33430     33440     33450     33460     33470     33480
```

pBSE62for.dat /rev
em_fun:scorfsaa

ID SCORFSAA standard; DNA; FUN; 61989 BP.
AC L47993;
NI g1019675
DT 14-OCT-1995 (Rel. 45, Created)
DT 14-OCT-1995 (Rel. 45, Last updated, Version 1) . . .

SCORES Init1: 64 Initn: 213 Opt: 74
59.8% identity in 87 bp overlap

```
       209       199       189       179       169       159
pBSE62 AGGGGAGGACTGATGGTTAGGTAGTAGTGACATGCTCAGAAATGTGACATTGCCATCGAA
                                       |||  | | ||  |  || ||     |  |
scorfs TCTTCCAAACATTAATTTGTTCCACGGGTCCATTTTTATAATTACGAAATAATTGTGCAT
         8030      8040      8050      8060      8070      8080

       149       139       129       119       109        99
pBSE62 TACTAACAAAGGGAGAAAGGAAGATTATCCCACTCTGAGTGCTATAATGTCTGAGAAAGG
       |  ||||||||||||||||  |   |||||  |  ||  |||    ||  ||  ||| ||
scorfs TGGTAACAAAGGGAGAAAGCGA---TATCCTAAACT-AGTAGGTTACAGT-TGATAATAT
         8090      8100          8110      8120      8130

        89        79        69        59        49        39
pBSE62 TCCATACAGTATTTCTAAAATAAAGATATCTGGCATAATGACAGCTGAGCCCCAGAAGAG

scorfs AAAAAGTTGGAAAATAGCTTAGAAATGAAGAAGAGAATAACGGCTTCTATGCGGTTGTGA
```

Appendix 2C
FASTA search on sequence data obtained for the proximal end of the 6.2kb fragment from the C6 cell line Figure 3.11. As with the distil end, many stretches of homology are to sequences in a cosmid from *C.elegans* (see text).

The best scores are:

	init1	initn	opt..
em_ro:mmtsxdna X99946 M.musculus 94kb genomic sequence e...	60	204	60
em_vi:as18466 /rev U18466 African swine fever virus, com...	65	179	65
em_hum1:hsdgcrcen /rev L77570 Homo sapiens DiGeorge synd...	64	177	64
em_fun:scchrxvi /rev Z71255 S.cerevisiae chromosome XVI ...	59	176	64
em_hum1:hsretblas L11910 Human retinoblastoma susceptibi...	67	175	71
em_fun:spac12b10 /rev Z70721 S.pombe chromosome I cosmid...	60	172	60
em_in:cek02a4 Z67883 Caenorhabditis elegans cosmid K02A4...	62	171	62
em_hum1:hs65o19 /rev Z73358 Human DNA sequence from cosm...	63	170	64
em_vi:hsgend /rev X64346 Herpesvirus saimiri complete ge...	60	170	60
em_ro:mmtcra /rev M64239 Mouse T-cell receptor alpha/del...	66	167	76
em_fun:spac12c2 /rev Z54140 S.pombe chromosome I cosmid ...	56	167	61
em_in:cec44e12 /rev U39647 Caenorhabditis elegans cosmid...	59	167	64
em_fun:scchrix_1 /rev Continuation of SCCHRIX from base ...	57	167	70
em_in:cec08b6 Z72502 Caenorhabditis elegans cosmid C08B6...	58	163	58
em_hum2:hsu66082 /rev U66082 Human contig of seven cosmi...	69	163	70
em_ba:mju67511 /rev U67511 Methanococcus jannaschii from...	54	162	104
em_in:ceb0464 /rev Z19152 Caenorhabditis elegans cosmid ...	56	162	86
em_in:cec47g2 Z49125 Caenorhabditis elegans cosmid C47G2...	53	162	53
em_in:dm25p /rev L49405 Drosophila melanogaster (clone: ...	57	161	61
em_hum1:hs232122 Z73986 Human DNA sequence from cosmid 2...	54	161	54
em_hum2:hstitin2b /rev X90568 H.sapiens mRNA for titin p...	66	160	67
em_hum1:hsfmr1s L29074 Homo sapiens fragile X mental ret...	61	160	76
em_or:chzmxx X86563 Z.mays complete chloroplast genome. ...	63	160	70
em_in:cec29a12 /rev Z73970 Caenorhabditis elegans cosmid...	55	159	56
em_hum1:hs358h7 Z77249 Human DNA sequence from PAC 358H7...	64	159	83
em_hum1:hsnfg9 /rev Z69719 Human DNA sequence from cosmi...	70	158	80
em_or:chevcg /rev M81884 Epifagus virginiana chloroplast...	71	157	72
em_in:cer01e6 /rev Z68118 Caenorhabditis elegans cosmid ...	57	157	57
em_or:chevcg M81884 Epifagus virginiana chloroplast comp...	71	157	72
em_in:cet06h11 Z49889 Caenorhabditis elegans cosmid T06H...	60	156	79
em_hum1:hsdnail15 X91233 H.sapiens IL15 gene. 9/95	54	156	55
em_hum1:hsids /rev L35485 Homo sapiens iduronate sulphat...	53	156	61
em_hum1:hs246o8 Z76735 Human DNA sequence from PAC 246O8...	54	154	75
em_in:cek10c8 Z74474 Caenorhabditis elegans cosmid K10C8...	67	153	85
em_om:ocu60902 /rev U60902 Otolemur crassicaudatus epsil...	70	150	84
em_in:cef56d1 U39997 Caenorhabditis elegans cosmid F56D1...	70	149	90
em_vi:hevzvxx X04370 Varicella-Zoster virus complete gen...	73	149	77
em_in:cet22g5 Z81127 Caenorhabditis elegans cosmid T22G5...	77	148	79
em_hum2:hsu230b10 Z68339 Human DNA sequence from cosmid ...	49	148	75
em_hum1:hs92m18a Z73359 Human DNA sequence from cosmid 9...	67	148	74

pBSE62rev.dat
em_ro:mmtsxdna

ID MMTSXDNA standard; DNA; ROD; 94459 BP.
AC X99946;
NI e1002263
DT 15-NOV-1996 (Rel. 49, Created)
DT 15-NOV-1996 (Rel. 49, Last updated, Version 2) . . .

SCORES Init1: 60 Initn: 204 Opt: 60
100.0% identity in 15 bp overlap

```
     290       300       310       320       330       340       349
pBSE62 TCTTTGGGTGAACCTATTAAGTCTGCCACGTTTTAATCTTTTTCAGAAAACCCNCAACCC
                                    ||||||||||||||||
mmtsxd GAAAAATGGAAAGGGTCAAAGCACAGGCATTTTTAATCTTTTTCAAGTTCTTTCAATTTG
   90440     90450     90460     90470     90480     90490

     350       360       369
pBSE62 ATTTANGATTAGAGGGTTCC

mmtsxd GTCTTTCGTTCCTCCAGTTTTCTGATTGTTGTCAATTTTATTATTCTTAAATTTTATGCA
   90500     90510     90520     90530     90540     90550
```

pBSE62rev.dat /rev
em_vi:as18466

ID AS18466 standard; DNA; VRL; 170101 BP.
AC U18466;
NI g780375
DT 29-APR-1995 (Rel. 43, Created)
DT 29-APR-1995 (Rel. 43, Last updated, Version 2) . . .

SCORES Init1: 65 Initn: 179 Opt: 65
94.4% identity in 18 bp overlap

```
      369       359       349       339       329       319
pBSE62   GGAACCCTCTAATCNTAAATGGGTTGNGGGTTTTCTGAAAAAGATTAAAACGTGGCAG
                                    |||||| ||||||||||
as1846 CTCGGTGACGGAGGTTATTTTCTGGAAATCGGTTTTTTGAAAAAGATTTTCAATGTGTTT
       92780     92790     92800     92810     92820     92830

      309       299       289       279       269       259
pBSE62 ACTTAATAGGTTCACCCAAAGACATGCCATACATATCCAAATGCCCAAATTAAGAGAGTT

as1846 GCGGGTTGAGTTGCTTTGCAGTCCATACAAGACATCAAAAAATTCAATCAGCAAAAACTT
       92840     92850     92860     92870     92880     92890
```

pBSE62rev.dat /rev
em_hum1:hsdgcrcen

ID HSDGCRCEN standard; DNA; HUM; 108400 BP.
AC L77570;
NI g1377755
DT 28-JUN-1996 (Rel. 48, Created)
DT 28-JUN-1996 (Rel. 48, Last updated, Version 1) . . .

SCORES Init1: 64 Initn: 177 Opt: 64
100.0% identity in 16 bp overlap

```
       99        89        79        69        59        49
pBSE62 GAAGATGCCTCCCAAGTATGAGGTAGTAACTAGAAACAGGAAGTAGATCTTAACCAAGCA
                                    ||||||||||||||||
hsdgcr TATATGGAGTATCTAAAGCAGTCAAACTCTTAGAAACAGGAAGTAGGACGGTGGTTGCCA
          73310     73320     73330     73340     73350     73360

       39        29        19         9
pBSE62 AGGAAGCAAAGACTATGTTNCCACNACCACAGAGAAAGGAA

hsdgcr GGCTGGAGAGAGAGATAAAAGGGAGGGGTAGTTACTCAATAGGTACTGAGTTTCAGTTTT
          73370     73380     73390     73400     73410     73420
```

pBSE62rev.dat /rev
em_fun:scchrxvi

ID SCCHRXVI standard; DNA; FUN; 165536 BP.
AC Z71255;
NI g1279666
DT 19-APR-1996 (Rel. 47, Created)
DT 10-MAY-1996 (Rel. 47, Last updated, Version 2) . . .

SCORES Init1: 59 Initn: 176 Opt: 64
73.3% identity in 30 bp overlap

```
       179       169       159       149       139       129
pBSE62 GTGGTTCCAATTACCTTATTGTAGAACCTATATACATCAGCTCTGAAAATCAGGAGATCA
```

```
                                          ||  |||||||  | |||||||||  || |
scchrx AAATATCAATGAAACCAACCAAAATCACAGTAATCATCAGCATTCAAAATCAGGCAATAA
          43290     43300     43310     43320     43330     43340

        119       109       99        89        79        69
pBSE62 GCATGAGAAAGGCTCGAGCTTTGAAGATGCCTCCCAAGTATGAGGTAGTAACTAGAAACA

scchrx CTGGCAAGATAGTTCGGTAAGCTTGCCAGCGAAAGCTGATTCACGTCTTAACATGATGGA
          43350     43360     43370     43380     43390     43400


pBSE62rev.dat
em_hum1:hsretblas

ID   HSRETBLAS  standard; DNA; HUM; 180388 BP.
AC   L11910;
NI   g292420
DT   06-MAR-1993 (Rel. 35, Created)
DT   12-DEC-1994 (Rel. 42, Last updated, Version 7) . . .


SCORES            Init1:    67 Initn:   175 Opt:    71
           66.7% identity in 39 bp overlap

          300       310       320       330       340       350
pBSE62 TGGGTGAACCTATTAAGTCTGCCACGTTTTAATCTTTTTCAGAAAACCCNCAACCCATTT
                                          ||||||||||   ||   : |||||||||
hsretb GATAGGACTTAACATTCCCCATACCCCTACAATCTTTTTCATCCAATATTTAACCCATTT
    70920     70930     70940     70950     70960     70970

          360      369
pBSE62 ANGATTAGAGGGTTCC
       :  |||  |
hsretb TTATTTACAATTCTTACCTGGCCCTGTTAGCATTATAGTTTGCAACTCCTGATAGACAAT
    70980     70990     71000     71010     71020     71030


pBSE62rev.dat /rev
em_fun:spac12b10

ID   SPAC12B10  standard; DNA; FUN; 37059 BP.
AC   Z70721;
NI   g1262413
DT   11-APR-1996 (Rel. 47, Created)
DT   11-APR-1996 (Rel. 47, Last updated, Version 1) . . .


SCORES            Init1:    60 Initn:   172 Opt:    60
          100.0% identity in 15 bp overlap

          69        59        49        39        29        19
pBSE62 AACTAGAAACAGGAAGTAGATCTTAACCAAGCAAGGAAGCAAAGACTATGTTNCCACNAC
                                          ||||||||||||||||
spac12 AACGGGCGCAGGGCCGTCGTAAGATATATGGCAAGGAAGCAAAGAAGCTTTCTCTAAGTT
    28290     28300     28310     28320     28330     28340

           9
pBSE62 CACAGAGAAAGGAA

spac12 TGAGCTTCCTGCTTGCTCCAATTTGACAATAGGTTTGACTGAATCCATTGTGCTCTAAGA
    28350     28360     28370     28380     28390     28400


pBSE62rev.dat
em_in:cek02a4
```

Appendix 2D
FASTA search on sequence data obtained for the proximal end of the 8kb *Sal-Sal* fragment from the D8 cell line Figure 4.7. As would be expected, the only strong homologies found were to CMV polyadenylation signal.

The best scores are: initl initn opt..

em_vi:hehcmvcg /rev X17403 Human Cytomegalovirus Strain ...1188 1188 1193
em_vi:hehs5ie11 M26973 Human cytomegalovirus immediate e...1153 1153 1154
em_vi:hehs5ier M11298 Human cytomegalovirus (strain Town... 972 972 973
em_vi:hehs5mie M21295 Human cytomegalovirus (HCMV) major... 364 364 368
em_vi:hehs5m03 M11630 Human cytomegalovirus (Towne) majo... 360 360 364
em_sy:cv457 U02457 Cloning vector pYEUra3, complete sequ... 244 328 497
em_sy:a13388 A13388 pSPT18 DNA sequence. 1/94 244 328 497
em_sy:cvcole1 /rev M69063 Cloning vector pAM34. 7/95 244 324 497
em_sy:cv13856 /rev U13856 pGEX-5X1 cloning vector, compl... 264 300 517
em_sy:cv13857 /rev U13857 pGEX-5X2 cloning vector, compl... 264 300 517
em_sy:cv13854 /rev U13854 pGEX-4T2 cloning vector, compl... 264 300 517
em_sy:cv13853 /rev U13853 pGEX-4T1 cloning vector, compl... 264 300 517
em_sy:cv13858 /rev U13858 pGEX-5X3 cloning vector, compl... 264 300 517
em_sy:cvpgex3 /rev M21676 Plasmid pGEX-1 expression vect... 264 300 517
em_sy:cv13851 /rev U13851 pGEX-2TK cloning vector, compl... 264 300 517
em_sy:cv13850 /rev U13850 pGEX-2T cloning vector, comple... 264 300 517
em_sy:agglutran /rev M97937 Plasmid pGEX-5G/LIC expressi... 264 300 517
em_sy:cv13849 /rev U13849 pGEX-1 Lambda T cloning vector... 264 300 517
em_sy:cv13855 /rev U13855 pGEX-4T3 cloning vector, compl... 264 300 517
em_sy:cv13852 /rev U13852 pGEX-3X cloning vector, comple... 264 300 517
em_sy:apdg56 X67018 Artificial DNA sequence (pGD56) of p... 244 292 497
em_sy:cvecoyst L11060 Cloning vector sequence (E. coli/y... 244 290 497
em_sy:sctram2 M74016 Cloning vector pYADE4 TRP1 and AMPr... 244 290 497
em_sy:ec1trb L07041 pMHNeo eukaryotic expression vector,... 244 288 497
em_sy:cvptl61t M29896 Cloning vector pTL61T, complete se... 244 288 497
em_sy:ec1tra L07040 pFNeo eukaryotic expression vector, ... 244 288 497
em_sy:cv19585 /rev U19585 Cloning vector pPROEX-1, compl... 248 284 507
em_sy:a28163 /rev A28163 Plasmid pKK177-3/GLUCPI_Arg6 DN... 244 280 497
em_un:i13349 /rev I13349 Sequence 5 from patent US 54361... 244 280 497

SCORES Init1: 1188 Initn: 1188 Opt: 1193
94.3% identity in 335 bp overlap

```
            359       349       339       329       319       309
14rev.  GCGCCTGATGCGGTATTTTCTCCTNACGCATCTGTGCGGTATTTCACACCGCATAAATTC
                                          | || ||||| | | || | |   |
hehcmv  TTTAATAATCACCTTGAAAGGACGCGTGGTTTTGCGCGGTTTCTTACGCGGGCT------
         170630    170640    170650    170660    170670

            299       289       279       269       259       249
14rev.  GAGCTCGGGAGTAGCGGATGCCCCGGGGAGAGGAGTGTTAGTAACCGCGACGCTGGTGGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
hehcmv  GAGCTCGGGAGTAGCGGATGCCCCGGGGAGAGGAGTGTTAGTAACCGCGACGCTGGTGGG
     170680    170690    170700    170710    170720    170730

            239       229       219       209       199       189
14rev.  GGTCGGCTTGTTAAGAGGGGCGCTGCTAACGCTGCAAGAGTGGGTTGTCAGCGTGGGGCC
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
hehcmv  GGTCGGCTTGTTAAGAGGGGCGCTGCTAACGCTGCAAGAGTGGGTTGTCAGCGTGGGGCC
     170740    170750    170760    170770    170780    170790

            179       169       159       149       139       129
14rev.  GGTGCTACTGGAATCGATACCGGCATGATTGACAGCCTGGGCGAGGATGTCACCTGATGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
hehcmv  GGTGCTACTGGAATCGATACCGGCATGATTGACAGCCTGGGCGAGGATGTCACCTGATGG
     170800    170810    170820    170830    170840    170850

            119       109        99        89        79        69
14rev.  TGATAAGAAGACACGGGAGACTTAGTACGGTTTCACAGGCGTGACACGTTTATTGAGTAG
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
hehcmv  TGATAAGAAGACACGGGAGACTTAGTACGGTTTCACAGGCGTGACACGTTTATTGAGTAG
     170860    170870    170880    170890    170900    170910

             59        49        39        29        19         9
14rev.  GATTACAGAGTATAACATAGAGTATAATATAGAGTATACAATAGTGACGTGGGATCCTCT
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
hehcmv  GATTACAGAGTATAACATAGAGTATAATATAGAGTATACAATAGTGACGTGGGATCCATA
     170920    170930    170940    170950    170960    170970


14rev.  AGAGTC
        |  |||
hehcmv  ACAGTAACTGATATATATATACAATAGTTTACTGGTCAGCCTTGCTTCTAGTCACCATAG
     170980    170990    171000    171010    171020    171030
```

Appendix 2D
FASTA search on sequence data obtained for the distil end of the 8kb *Sal-Sal* fragment from the D8 cell line Figure 4.8. The only strong homologies found were to mouse satellite sequence.

The best scores are:

		init1	initn	opt..
gb_ro:s72350	S72350 {satellite DNA sequence} [mice, BALB...	556	894	918
em_ro:s72350	S72350 {satellite DNA sequence} [mice, BALB...	556	894	918
em_ro:rslin4a	M13101 Rat long interspersed repetitive DN...	553	874	912
em_ro:rnl1rto2b	X61295 R.norvegicus L1 retroposon, ORF2 ...	553	867	935
em_ro:rnlined	X53581 R.norvegicus long interspersed repe...	567	837	928
em_ro:s80119	S80119 reverse transcriptase homolog {L1 re...	527	834	928
em_ro:rslin3a	M13100 Rat long interspersed repetitive DN...	527	830	928
em_ro:mm15647	U15647 Mus musculus (ORF 1) and reverse tr...	567	829	931
em_ro:mml1a1	M29324 Mouse L1Md-A13 repetitive sequence. ...	567	828	931
em_ro:rnl1rto2a	X61294 R.norvegicus L1 retroposon, ORF2 ...	520	823	914
em_ro:rnvaoxy /rev	X59496 Rat genes for vasopressin, oxy...	534	809	928
em_ro:rninsi /rev	J00747 Rat insulin-I (ins-1) gene. 7/95	527	803	912
em_ro:s72355	S72355 {satellite DNA sequence} [rats, live...	525	791	865
gb_ro:s72355	S72355 {satellite DNA sequence} [rats, live...	525	791	865
em_ro:mmrslima	M13002 Mouse L1Md-A2 repetitive element w...	567	766	917
em_ro:rnrl115	X03093 Rat LINE1 5' sequence 3kb upstream ...	274	754	850
em_ro:mml1fte	X57795 M.musculus F-type L1 element DNA. 5/91	539	715	861
em_ro:mmbamhia	M13164 Mouse BamHI family (ER-1) repeat, ...	222	697	842
em_ro:s72351	S72351 {satellite DNA sequence} [Mastomys c...	565	685	884
gb_ro:s72351	S72351 {satellite DNA sequence} [Mastomys c...	565	685	884
em_sy:agmusint /rev	M16571 Mouse tumor cell line DNA int...	532	655	672
em_ro:mml1m9	M29325 Mouse L1Md-9 repetitive sequence. 7/91	560	641	887
em_ro:mmmopc /rev	M11515 Mouse MOPC 41 plasmacytoma P-1-...	567	641	931
em_ro:rnrl13	X03095 Rat LINE1 sequence RL1.3 (random gen...	535	639	880
em_ro:cgphc312	X79295 C.griseus repetitive DNA, clone pH...	459	549	725
em_ro:pml1pm62x	M97517 Deer mouse (L1Pm62) gene. 12/92	280	527	733
em_ro:rncyp2a1 /rev	M33312 Rat hepatic steroid hydroxyla...	276	519	604

LOCUS S72350 474 bp DNA ROD 27-JAN-1995
DEFINITION {satellite DNA sequence} [mice, BALB/cA Jc1, liver, Genomic, 474 nt].
ACCESSION S72350
KEYWORDS . . . .

SCORES Init1: 556 Initn: 894 Opt: 918
80.3% identity in 355 bp overlap