APPROACHES TO THE CLONING OF HIGH ACTIVITY

REGULATORY ELEMENTS FROM MYELOMA CELLS

by

Nicholas R. Sutcliffe BSc(Hons)

Dept of Biochemistry

University of Leicester

June 1995

UMI Number: U543391

UMI®

Dissertation Publishing

ProQuest®

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

Acknowledgements

For Sue

Approaches to the Cloning of High Activity Regulatory Elements from Myeloma Cells

N Sutcliffe 1995

Cloning strategies were developed to isolate genomic sequence from the site of plasmid vector integration of myeloma J558L transformants.

Two potential approaches were examined. Ligation-mediated anchor PCR was shown to be capable of generating amplification products that included sequence from the integrated plasmid. However, when these were cloned and analysed, they were found to contain various arrangements of primers and plasmid sequence, but no genomic flanking sequence. Further refinement of the procedure failedto generate any products containing flanking sequence.

λ-phage libraries were constructed from transformant genomic DNA. These were screened for clones containing plasmid sequence. A clone was identified, purified and analysed. Although it did contain both plasmid and unknown genomic sequence, its size and arrangement did not tally with information previously gleaned about the integration site and so its significance was hard to establish.

Ligation-mediated anchor PCR has been shown to be a potentially quick and convenient cloning method but one which is extremely technically demanding. It is unclear why it failed in this case to generate useful product. λ-phage cloning is a more laborious approach which did produce results. Although the clone generated was not obviously useful, the potential of the procedure has been demonstrated and it should form the basis of future investigations.

Contents

Chapter I

Introduction

1.1 Introduction

The emergence of molecular biology in the last 15 years has had major consequences for the commercial production of proteins, such as insulin, which have therapeutic and other uses. Proteins that previously required tedious extraction and purification from large quantities of native tissue, can now be produced cheaply and effectively using recombinant means. Because of the considerable cost savings, this is commercially attractive and produces proteins free of potentially dangerous contaminants such as retroviruses.

In order to produce a recombinant protein, the gene or cDNA coding for it must first be identified and isolated. It must then be engineered into an expression vector containing suitable regulatory elements, which enable the gene to be expressed in the host system. The vector containing the recombinant gene must then be passed into the host, where expression takes place. The recombinant protein produced can be harvested by collecting a supernatant or by lysing the host and purifying the protein of interest.

There is a wide choice of possible hosts, ranging from bacterial systems such as *Escherichia coli*, which are relatively easy to grow in large quantities, to insect and animal-cell systems. Large eukaryotic proteins often require extensive post-translational processing in order to be fully active and suitable mechanisms to achieve this are only available in eukaryotic expression systems. This makes these systems particularly attractive in the production of therapeutic proteins.

Before looking in more depth at these expression systems, it is important to consider those aspects of mammalian gene expression that underpin recombinant protein technology because high levels of recombinant

protein depend on the optimisation of gene expression at every stage at which it is regulated.

1.2   DNA Structure

Three structures have been deduced for the DNA molecule. A and B form DNA are both right-handed double helices [Dickerson et al., 1982]. Both have two anti-parallel strands that wind around each other, linked by purine-pyrimidine hydrogen bonds. B form has 10.5 bp per turn and a marked major and minor groove whilst A form DNA has 11-12 bp per turn, a deeper major and a shallower minor groove. A third DNA structure [Rich and Wang, 1984], Z form, has base-paired, anti-parallel strands wound in a left-handed helix. This has 12-13 bp per helical turn and whilst the major groove is flat, the minor groove is deep. The presence of Z-DNA in the eukaryotic chromosome can be shown by Z-DNA-specific antibodies and is thought to occur when the C-5 molecule of cytosine is methylated or in stretches of purine-pyrimidine repeats.

DNA structure at this level has considerable relevance to gene expression. DNA binding proteins recognise specific DNA sequences through interactions with base residues in the major and minor grooves. Changes in the topological structure of the DNA may inhibit or enhance such interactions [Atchison, 1988]. Because hydrogen bonding between adenine and thymine is less stable than between guanine and cytosine, AT-rich regions have a looser structure and can partially unwind under certain conditions. Many 5' flanking regions are AT-rich, suggesting that a local relaxing of structure may affect the binding of transcriptional regulatory elements.

Supercoiling can also impose a torsional strain on the DNA molecule [Negri et al., 1994]. This can occur in circular molecules, such as plasmids, and stretches of DNA with anchored ends. Naturally occurring DNA tends to be underwound and so contains negative supercoils. This makes factors such as interactions with nucleosomes, AT breathing and Z-DNA formation more energetically favourable.

1.3 Chromatin

Native eukaryotic DNA is invariably associated with a range of proteins and in this form it is called chromatin [Travers, 1994 ; van Holde, 1989]. Histone proteins are an important structural element of chromatin. Histones H2A, H2B, H3 and H4 are small basic proteins with sequences that are highly conserved amongst eukaryotes. Two molecules of each of these histones form an octamer around which is wrapped 1.8 turns of DNA comprising about 145 bp. Each of these nucleosome core particles is separated from its neighbours by about 60 bp of linker DNA. In low ionic strength solution, these particles appear as 'beads on a string' [Richmond et al., 1984 ; Thoma et al., 1979]. Histone H1 is a lysine-rich histone that binds to the linker sequence and stabilises the chromatin structure. In the presence of histone H1, chromatin can form a highly compact 30 nm diameter solenoid fibre containing about 7 nucleosomes per turn [Schwarz and Hansen, 1994 ; Hansen et al., 1991].

This compaction of chromatin results from both nucleosome-nucleosome interaction and the neutralisation of the negatively charged DNA in the linker region by cations.

Stretches of chromatin that lack H1 appear to exist in an equilibrium of extended and compacted states. While aggregation can occur without histone H1 under appropriate ionic conditions, the presence of histone H1

seems to make a large energetic contribution to chromatin folding [Yao et al., 1991]. Chromatin compaction is a salt-dependent structural transition that appears to be an intrinsic property of chromatin in solution. Another such transition is the dissociation of chromatin via the release of intact octamers. This presumably reflects the strength of nucleosome-histone octamer interactions compared to nucleosome-DNA interaction. Individual histones can also dissociate via the initial release of H2A/H2B histone dimers followed by an H3/H4 histone tetramer [Ausio, 1992]. This may be significant in transcription. Certain evidence indicates that, while the intact nucleosome inhibits the passage of the RNA polymerase, the presence of a H3/H4 histone tetramer does not [Almouzni et al., 1991]. Thus the dissociation of H2A/H2B histone dimers from the nucleosome, leaving the remaining histones in place, may facilitate transcription. However, most current research indicates that the whole histone octamer is displaced by the passage of the RNA polymerase, after which it can rebind either at the same or a different site [Kornberg and Lorch, 1995].

It is thought that the 30 nm solenoid is inactive and some form of decondensation is necessary for transcription to take place [Garrard, 1991]. This may be a point of regulation and histone modification may represent a control mechanism [Wolffe, 1994]. Acetylation of the octamer core has been shown to affect the linker DNA [Ridsdale et al., 1990]. This has an effect on the higher-order chromatin structures stabilised by histone H1 [Davie and Hendzel, 1994]. Phosphorylation of the linker histone may also have a role in stabilising or destabilising higher order chromatin structure. The site of phosphorylation may govern the nature of the effect [Churchill and Travers, 1991]. Numerous other factors may also affect the stability of chromatin structure.

1.4 Chromatin and Gene Expression

Nucleosomes also seem to have a role in regulating gene expression by competition with trans-acting factors. This may occur through a pre-emptive mechanism whereby the first molecules to bind to the DNA after replication cannot be displaced [Becker, 1994; Felsenfeld, 1992; Wolffe, 1988; Lorch et al., 1987]. If histones are the first molecules to bind the DNA, the gene locus concerned is inactive. If trans-acting factors bind first, the locus is activated. The presence of histones irreversibly blocks the binding sites for trans-acting factors. This means that regulatory elements can only be activated immediately after replication, before the histones are in place [Lu et al., 1994].

Other possible mechanisms allow genes to be activated without replication. Histones and binding factors are in dynamic equilibrium and trans-acting factors can still bind the DNA in the presence of histones [Hager et al., 1993; Almer and Hørz, 1986]. The protein binding site may be relatively accessible: in the linker DNA between nucleosomes; at the points of entry or exit from the nucleosome; or on the surface of the nucleosome itself. Binding sites buried within the nucleosome may be made accessible to one trans-acting factor by the cooperative action of another factor, binding at another site [Vettesedadey et al., 1994; Chasman et al., 1990]. The wound nature of chromatin makes interaction possible between sites well separated in linear sequence, but brought together by the higher order structure of chromatin.

Chromatin condenses into a 30 nm fibre with the histone H1 molecule [Bartolome et al., 1994]. This fibre appears to have only a limited helical content and may lack regular structure [van Holde and Zlatanova, 1995]. This fibre is further organised into looped domains, apparently fixed at the ends to a proteinaceous nuclear matrix. These loops may have a role in

containing topological strains, and also in the modulation of gene expression [Getzenberg, 1994]. The loop may represent a unit whose genes are all expressed according to the regulatory elements it contains, and which are unaffected by regulatory elements beyond the boundaries of the loop [Sippel et al., 1993]. A number of putative boundary elements and Matrix Attachment Regions (MARs) have been identified [Eissenberg and Elgin, 1991; Levy-Wilson and Fortier, 1989; Stief et al., 1989; Bode and Mass, 1988]. MARs may also be involved in regulating gene expression [Boulikas, 1994; Forrester et al., 1994].

Recent studies in yeast indicate that a large multi-subunit complex of about 2 MDa, called the SWI-SNF complex, plays a significant role in the expression of many genes by disrupting chromatin structure and facilitating the binding of protein factors [Peterson and Tamkun, 1995]. Counterparts of the SWI-SNF complex may exist in higher eukaryotes.

At another level of chromosomal compaction, these chromosomal loops undergo further folding in the chromosome. The light and dark bands characteristic for each chromosome occur as a result of variations in the final compactness of coiling. It is not known whether this variation reflects functional differences between chromosome regions. In the average chromosome, active chromatin capable of being transcribed, represents about 10% of the total amount. It is characterised by an increased sensitivity to DNAase I which presumably reflects a more open structure than the highly condensed heterochromatin [Lawson et al., 1982].

Because chromatin provides the template from which RNA is transcribed, chromatin structure is an important factor in the regulation of gene expression [Paranjape et al., 1994]. The process is largely controlled through promoter and enhancer elements, which contain a number of cis-acting regulatory sequences capable of binding to regulatory protein factors.

1.5 Promoter Elements and Transcription Initiation

Eukaryotic cells contain three distinct RNA polymerases. RNA polymerase I is located in the nucleolus and synthesises only a rRNA precursor transcript. RNA polymerase III transcribes the tRNA genes and the 5S rRNA gene and is located in the nucleoplasm. Finally, RNA polymerase II transcribes mRNA precursors and hnRNA and so is responsible for the transcription of all genes coding for both native and recombinant protein products [Eick et al., 1994].

Recognition signals in the DNA sequence are required to allow the RNA polymerase II and its associated transcription factors to position themselves accurately on the DNA to initiate transcription. Comparison of the 5' flanking sequences of many DNA polymerase II genes indicates certain common features. A feature known as the TATA box is commonly situated about 20-30 bp upstream from the transcription start site. This element appears to be important for the accurate initiation of transcription. The binding of the TATA binding protein (TBP), also referred to as TFIID, to the TATA box is an important step in the initiation of transcription. Mutation of a single base has a deleterious effect on transcription. Incorporation of the TATA sequence into the nucleosome also inhibits transcription, presumably by affecting the access of trans-acting factors [Imbalzano et al., 1994]. Similar elements are found in prokaryotes at position -10 (10 bp upstream of the start site) and in yeast somewhere between -30 and -90. TATA boxes are commonly flanked by GC-rich regions.

While the TATA box is necessary for transcription it is not usually sufficient in itself. Other promoter sequences are located further upstream

between positions -40 and -110. These sequences are less conserved than the TATA box. At least two such domains, in addition to the TATA box, seem to be required for the initiation of transcription. A common element in many eukaryotic promoters is a region containing the CCAAT motif. GC-rich motifs are also common, particularly in constitutive genes.

RNA polymerase II cannot initiate transcription on its own. Other protein binding factors are required. The diversity in the sequence and position of promoter elements indicates that an interaction with a large complex composed of many different protein factors is necessary for transcription initiation. Many of these factors have been characterised. The TATA box-binding proteins from several species have been extensively studied. In all cases, it is a monomeric protein that binds in the minor groove, undergoing a conformational change as it does so which bends the DNA at that position [Conaway and Conaway, 1993].

A 100 kb protein called Sp1 binds to GC-box elements (Consensus sequence; GGGGCGGGGC) [Mitchell and Tjian, 1989]. This protein contains three zinc-finger DNA-binding regions and two glutamine-rich activation regions, which presumably interact with other protein elements in the initiation complex [Courey et al., 1989; Courey and Tjian, 1988]. Sp1 is important in the function of the Simian Virus 40 (SV40) early promoter which contains six Sp1 binding sites between the TATA box and upstream enhancer elements [Kadonaga et al., 1986]. It is also found in the thymidine kinase promoter from the herpes simplex virus which contains two sites [Jones et al., 1985]. Sp1 can act from distal enhancer sites and interact synergistically with proximal promoter sites. It can also self associate, through its glutamine domains, to bring together well separated DNA domains. This suggests that a DNA looping mechanism may play an important role in enhancer function [Su et al., 1995].

In summary, the promoter region contains a number of binding sites for proteins involved in the initiation of transcription, enabling an active transcriptional assembly to be built up at the transcriptional start site.

1.6 Enhancer Elements and Positive Regulation

Enhancers are a second class of elements that play an important role in the regulation of gene expression. Although the distinction between promoter and enhancer elements is somewhat blurred, there are certain characteristics that typify enhancers.

Elements characterised as enhancers are not promoter-specific and can regulate the activity of any adjacent i.e. cis promoter [Banerji et al., 1981]. They operate in an orientation- and position-independent manner and so can be inverted or moved upstream or downstream of the promoter without affecting activity. Although most enhancer elements are within 500 bp of the promoter, they can be located 10 kb or more away [Pinkert et al., 1987]. Enhancers can also be located within the transcriptional unit - as shown by the immunoglobulin heavy-chain enhancer which is located in the intron between the VDJ exon and the C region exons [Queen and Baltimore, 1983].

Many enhancers confer tissue-specific expression, presumably by acting as binding sites for tissue-specific transcription factors [Tjian and Maniatis, 1994]. Other enhancers modulate expression in response to external stimuli. The glucocorticoid response element (GCE) is recognised by the glucocorticoid receptor/glucocorticoid complex and so gene expression is regulated in response to the glucocorticoid steroid hormones [Karin et al., 1984]. Other enhancers modulate the response to physiological changes such as heat shock [Baumann et al., 1987] .

1.7 Negative Regulation

Whilst most regulatory elements have a positive effect on the expression of the genes they regulate, negative regulation also exists. The presence of actively transcribed tRNA genes upstream of DNA polymerase II promoters has been shown to repress expression [Hull et al., 1994]. However, repression is more controlled by cis-acting regulatory elements.

Elements that can exert a negative effect have been identified in many gene loci [Garzon and Zehner, 1994; Hoyle et al., 1994; May et al., 1994]. There are two types of repressors (negative regulators). Passive repressors decrease the enhancing effect of positive transcription factors, whilst active repressors directly inhibit transcription through their own activity [Cowell, 1994].

Passive transcriptional repression can take one of two forms. Negative transcription factors can compete with positive factors for binding sites within promoter and enhancer elements. Several genes appear to be repressed by the binding of a factor to the GC-box region of the promoter in direct competition with the Sp1 factor [Kageyama and Pastan, 1989]. In addition to negative transcription factors, factors that normally activate gene expression can also repress in certain circumstances. Retinoic acid-induced gene expression is repressed by the transcription factor AP1, which competes for DNA binding sites with the retinoic acid receptor [Schule et al., 1990].

Passive repression can also occur when repressor proteins bind directly to activating factors to form a protein complex with a reduced affinity for DNA. The Id protein can bind to various positively acting helix-loop-helix transcription factors. The resultant hetero-dimers have a reduced affinity for DNA and no positive regulatory effect [Cowell, 1994]. AP1 transcription factors and the glucocorticoid receptor appear to interact to repress

transcription. A protein-protein complex is formed that has no DNA binding ability. The cellular oncogene products c-Jun and c-Fos, which largely make up the AP1 complex, can repress the transcription of glucocorticoid receptor-activated genes. In turn, the glucocorticoid receptor can repress the AP1-activated collagenase gene [Yang Yen et al., 1990]. The formation of a hetero-dimer between transcription factor and repressor may not affect the DNA binding ability of the factor, but instead may prevent transcriptional activation by blocking the protein domain responsible. An example is the interaction of transcription factor GAL4 with repressor GAL80 in *Saccharomyces cerevisiae*.

Other repressor proteins have more direct modes of action than the inhibition of a transcriptional activator. Examples include the Wilms tumour gene product WT1 [Madden et al., 1993] and the human bZIP protein E4BP4 [Cowell et al., 1992]. Like other transcription factors, these repressors seem to contain clearly defined DNA binding and transcriptional repression domains. Domains from different factors can be incorporated into recombinant factors that possess the DNA binding characteristics of one protein and the activity of another. The repressor domains of many factors, including WT1, are rich in alanine, glutamine and/or proline and contain few charged amino acids. Other repressor domains, such as E4BP4 seem to contain many charged side chains. This seems to indicate different classes of repressor protein, based on the structure of the repressor region. Some active repressors may interact directly with general transcription factors such as TFIIB, others may interact with other proteins in the transcription initiation complex.

Several mechanisms have been suggested for active repressors. They may interact with the pre-initiation complex and prevent its formation. This would explain the inhibition of basal transcription caused by some repressors. Alternatively, repressors may interact with DNA-bound

-11-

transcription factors to block activating domains. A third possibility is that the repressor alters chromatin structure to repress transcription. This may involve repositioning nucleosomes over key protein binding sites [Wolffe, 1994] or encouraging a more compact chromatin structure, perhaps through facilitating histone H1 binding.

1.8 Transcription

Transcription initiation is a multistage process requiring the complex interaction of at least five factors and an ATP cofactor. Each stage requires the presence of particular factors and cofactors. It occurs at the promoter region described in Section 1.5, where the required protein binding sites are clustered.

The first stage is the formation of an initial complex, when the TATA-binding element binds to the core region of the promoter. This complex is recognised by RNA polymerase II. The polymerase binds initially to the TATA-binding element, in the presence of at least four other factors, to form the site selected complex. Protein/DNA interactions then extend downstream to the transcription-initiation site to form the inactive pre-initiation complex. This complex, which is activated by an ATP-dependent step, can actively transcribe in the presence of ribonucleoside triphosphates [Conaway and Conaway, 1993] .

Immediately after it is synthesised, the 5' end of the RNA chain is capped with a 7-methyl guanosine residue attached via a 5' to 5' triphosphate linkage. The 2' hydroxyl groups of the initial nucleotide, and occasionally the second nucleotide, are then methylated. The role of the cap appears to be threefold: it increases mRNA stability by protecting against 5' - 3' exonucleases; it facilitates splicing, possibly by providing a recognition signal for the formation of a ribonucleoprotein complex; it is

also a recognition signal for the initiation of protein synthesis by the ribosomes.

1.9 Transcription Termination

The termination of eukaryotic transcription tends to occur well beyond the 3' end of the mature mRNA. Many genes produce an initial transcript that terminates some distance downstream (100 - 4000 bp) of the subsequent site of polyadenylation. It is not known what eventually causes the RNA polymerase II to stop transcribing and detach from the gene. These long initial transcripts are rapidly processed into the normal 3' end of the mRNA [Hawkins, 1991].

There may be 1000 or more base pairs between the termination codon that marks the end of the sequence that is translated into protein and the 3' end of the mRNA. This region contains certain sequences that affect termination. An endonuclease that recognises the sequence AAUAAA, which is highly conserved in eukaryotic mRNAs, cleaves the 3' end of the primary eukaryotic transcript. The context of this sequence element is important in directing cleavage specifically to this single site.

Termination and 3' processing seem to depend on the interaction of trans-acting factors with cis-acting sequence. Correct processing of the 3' end of the mRNA is important to its overall stability, but it is unclear whether this represents a means of regulating gene expression.

Histone genes, which are small and have no introns, possess a highly conserved 23 bp sequence that is important for termination. This sequence contains an inverted repeat capable of forming a stable hairpin-loop structure [Birchmeier et al., 1982]. Also important is an 80 bp region adjacent to the termination site. This contains the conserved sequence CAAGAAAGA [Birchmeier et al., 1983]. It is thought that snRNPs (small

nuclear ribonucleoproteins) are involved in the termination process through interaction with the hairpin-loop or other sequences close to the termination site [Platt, 1986].

1.10 Post Transcriptional Processing

A template-independent poly-A polymerase adds a poly-A tail about 300 bases long to the 3' end of the RNA after transcription. This is important for the stability of the mRNA transcript and the efficiency with which it is translated.

Introns are removed from the RNA sequence through the 'splicing' process. This is dictated by the presence of specific sequences at the splice points and is catalysed by snRNPs. The snRNPs and mRNA precursors (hnRNA) together form the 'spliceosome' complex in which splicing occurs (Padgett et al., 1986).

The hnRNA is associated with nuclear RNPs throughout its time in the nucleus. Upon export to the cytosol, these are replaced by other RNPs, including ribosomes. Almost all RNA processing occurs in the nucleus after which mRNAs are exported to the cytosol where translation occurs. The role of mRNA export from the nucleus, as a point of regulation of gene expression, is unclear. Similarly, the factors that affect mRNA stability in the cytosol are not well understood so its significance as a control of gene expression is difficult to assess. It is clear, however, that mRNAs coding for particular products, such as those involved in the cell cycle, must have short life-spans while others, such as reticulocyte haemoglobin mRNA, are more stable.

1.11 Regulation of Protein Production

The control of eukaryotic translation seems to play a part in the regulation of gene expression. Phosphorylation of ribosomal proteins correlates with increased levels of translation in cells exposed to growth factors. Modification of the initiation factor eIF2 in reticulocytes links translation of mRNA with the availability of haem groups. Other mechanisms regulate the translation of particular mRNAs.

Almost all the stages of protein production, from transcription through translation to localisation and secretion, are subject to regulation. For the biotechnologist intent on the production of recombinant protein they fall into two categories: detailed knowledge of regulation at the DNA level (i.e. transcription) is important because the recombinant gene must be highly expressed; a knowledge-based approach ensures that the most effective regulatory elements, selected with the host system in mind, accompany the gene on the vector.

Other levels of regulation fall into a second category because they are largely dictated by the particular expression host chosen. They are far less amenable to optimisation, since they involve complex and delicate cellular processes. Problems in this area tend to be solved by a change in expression host, which functions largely as a 'black box' once the expression vector has been transformed into it. Hence the emphasis on regulation at the transcriptional level which can be engineered to optimise gene expression.

1.12 Position Effects

A common observation in experiments involving the transfection of mammalian cells, or creation of transgenic mice, is the variability of expression of the inserted gene. Individual transfectants or mouse-lines express the foreign gene to very different levels, despite containing the same expression vector [Al-Shawi et al., 1989]. The number of copies of the vector that have integrated does not dictate the level of expression, as might be expected.

The explanation of this phenomenon lies in the essentially random nature of the integration event. The vector may integrate into the genome at any location. It is the site of integration that is important in determining the level of expression of the recombinant gene. If the vector integrates into a transcriptionally inactive locus, there will be little or no expression of the genes it contains. If it integrates into an active locus, its genes will be highly expressed. This is independent of both the regulatory elements contained on the vector and the vector copy number. This dependence of expression on the site of integration is called the 'position effect'.

The position effect poses problems in transfection experiments. After the transfection of a cell-line, transfectants in which the vector has successfully integrated into the genome can be readily identified by screening for expression of the selectable marker gene contained in the plasmid. This screen takes no account of expression levels, as only minimal levels of marker expression are sufficient for the cell to withstand the selection. A second screen, based on detection of the level of recombinant product, is necessary to distinguish those few transfectants with high level expression from the rest. This screening is costly, time consuming, and often technically difficult [Hudson, 1989; Hudson and

Harrison, 1987]. If the position effect could be overcome it would be of considerable benefit to the biotechnological exploitation of mammalian cells and transgenic animals.

1.13 The β -globin Locus Control Region

The first hint of the presence of regulatory regions capable of overcoming the position effect came from studies of Dutch patients with thalassemias. Thalassemias are a group of disorders arising from impaired globin production. In the case of this particular disorder, both the entire β-globin locus and its known regulatory elements were discovered to be intact [Curtin and Kan, 1987]. These included two tissue-specific enhancers. One was situated 550 - 800 bp downstream of the β-globin polyadenylation site, and the other in the second intron [Kollias et al., 1987].

Patients were found to be heterozygous for a 100 kb deletion upstream of the β-globin locus. The mutant locus was also shown to be in an inactive, condensed chromatin configuration [Driscoll et al., 1989; Kioussis et al., 1983]. This suggested that this locus had lost a key control element located in the deleted upstream region. This element was found to consist of 4 DNAase I hypersensitive sites upstream of the ε-globin gene and one site downstream of the β-globin gene. Whilst low level DNAase I sensitivity is a general characteristic of transcriptionally active regions, DNAase I hypersensitivity is characteristic of regulatory elements. These sites were only present in erythroid tissue when a β-globin gene was being expressed [Tuan et al., 1985]. If murine erythroleukaemia (MEL) cells and human non-erythroid cells are fused to form hybrids, the previously inactive β-globin locus from the human non-erythroid parent is reorganised into an active state, with stable hypersensitive sites forming both upstream and downstream of the β-globin genes [Forrester et al.,

1987]. MEL-cell derived trans-acting factors activate the human β-globin genes and produce distinct, erythroid-specific chromatin structures.

The regions containing these sites were placed on a 38 kb construct with the β-globin gene [Grosveld et al., 1987]. When transgenic mice were made containing this construct, tissue-specific expression was detected that was both position independent, and related to copy number. Expression was at levels comparable to those of the endogenous murine β-globin gene. This 'minilocus' construct, therefore, contained all the sequences necessary to specify position-independent expression of the β-globin locus [Ryan et al., 1989]. These sequences are called the Locus Control Region or LCR.

The activity of the LCR, which resides in the four upstream 5' hypersensitive sites (HS4 to HS1), can be contained in a 6.5 kb microlocus [Talbot et al., 1989] and in a 1 kb construct. Each confers erythroid-specific, position-independent expression on any linked heterologous gene. Expression is usually at levels equivalent to endogenous β-globin levels. These constructs are of considerable value in the construction of expression vectors for MEL host systems [Needham et al., 1992].

The β-globin LCR has been extensively studied [Bresnick and Felsenfield, 1994]. Each hypersensitive site consists of a core region of about 300 bp which contains several protein binding sites. HS2, HS3 and HS4 are all capable, on their own, of directing position-independent expression, albeit at reduced levels [Fraser et al., 1990]. This activity may require the presence of more than one copy of the hypersensitive site [Ellis et al., 1993]. Both the LCR and its constituent hypersensitive sites, with the exception of HS2, produce an effect only when they are stably integrated into the genome. They do not affect expression in transient assays. This distinguishes the LCR from an enhancer and suggests a mechanism that may involve chromatin structural changes.

Each of the hypersensitive sites in the β-globin LCR has been extensively characterised and assessed for its individual activity.

5'HS1, sited 6.1 kb from the ε-gene promoter, contributes little to the enhancer activity of the LCR and seems to play an insignificant part in adult β-globin gene expression.

5'HS 2 is 10.9 kb from the ε–gene promoter. It is active in both stable and transient assays and is responsible for 40 - 50 % of the enhancer activity of the LCR [Fraser et al., 1990; Curtin et al., 1989]. It also seems to be capable of directing the developmental switching of β-like globin genes [Morley et al., 1992]. It contains two Jun/Fos consensus binding sequences that bind the transcription factor NF-E2. This is important for enhancing activity, but not for position independence. While the 3' site contributes to this activity, the 5' site is essential. These sites appear to interact with three downstream elements which bind GATA1, H-BP and J-BP [Talbot and Grosveld, 1991]. Elements that direct position independence are located within a 373 bp core and appear to include duplicate NF-E2 binding sites [Caterina et al., 1991]. Both HS2 and HS3 appear to operate through multimeric protein/DNA complexes. No single binding site appears to be completely responsible for activity [Caterina et al., 1994].

5' HS3 is situated 14.7 kb upstream of the ε-gene promoter. The 300 bp core region of HS3 is capable of conferring position independent expression at about 30 % of the level of the full LCR. It contains binding sites for ubiquitous factors such as SP1 and erythroid specific factors such as GATA1 and NF-EF2 [Strauss and Orkin, 1992; Talbot et al., 1990]. A triple repeat of GATA1 sites and G-rich sequences each about 30 bp apart [Philipsen et al., 1993] appear to contribute to activity.

5'HS4 is 18 kb away from the ε-gene promoter. Activity can be localised to a 280 bp region [Pruzina et al., 1991]. It contains a number of binding sites for erythroid-specific and ubiquitous proteins, including a CCAAT-

like binding motif that appears to bind the factor CP-2. This appears to account for most of the sites activity [Walters et al., 1991].

A fifth site, HS5, seems to have a role as a chromatin insulator, restricting regulatory effects to the β-globin locus [Li and Stamatoyannopoulas, 1994].

1.14 Other LCRs

Since the discovery of the β-globin LCR, elements which confer position-independent expression at high levels have been discovered in other gene loci. While many such elements are associated with hypersensitive sites, they often bear little resemblance to the β-globin LCR and are usually situated close to the gene they regulate. There are several examples of well characterised LCR-like elements in other gene loci.

Two DNAase I hypersensitive sites have been identified in the 3' flanking region of the human CD2 gene, which codes for a T-cell surface glycoprotein. These sites are 0.5 - 1 kb downstream of the 3' end of the gene and are T-cell specific [Greaves et al., 1989]. A fragment containing these sites is capable of conferring high level, position-independent expression on any linked gene. Expression is T-cell specific and so this element has very similar properties to the β-globin LCR, although it is only linked to the expression of a single gene.

The rat WAP (whey acidic protein) can be expressed at high levels in a position-independent manner using a construct that includes the 2.9 kb coding region, 949 bp of promoter sequence and 70 bp of 3' flanking region [Dale et al., 1992]. Replacement of the 3' region with heterologous poly-A sequence results in a sharp decrease in expression levels, although expression remains position-independent. Further deletion of 91 bp of the highly conserved WAP 3' untranslated region renders expression

position-dependent. Position independence can be restored by reinserting the deleted sequence, although it must be in the correct orientation.

Transgenic mice express the human a1(I) collagen gene at high levels in a position-independent manner when a 2.3 kb 5' upstream flanking sequence is present [Slack et al., 1991].

The human major histocompatability complex class I gene HLA-B7 requires 660 bp of 5' upstream sequence to be expressed in an efficient position-independent manner [Chamberlain et al., 1991].

A hypersensitive site 4.8 kb upstream of the PEPcK gene has been shown to confer position-independent, tissue-specific expression at high levels in transgenic mice [Cheyette et al., 1992].

A 4.5 kb construct containing the chicken βA-globin gene with its downstream enhancer is capable of conferring tissue-specific, position-independent expression [Bonifer et al., 1994; Reitman et al., 1990]. Activity is probably derived from a number of elements in the region.

A 2.8 kb region containing two hypersensitive sites 5' of the gene coding for the rat transcription factor LAP-(C/EBPb) confers high-level, position-independent expression on linked genes [Talbot et al., 1994]. Its activity is not confined to the liver and certain constructs are active in other tissues. This is unusual as most LCRs are extremely tissue-specific.

The human α-globin LCR has strong similarities with the β-globin HS2 in both position and structure [Higgs et al., 1990]. However it does not appear to exert any long-range effect on chromatin structure or generate DNAase I hypersensitive sites. This may be because, unlike the β-globin LCR, it is located in a constitutively open chromatin region [Craddock et al., 1995]. Many conserved features are also recognisable in the mouse α-globin LCR [Kielman et al., 1994].

Further LCR-like regions have been identified in chicken lysozyme [Huber et al., 1994; Bonifer et al., 1990], human adenosine deaminase

[Aronow et al., 1992], human keratin [Abe and Oshima, 1990] and murine metallothionein [Palmiter et al., 1993].

1.15 Matrix Attachment Regions

Chromosomes are associated with a proteinous structure known as the nuclear scaffold or matrix. Sequences that seem to mediate the physical attachment to this matrix are known as MARs (Matrix Attachment Regions) or SARs (Scaffold Attachment Regions), depending on the experimental procedure used to isolate them (Mirkovitch et al., 1984, Bowen, 1981).

It seems that such elements play a role in the regulation of gene expression. The presence of SARs has been observed to stimulate expression, albeit without any concurrent position-independent expression [Poljak et al., 1994].

Boundary sequences known as *scs* elements have been identified in *Drosophila melanogaster* [Kellum and Schedl, 1991]. These have been shown to block the activity of an enhancer when situated between enhancer and promoter, suggesting an insulating role for such elements. More evidence for this role is provided by MARs in the apolipoprotein-B domain. These are located between DNAse sensitive and insensitive regions and have been shown to have an insulating effect which protects the sequence between them from the regulatory effects of adjacent chromatin. They also seem to confer positon-independent expression [Kalos and Fournier, 1995]. MARs are often AT-rich and are commonly associated with regulatory elements [Gasser and Laemmli, 1986]. A good example of the apparent activity of MARs is the chicken lysozyme gene.

MARs known as A-elements have been identified at the boundaries of the chicken lysozyme gene, defining the boundaries of an area of general

DNAase 1 sensitivity [Phi-Van and Stratling, 1988]. The 5' and 3' lysozyme A-elements are capable of conferring position-independent, high-level expression when flanking a reporter gene with the lysozyme promoter and enhancer elements [Stief et al., 1989]. An A-element can also repress expression if situated between enhancer and promoter. Although this appears to be an LCR-like activity, it is not associated with DNAase hypersensitivity. Since activity is only apparent after stable integration, these elements may serve to insulate the locus they define from external regulatory influences.

It has recently been shown that the β-globin MAR, located at the 5' boundary of the β globin LCR can, in cis with the polyoma enhancer, direct high-level, position-independent expression. This element is not required for LCR activity and the effect indicates that there may be several mechanisms, including matrix attachment, involved in position independence [Yu et al., 1994].

It is possible that these regions mark boundaries of the looped euchromatin domains which form transcriptional units, insulating them from neighbouring gene loci [Eissenberg and Elgin, 1991]. They may also serve to bring together transcription factors from distant sites through a looping out mechanism to facilitate transcription [Boulikas, 1995]. The precise role of these boundary elements is still vague as they are often difficult to define experimentally and data from experiments using different protocols tend to be contradictory [Cockerill, 1990].

Whilst MARs and SARs are defined experimentally by their attachment to the nuclear scaffold, it is also possible to define boundary elements by a functional assay that measures the ability of an element to insulate a gene from the effects of nearby regulatory elements. In this way, regions that may define chromatin loops as regulatory units can be identified [Li and Stamatoyannopoulas, 1994].

1.16 Recombinant Protein Expression Systems

Recombinant proteins have been expressed in many different host systems ranging from bacteria to microbial eukaryotes, mammalian cells and transgenic animals [Kingsman and Kingsman, 1988].

1.16.1 Bacterial Systems

The most commonly described recombinant expression system employs *Escherichia coli* [San et al., 1994]. This was the original system of choice because it was easy to manipulate and genetically well understood.

E. coli has many advantages as an expression host. Regulatory elements have been well characterised, so expression vectors are relatively easy to design for specific purposes. Inducible elements can be used in a vector to switch on expression at a particular stage in the growth of a culture. This is useful if high-level expression of the recombinant protein has an adverse effect on cell growth. *E coli* can represent a cheap means of producing large quantities of protein since recombinant product can represent up to 25% of the total cellular protein.

There are also problems associated with high-level expression in *E. coli*. Many eukaryotic proteins become denatured and form aggregates with nucleic acids in particles called inclusion bodies [Williams et al., 1982]. These proteins then have to be solubilised and renatured. Purification also poses problems. The reagents commonly used in this process, such as SDS, urea and guanidine HCl are difficult to remove after use. Endogenous *E. coli* proteins, particularly potentially harmful endotoxins, also have to be removed. Purification methods for isolating

recombinant eukaryotic protein from *E. coli* are often time consuming and difficult [Marston, 1986].

More fundamental problems also exist. The disulphide linkages that eukaryotic proteins require for an active conformational state, may not form in the reducing environment in the prokaryotic cell. Endogenous prokaryotic proteases may also digest the foreign recombinant protein. Although fusion to a native protein can prevent the recognition of foreign protein, the extra protein element must be subsequently removed. In addition, prokaryotic proteins are synthesised with a N-terminal, N-formylmethionine residue which may affect the protein activity if it is not removed [Liang et al., 1985].

Although there are significant advantages to the secretion of recombinant protein rather than its accumulation within the cell, *E coli* does not readily secrete protein. The incorporation of signal sequences to direct secretion tends instead to direct the protein to the periplasmic space, where it is prone to precipitation and proteolytic digestion [Chang et al., 1986]. Certain sequences have been shown to be effective in promoting secretion [Hudson, 1989]. Secretion has been observed for eukaryotic proteins fused to the Staphylococcal protein-A signal sequence. Sequence from the haemolysin-A C-terminal is also effective when fused to the recombinant gene. Lysis peptides, active in colicin secretion, can also facilitate the secretion of recombinant protein. Recently, a novel protein secretion factor, isolated from a species of *Vibrio*, has been shown to direct secretion in *E. coli* efficiently [Tokugawa et al., 1994].

In some situations, *E. coli* lacks the necessary biochemical pathways for phenotypic expression of cloned genes. Other bacteria can also be used as expression hosts. For example, *Bacillus subtilis* is capable of high-level secretion of recombinant protein [Old and Primrose, 1985], although

problems with endogenous proteases and genetic instability have limited its use.

1.16.2 Microbial Eukaryotes

Many of the problems of expression in prokaryotic hosts such as *E. coli* can be overcome if eukaryotic hosts are used. The simplest eukaryotes are microbial yeasts and algae, of which the most extensively studied is the yeast *Saccharomyces cerevisiae*, which has been exploited since ancient times in brewing and baking. Many basic cellular processes are similar in both yeasts and higher eukaryotes. *S. cerevisae* can be grown easily in culture to high cell densities and is relatively easy to manipulate genetically [Schultz et al., 1994]. Because it can be cultured in a haploid form, mutant strains can be easily identified. Selectable markers can be used to complement auxotrophic mutants. When mutants cannot be isolated, such as with industrial polyploid strains, dominant selectable markers such as thymidine kinase can be used [McNeil and Friesen, 1981].

Vectors in *S. cerevisae* can be integrative or autonomous, depending on whether replication is autonomous or not. The former, such as YIp1 [Struhl et al., 1979] rely on homologous recombination to integrate into the genome. The latter may contain combinations of yeast chromosomal sequences such as ARS (autonomous replicating sequences), CEN (centromeric sequences) and TEL (telomeric sequences). These sequences can also be used to make Yeast Artificial Chromosome (YAC) vectors capable of bearing inserts of up to 1000 kb [Burke et al., 1987]. Most expression vectors are based on the endogenous yeast plasmid, the 2 μm circle.

Although some heterologous eukaryotic regulatory elements are active in yeast, the expression of recombinant genes normally requires the addition of yeast-specific elements. Efficient promoters are commonly derived from glycolytic enzymes. Since long RNA transcripts are unstable, 3' non-coding sequences containing transcription termination elements are frequently incorporated.

Continuous high level production of recombinant protein can decrease the growth rate of cells in culture or have toxic effects. Inducible promoters are commonly used. The GAL10 promoter can be induced 1000-fold by the addition of galactose and is commonly used in industrial processes.

S. cerevisae can accurately modify translated proteins through phosphorylation and fatty acylation. Glycosylation occurs efficiently, but differs from that in higher eukaryotes in that only mannose residues are added to the core oligosaccaride [Staneloni and Leloir, 1982]. Secretion of some recombinant proteins (eg lysozyme) can be achieved using a foreign signal-peptide sequence. This is frequently inefficient. Improvements have been made using signal sequences from *S. cerevisae* genes such as invertase and α-factor [Chang et al., 1986; Brake et al., 1984].

Pichia pastoris is a methylotrophic yeast that is widely used to express recombinant proteins. Expression is usually controlled by the methanol induced promoter *AOX1*. This controls the expression of the endogenous alcohol oxidase enzyme and is capable of generating high levels of recombinant gene expression.

Other microbial eukaryotes that have been studied with regard to genetic manpulation and possible use in expression systems include: the yeast *Schizosaccharomyces pombe*; the filamentous fungi *Neurospora crassa* and *Aspergillus nidulans*; the slime mould *Dictyostelium discoideum*; and the green alga *Chlamydomonas reinhardi.*

Another commonly used expression system employs baculovirus-based vectors to drive the expression of recombinant proteins in insect cells. These are easier and cheaper to use than mammalian cell systems but provide many of the same post-translational modifications, to generate recombinant proteins that are functionally and immunogenically similar to the native protein.

1.16.3 Mammalian Cells

For production of authentic recombinant mammalian protein, a mammalian cell based expression system has many advantages [Lopez et al., 1994; Werner, 1994]. Such systems are either transient or stable. Transient systems, such as the COS system (CV-1, origin, SV40 cells), express genes for a fixed time period after transfection, after which the cells are harvested and the recombinant products analysed [Gluzman, 1981]. In stable systems, the expression vector stably integrates into the cell genome and gene expression occurs throughout the life of the culture [Hudson, 1989].

Cultures of mammalian cells are difficult and time consuming to set up and maintain [Freshney, 1983]. Immortalised cells in continuous culture display many of the characteristics of tumour cells. Genetic instability and ploidal variations are common.

Immortalised cell lines can be derived from various types of cell and at particular stages in the differentiation process. For example, cell lines have been created from each stage of β lymphocyte differentiation [Mather et al., 1984]. Conditions in the culture have a strong influence on the properties of the cells. For instance, while manyt cell lines require a solid support on which to grow, others, such as lymphoblastoid cells, will grow free in suspension.

The transfection of vectors into mammalian cell lines can be achieved in a number of ways [Kingsman and Kingsman, 1988]. DEAE dextran is commonly used for transient transfection experiments. Other techniques employ physical means (electroporation, protoplast fusion, and microinjection) or chemical treatments (calcium phosphate and dimethylsulfoxide/polycations). Transfections can be selected using complementation markers such as thymidine kinase, or dominant markers such as the *neo* gene that confers resistance to the antibiotic G418.

Plasmid vectors commonly contain regulatory elements from viruses such as SV40 and Rous Sarcoma Virus (RSV) [Southern and Berg, 1982]. Viruses themselves are also used as vectors and different types have been successfully employed. Integrative vectors include papovaviruses such as SV40 and retroviruses such as MoMLV. Other viral vectors are stably maintained extrachromosomally as autonomous episomes eg papillomaviruses, Epstein-Barr virus and BK virus.

1.16.4 Transgenic Animals

Cultured cells lack many of the features of a living organism. To study many aspects of gene expression it is necessary to manipulate whole organisms. The mouse, and the fruit fly *Drosophila melanogaster* are the most commonly studied organisms.

Transgenic mice are the best model for mammalian gene expression. They are created by the microinjection of vector DNA into the male pronucleus of fertilised eggs. These are transplanted into a pseudopregnant female to yield live offspring. Offspring are identified in which the vector has integrated successfully into the genome. Lines of transgenic mice can be maintained for as many generations as is required.

Expression of inserted genes can be studied at various developmental stages in a variety of tissues.

Transgenic animals can also be used for the efficient production of recombinant protein [Gillman, 1994]. They are created using vectors that contain the gene encoding the desired protein under the control of lactation-specific regulatory elements. The protein is produced in the milk and can be harvested and purified. Commercial production systems are being developed using sheep, goats and cows, producing therapeutic proteins such as alpha-1-antitrypsin, tissue plasminogen activator and antithrombin III. Production costs are considerably cheaper than for cell based systems. However, there is currently little commercial pressure to adopt transgenic systems because production costs represent only 20% of the selling price of a protein.

1.17 Myeloma Cells

Myeloma cells are immortalised cells derived from terminally differentiated B-cells called plasma cells which are responsible for the secretion of immunoglobulin. They possess several features that are potentially useful in an expression host [Traunecker et al., 1991].

Myeloma cells secrete immunoglobulin at high levels. Up to 20% of the total cell protein synthesised is secreted as immunoglobulin [Laskov and Scharff, 1970]. Secretion is constitutive and occurs independent of stimulus. There is no intracellular accumulation of product since it is rapidly secreted after synthesis. The highly developed secretory apparatus required for this process could be used to produce recombinant protein. The secretion of protein would not then limit the amount of product available. In addition, since virtually all secreted myeloma protein is

immunoglobulin, recombinant protein secreted into the medium should be relatively pure and contamination-free, facilitating purification.

Myeloma cells can grow freely in suspension and are relatively easy to culture. Techniques for the large-scale culture of monoclonal antibody-producing hybridoma cells are well established and suitable for myelomas. Extensive genetic analysis of the immunglobulin loci has been carried out and the mechanisms of immunoglobulin expression are well understood. This information may be useful in designing expression vectors for use in myeloma [Wendle and Buckel, 1987]. Thus myeloma cells may be useful expression hosts but their potential remains to be developed.

1.18 Project Background

Expression vectors for use in myeloma cells have been developed in the laboratory over a number of years [Hudson, 1989; Hudson and Harrison, 1987]. A series of plasmid vectors derived from the pSV plasmid were designed using various combinations of regulatory elements and polyadenylation signals, and containing the chicken lysozyme cDNA [Harrison et al., 1995]. These were used to transfect mouse-derived J558L cells [Oi et al., 1983].

From these transfection experiments, two particular transfectants were identified. These were designated C6 and D8 and were shown to produce 50 - 100 times more lysozyme than other transfectants. Expression was on an equivalent level, in molar terms, to that of endogenous immunoglobulin.

Further genetic analysis [Glassford, 1993] indicated unique integration sites in both cases and a copy number of one for C6 and two for D8. This expression was considered to be due to the influence of an endogenous regulatory element at the site of vector integration. This element was

supposed to be comparable to the β-globin LCR and would be useful to isolate and include in an expression vector. Efforts were made to clone regions flanking the vector in the transfectant genome. A range of approaches were attempted: including plasmid library construction, plasmid rescue [Perucho et al., 1980]; inverse PCR [Ochman et al., 1988] and ligation mediated anchor PCR [Loh, 1991]. Though none were successful, it was considered that the latter had shown some potential. The current project focussed initially on this approach and considerable progress was made in developing this technique, as described in Chapter 3. Subsequent work, described in Chapter 4, concerned the construction and screening of λ–bacteriophage genomic libraries of the D8 cell line.

LCR-like elements have been identified in a number of genes in different species and seem to represent a level of regulation that involves chromatin structural alterations and chromosomal domains. With the exception of the β-globin LCR itself [Needham et al., 1992], LCRs have not been exploited in expression systems. The transfectant lines C6 and D8 provide an approach to the isolation of a highly active regulatory element that may have applications in the development of myeloma based expression systems.

Chapter II

Materials and Methods

2.1 General

The water used throughout this work was treated with the Milli-Q purification system. Water used in DNA manipulation work was further autoclaved at 121°C and 1.06 bar for 25 mins. Water used for PCR was autoclaved and stored separately from other reagents.

All solutions, glassware and plastics used in DNA manipulations, bacterial and myeloma cell work were either supplied sterile or autoclaved for 25 mins. PCR reagents were stored separately. Solutions were generally manipulated using Gilson pipettes.

Latex gloves were worn during DNA manipulation experiments to avoid contamination with nucleases.

The methods used have been described elsewhere [Sambrook et al., 1989] except where stated.

All myeloma cell lines were maintained in the laboratory by Sarah Munson and Sarah Chapman. They were grown in suspension in high glucose (4.5 g/l), Dulbecco's Modified Eagle medium, penicillin (100 units/ml) and streptomycin (100 µg/ml) and 10% (v/v) horse serum. Selection was done with XHMPA (xanthine, 250 mg/ml; hypoxanthine, 0.1 mM; and mycophenolic acid 6 mg/ml). Cells were cultured in incubators at 37°C in a humidified atmosphere of 10% CO_2 in air.

All oligonucleotide primers were made in the Leicester University Biochemisry Department by Debra Langton. An Applied Biosystems 380B oligonucleotide synthesiser was used, which utilised phosphoramidite elements with a controlled-pore glass support, following Applied Biosystems protocols.

The following suppliers were used for laboratory reagents. All reagents used were of analytical grade unless stated otherwise. Manufacturers' instructions were followed where appropriate.

Amersham International plc., Amersham, England.

BDH Chemicals Ltd., Poole, England.

Bio-Rad Laboratories Ltd., Hemel Hempstead, England.

Boehringer Corporation Ltd., Lewes, England.

Cambio Ltd., Cambridge, England.

Fisons Ltd., Loughborough, England.

Gibco-BRL plc., Paisley, Scotland.

Pharmacia LKB Biotechnology AB, Uppsala, Sweden.

Promega Corporation, Madison,Wisconsin, USA.

Stratagene Ltd., Cambridge, England.

Unipath Ltd., Basingstoke, England.

United States Biochemical Corporation, Cleveland, Ohio, USA.

2.2 Bacteria

2.2.1 Bacterial Strains

The following *E. coli* strains were used in this work:

<u>Strain</u>	<u>Genotype</u>	<u>Reference</u>
TG1	*supE hsdΔ5 thiΔ(lac-proAB)* F′[*traD36 proAB⁺ lacI*q *lacZΔM15*]	[Gibson, 1984]
MC1061	*hsdR mcrB araD139 Δ(araABC-leu)7679 DlacX74 galU galK rpsL thi*	[Meissner et al., 1987]
DH10B	F⁻ *araD139 Δ(ara, leu)7679 ΔlacX74 galU galK rpsL deoR f80dlacZΔM15 endA1 nupG recA1 mcrA Δ(mrr hsdRMS mcrBC)*	[Grant et al., 1990]
SRB	*e14-(mcrA), Δ(mcrCB-hsdSMR-mrr)171, sbcC recJ, uurC, umvC,::Tn5(kan r), mpE44, lac, gyrA96 relA1, thi-1, endA1[F′proAB lacI*q *ZΔM15]*	[Elgin et al., 1991]
TG2	*SupE, hsdD5thiΔ(lac-proAB)Δ(sr1-recA)306::Tn10 (tet r), F′[traD36proAB+lacIq], laczDM15Tn10(tet* r*)*	[Gibson, 1984]
DH5	*SupE44,hsdk17,recA1,endA1,gyrA96,thi-1,relA1*	[Hanahan, 1983]
C600	*F-, thi-1, thr-1, leuB6, lacY1, tonA21, supE44*	[Jendrisak et al., 1987]

2.2.2 Bacterial Culture

Cells were routinely grown up in Luria-Bertani medium (LB):

Bacto-tryptone	1% (w/v)
Yeast Extract	0.5% (w/v)
NaCl	1% (w/v)

Solid medium for agar plates was made up as above, but with the addition of 2% (w/v) agar.

The *E. coli* SRB (Section 2.2.1) cells used in λ-phage cloning experiments were cultured with NZCYM medium:

NZ Amine	1% (w/v)
NaCl	0.5% (w/v)
Yeast extract	0.5% (w/v)
Casamino acids	0.1% (w/v)
$MgSO_4.7H_2O$	0.2% (w/v)

Medium for agar plates also contained 2% (w/v) agar.

A bacterial medium called SOB was also used in transformation experiments;

Tryptone	2% (w/v)
Yeast extract	0.5% (w/v)
NaCl	0.05% (w/v)

After autoclaving, KCl was added to 2.5 mM and $MgCl_2$ to 10mM.

For selective bacterial growth appropriate antibiotics were added from stock solutions;

Antibiotic	Storage	Stock Conc	Final Conc
Ampicillin	Water -20°C	50 mg/ml	50 µg/ml
Chlor-amphenicol	Ethanol -20°C	34 mg/ml	34 µg/ml
Kanamycin	Water -20°C	10 mg/ml	10 µg/ml
Tetracyclin	Ethanol -20°C	5 mg/ml	5 µg/ml

Liquid cultures were grown in an orbital shaking incubator at 37°C. Agar plates were grown inverted in an incubator at 37°C .

2.3 Preparation of Cells Competent for Plasmid Uptake

Host strain bacterial cells were grown overnight at 37°C on LB agar plates without selection. A liquid SOB culture was then inoculated using a discrete colony and shaken overnight at 37°C. Cells were then subcultured in warm SOB in the ratio 1/100 v/v. 50ml of this culture was grown for 2 - 4 hr at 37°C until the optical density was in the range A_{520} 0.3 - 0.5. It was then centrifuged for 5 mins at 4000 rpm at 4°C and the supernatant decanted and discarded. The pellet was resuspended in 40 ml of transformation buffer I (30 mM KOAc; 100 mM RbCl; 10 mM $CaCl_2$; 50 mM $MnCl_2$; 15% v/v glycerol; adjusted to pH 5.8 with 0.2 M CH_3COOH) and left on ice for 5 mins.

Cells were then centrifuged for 8 mins at 3000 rpm at 4°C and the supernatant discarded. The pellet was resuspended in 4 ml transformation

buffer 2 (10 mM MOPS (3-[N-morpholino]propanesulfonic acid); 75 mM $CaCl_2$; 10 mM RbCl; 15% (v/v) glycerol; adjusted to pH 6.5 with 1 M KOH). Aliquots of 50 μl were dispensed in 1.5 ml microfuge tubes and flash frozen in dry ice prior to storage at -70°C.

2.4 Transformation of Competent Cells

The plasmid or ligation product to be used was kept on ice for 5 mins prior to transformation. An aliquot of competent cells was taken from the -70°C freezer, thawed at 37°C and kept on ice for 5 mins. DNA was then added to the cells and carefully mixed in with a Gilson pipette. The cells were left on ice for 30 mins then placed in a 42°C waterbath for 2 mins. 1 ml of LB medium was then added and the cells kept at 37°C for 60 mins. Cells were then split and plated out on selective plates as required. Transformation efficiencies, for closed circular pUC based plasmid, of around 10^6 cfu/μg were typically obtained.

2.5 Storage of Bacterial Strains

A single bacterial colony was grown overnight in 10 ml of LB medium plus antibiotic. Cells were then spun at 4000 rpm for 5 mins and the supernatant discarded. The cells were then resuspended in 1 ml of 10 mM $MgSO_4$ and spun at 3000 rpm for 8 mins. The supernatant was removed and cells resuspended in 1 ml of 50% (v/v) glycerol, 50% (v/v) 10 mM $MgSO_4$. The suspension was transferred to a sterile bijou bottle and stored at -70°C.

2.6 Myeloma Cell Lines

All the myeloma cell lines used were derived from the cell line J558L [Oi, 1983]. This is a spontaneous heavy-chain-loss-variant myeloma cell line made from the J558 cell line [Lundblad, 1972]. It synthesises and secretes only immunoglobulin light-chain.

The cell line C6 was transfected with the plasmid pLysSV40.gpt. The cell line D8 is a J558L cell line that was transfected with the plasmid pLysCMV.gpt (see Figs 3.1 and 3.2 for plasmid maps) [Harrison, 1995; Hudson, 1989; Hudson 1987].

2.7 Minipreparation of Plasmid DNA

An overnight culture of plasmid-bearing cells in liquid LB medium was grown up on an orbital shaker at 37°C. 1.5 ml of cells were removed, spun in a microcentrifuge, and resuspended in 100 μl GTE (50 mM glucose; 10 mM EDTA; 100 mM Tris-Cl pH 8). After 5 mins at room temperature, 200 μl of 0.2 M NaOH/1% (w/v) SDS was added and the tube gently shaken. The tube was put on to ice for 5 mins then 150 μl of 3 M K, 5 M Ac was added and the tube put on ice for another 5 mins. The contents of the tube were then centrifuged in a microcentrifuge and the supernatant transferred to another tube. This crude DNA extract was then phenol extracted and ethanol precipitated as described in Section 2.9.

Plasmid DNA was also purified using the Promega 'Magic' minipreps system.

2.8 Midipreparation of Plasmid DNA

Cells containing plasmid were grown overnight in 100 ml of liquid LB medium plus appropriate antibiotic selection at 37°C. The cells were then spun at 4000 rpm for 10 mins and resuspended in 4ml GTE. After 5 mins on ice, 8 ml of 0.2 M NaOH/1% (w/v) SDS was added and the cells returned to ice for 5 mins. 6 ml of 3 M K, 5 M Ac was added and the tube placed on ice for a further 10 mins.

The tube containing the cells was then spun at 4000 rpm for 15 mins and the supernatant filtered through glass wool into a fresh tube. 17 ml of propan-2-ol was added and the tube cooled to -70°C for 15 mins. After this, the tube was spun at 4000 rpm for 15 mins and the supernatant discarded. The pellet was resuspended in 2 ml of TE (10 mM Tris-Cl pH 8; 1 mM EDTA pH 8), to which 2.5 ml of LiCl was added. The tube was cooled to 0°C for 60 mins then spun at 4000 rpm for 15 mins.

The supernatant was transferred to a fresh tube, to which was added 10 ml ethanol. After cooling to -70°C for 10 mins, the tube was spun at 4000 rpm for 15 mins and the supernatant removed. The pellet was washed with 70% (v/v) ethanol and dried in a vacuum dessicator. The pellet was then resuspended in 400 μl of TE to which was added 10 μl RNAase (10mg/ml) and 20 μl 10% (w/v) SDS. After heating to 70°C for 10 mins, the plasmid DNA was phenol extracted and ethanol precipitated (Section 2.9).

2.9 Phenol/Chloroform Extraction

To purify a crude DNA extract from contaminating proteins, an equal volume of neutralised phenol was added to the DNA solution and the two solutions mixed by vortexing. The tube was then centrifuged to separate the aqueous and phenol layers. The aqueous layer (top) was removed to a fresh tube and the process repeated. When the aqueous phase had been removed for a second time, an equal volume of chloroform : iso-amyl-alcohol (24 : 1) was added to remove the last traces of phenol.

2.10 Ethanol Precipitation

To precipitate DNA from solution in order to purify or concentrate it, 2 - 3 volumes of ethanol were added, along with 0.1 volume 0.3 M NaOAc pH 5.2. The solution was then cooled to -20°C for 60 mins and spun for 10 mins at 13 000 rpm in a benchtop centrifuge. The pellet was washed in 70% (v/v) ethanol and resuspended in a suitable volume of TE for future manipulations.

2.11 Quantitation of DNA

DNA in solution was quantitated in two ways. Either a sample was run in an agarose gel alongside tracks containing known amounts of DNA and the relative band densities compared, or its value was obtained spectrophotometrically. A sample of DNA was placed in a quartz curvette and its absorbance at 260 nm (A_{260}) measured in a spectrophotometer. The amount of DNA was then calculated, since 50 μg/ml double stranded (ds)

DNA has an A_{260} of 1. An assessment of the purity of the DNA was obtained by looking at the A_{260}/A_{280} ratio, which should be 1.8 for a pure dsDNA solution. Samples whose absorbance varied markedly from this were assumed to contain contaminating proteins.

2.12 Isolation of Myeloma Genomic DNA

Approximately 1×10^8 cells were centrifuged at 1200 rpm for 3 mins and then washed in 5 ml of PBS (8.0 g NaCl; 0.2 g KCl; 1.44 g Na_2HPO_4; 0.24 g KH_2PO_4 per litre of water pH 7.4) followed by a further wash with 10 ml PBS. The cells were repelleted and resuspended in 9 ml of Proteinase K solution (0.4 M Tris-HCl pH 7.8; 0.1 M EDTA; 500 µg/ml Proteinase K). 1 ml of 10% (w/v) SDS was added after resuspension, and the tube left at 37^oC until the lysate became clear.

An equal volume of phenol was added and the mixture gently agitated for 3 hrs. The phases were separated at 1200 rpm for 3 mins, the phenol phase removed and another volume of phenol added. The two phases were mixed and centrifuged. An equal volume of chloroform : isoamyl alcohol (24 : 1) was added to the aqueous phase, mixed and the two phases separated by centrifuging. The aqueous phase was removed and placed into a universal tube. 0.2 volumes of 10 M NH_4OAc were added, then 2 volumes of ethanol. The sample was gently mixed, then centifuged at 5000 rpm for 5 mins at room temperature. The pellet was washed in 70% ethanol (v/v) then dried in air. The sample was not completely dried out, as this would have made resuspension difficult. The pelllet was resuspended in 4 ml TE and rotated mechanically overnight to ensure the DNA had dissolved.

The concentration of the DNA was determined using a spectrophotometer and the ratio of $A_{260}/280$ determined to check for

protein or RNA contamination, with a ratio of >1.9 being considered acceptable. From 1×10^8 cells, between 1 and 2 mg of DNA was isolated.

The extent of shearing of the DNA was checked by running an aliquot of the DNA on a 1% (w/v) agarose gel.

Genomic DNA was also prepared using the Scotlab 'Nucleon Extraction' genomic DNA preparation kit.

2.13 General DNA Manipulations

2.13.1 Restriction Endonuclease Digests

Type II restriction endonucleases are enzymes that cut dsDNA at specific recognition sites determined by the surrounding sequence. Restriction digests of plasmid and genomic DNA were carried out in a 1 x Reaction Buffer supplied with each enzyme with the addition of 0.1 mg/ml BSA (Bovine Serum Albumin). Length of incubation varied according to the sample, ranging from 1 - 2 hrs for plasmid DNA, to overnight genomic digests. Reactions were incubated at 37°C except where the enzyme suppliers instructions stated otherwise.

2.13.2 Enzyme Inactivation

Some experiments required restriction endonuclease to be inactivated after digestion. Three methods were commonly employed to do this. If the enzyme was heat sensitive, the digest was heated to 65°C for 20 mins or 70°C for 10 mins. If the enzyme was not heat sensitive, the digest was either subjected to phenol/chloroform extraction followed by ethanol precipitation, or put down a 'Magic Clean Up' column.

-43-

2.13.3 Ligation

In order to ligate two pieces of DNA with compatible 'sticky ends', the two were mixed with an insert : vector ratio of about 3 : 1. The solution was heated to 45°C, cooled on ice, then BSA was added to 0.1 mg/ml, ATP to 1mM and ligation buffer to 1 x (50 mM Tris-Cl pH 7.6; 10mM MgCl$_2$; 1mM DTT; 5% PEG-8000 (w/v)). T4 DNA ligase was added (0.1 U) and the reaction left overnight at 16°C. Controls for this reaction were plasmid and insert on their own.

2.13.4 Blunt Ending

For some cloning purposes it was desirable to blunt the overhangs left by restriction endonuclease digestion. This was done by adding a dNTP (dATP, dGTP, dCTP, dTTP) solution, each to 50 mM, then adding 1 U of Klenow DNA polymerase per μg of DNA. This was then left at room temperature for 15 mins before inactivation by heating to 75°C for 10 mins or by phenol extraction.

2.13.5 Calf Intestinal Phosphatase

Calf Intestinal Phosphatase was used to remove 5' phosphate groups from pieces of DNA. This is useful because it prevents self-ligation during cloning reactions. To a DNA solution were added BSA to 0.1 mg/ml and buffer to 1 x (1 mM ZnCl$_2$; 1 mM MgCl$_2$; 10 mM Tris-Cl pH 8.3). This reaction was incubated at 37°C for approximately 3 hrs. The enzyme was then inactivated either by heating or by phenol extraction.

2.13.6 T4 Polynucleotide Kinase

Since the oligonucleotides used to prime PCR reactions lacked 5'- phosphate groups, T4 polynucleotide kinase was used to catalyse the transfer of the γ-phosphate of ATP to the 5'-termini of the oligonucleotides in order to facilitate cloning.

10 x T4 polynucleotide kinase buffer (0.5 M Tris-HCl, pH 7.6; 0.1 M $MgCl_2$; 50 mM DTT; 1 mM spermidine; 1 mM EDTA pH 8.0) was diluted to 1 x in a solution of 100 to 500 ng of DNA, along with ATP to 1 mM and 1 μl (10 units/μl) of T4 polynucleotide kinase. The reaction was incubated at 37°C for 1 hr and the enzyme inactivated by heating to 70°C for 10 mins.

2.13.7 Proteolytic Enzymes

Two proteolytic enzymes were used to remove contaminating proteins from a DNA solution. These were Pronase E and Proteinase K.

Pronase E is a mixture of serine and acid proteases from *Streptomyces griseus*. It was made up to a concentration of 20 mg/ml in 10 mM Tris-HCl pH 7.5, 10 mM NaCl, and self digested for 1 hr at 37°C in order to remove any DNAase or RNAase contaminants. The self-digested enzyme was then stored in small aliquots at -20°C. Pronase E was used at a concentration of 1 mg/ml in a reaction buffer composed of 0.01 M Tris-HCl pH 7.8, 0.01 M EDTA, 0.5% (w/v) SDS. The reaction was performed at 37°C for 1 hr.

Proteinase K is a highly active protease of the subtilisin type from the mold *Tritirachium album* Limber. Stock solutions were made up to a concentration of 20 mg/ml in H_2O. The enzyme was used at a concentration of 50 μg/ml in a reaction buffer of 0.01 M Tris-HCl pH 7.8,

0.005 M EDTA, 0.5% (w/v) SDS. The reaction was carried out at 37°C for 1 hr.

After proteolytic digestion with either of these enzymes, the DNA solution was phenol extracted to remove any traces of protease prior to subsequent manipulations.

2.13.8 Spun Column Chromatography

Spun column chromatography was used to separate dsDNA from unincorporated dNTPs and oligonucleotides, or from short sequences of DNA generated when a plasmid polylinker was cut at two restriction sites. Low-molecular weight substances are retained by the column while the large dsDNA molecules pass through.

The bottom of a 1 ml disposable syringe was plugged with a small amount of sterile glass wool and the syringe filled with sterilised Sephadex G-50 equilibrated in 1 x TE buffer. The buffer was allowed to flow through the syringe and more resin was added until the syringe was completely full. The syringe was inserted into a 15 ml disposable plastic tube and centrifuged at 2000 rpm for 4 mins at room temperature in a benchtop centrifuge in order to pack down the resin. More resin was added and the syringe re-centrifuged until the volume of the packed column was approximately 0.9 ml and remained unchanged after centrifugation.

100 μl of 1 x TE buffer was added to the column and the syringe centrifuged again. This step was repeated until the volume of buffer being eluted from the column was the same as that applied. The DNA sample was added to the top of the column in a total volume of 100 μl (in 1 x TE buffer) and the syringe centrifuged as before. The effluent from the bottom of the syringe was collected in a de-capped microcentrifuge tube placed inside a 15 ml tube.

An alternative method used to purify DNA from oligonucleotides, proteins and short DNA sequences was to use Chroma+TE spin columns. These were used according to the manufacturers instructions.

2.14 Agarose Gel Electrophoresis

2.14.1 Analytical Gel Electrophoresis

Electrophoresis separates molecules of DNA according to size. The molecular weight of the DNA molecule is proportional to the log of the distance moved.

In order to examine DNA of less than 25 kb, the sample was run in a 1% agarose gel in 1 x E buffer (40 mM Tris-HCl pH 8.0; 30 mM acetic acid; 1 mM EDTA). The gel was prepared by melting the agarose in a microwave and pouring it into a mould made by wrapping tape around a perspex template. Before the gel set, a comb was inserted, to form wells into which the sample could be loaded when the gel had set. When set, the comb was removed from the gel and transferred from the mould to a purpose built perspex electrophoresis tank filled with 1 x E buffer, so that the buffer just covered the gel.

The DNA sample to be run was resuspended in an appropriate volume, according to well size, usually about 10µl. To this was added 2 µl 6 x Loading Buffer (0.25% (w/v) bromophenol blue; 0.25% (w/v) xylene cyanol; (v/v) 30% glycerol). The sample was then pipetted into a well of the gel. Samples were commonly run at 100 v for 1 hr, although the exact conditions varied according to the gel and the resolution required.

At the end of the run the gel was removed from the electrophoresis tank and placed in a staining solution composed of 0.5 mg/ml ethidium

-47-

bromide in 1 x E buffer. After about 30 mins of staining, the gel was visualised by placing it on a UV transilluminator (312 nm). Ethidium bromide fluoresces, and since it binds to DNA, bands can be seen.

Photographs of the gel were then taken using Kodak T_{max} film. Sizes of DNA bands could be established by running a marker track of DNA fragments of known sizes alongside the sample track. Commonly used were λ bacteriophage DNA cut with Hind III (7 DNA bands of sizes; 23.1, 9.4, 6.6, 4.4, 2.3, 2.0, and 0.56 kb) and 1 kb marker. (18 DNA bands of sizes; 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.6, 1.0, 0.5, 0.4 and 0.3, 0.2 and 0.1 kb).

2.14.2 Pulse Field Gel Electrophoresis

Standard electrophoresis does not resolve DNA fragments of greater than about 25 kb. To resolve fragments of this size, and to improve the resolution of smaller fragments, pulse field electrophoresis on a CHEF apparatus was used. Gels were typically 1% (w/v) agarose in 0.5 x TBE (0.045 M Tris-borate, 0.001 M EDTA). DNA was either in solution, or when dealing with high molecular weight fragments, embedded in agarose plugs. The electrophoresis parameters varied according to the size of DNA and resolution required. For DNA of 25 kb or below, 200 v for 5 hrs was used with a pulse time that ramped from 1 sec to 5 secs over the duration of the run.

2.14.3 Preparative Gel Electrophoresis

It was often necessary to isolate and purify a particular piece of DNA prior to cloning. The DNA was run out on an agarose gel in 1 x E buffer as normal and visualised by staining in ethidium bromide. The band of

interest was then cut out using a clean scalpel and placed in a sterile eppendorf. Two methods were used to extract the DNA from the agarose:

2.14.3.1 Gelase

This method required that the gel be composed of low melting point agarose. The gel slice containing the DNA was weighed and then soaked in three volumes of 1 x buffer (40mM bis-Tris pH 6.0; 40 mM NaCl) for 1 hr. The excess buffer was then removed and the gel slice melted completely in a $70^\circ C$ water bath. This is important to ensure that all the microcrystalline bundles of helical coils in the agarose are dissociated so the enzyme has access to all bonds.

The gel slice was equilibrated to $45^\circ C$ and 1 unit of gelase enzyme was added for every 600 mg of gel. This was incubated at $45^\circ C$ for 1 hr. DNA was obtained by ethanol precipitation at room temperature using 5 M NH_4OAc instead of 3 M NaOAc, to prevent the possible co-precipitation of oligosaccharide digestion products.

2.14.3.2 Qiaex kit

For extraction using the Qiaex kit, any type of agarose could be used. The band in the gel was excised, soaked in 300 µl of QX1 per 100 mg gel. 10 µl of the Qiaex resin was added, mixed and the tube incubated at $50^\circ C$ for 10 mins. The tube was then centrifuged and the resin pellet washed twice in 500 µl QX2 and twice in 500 µl QX3. The pellet was then air dried. DNA was obtained by adding TE heated to $50^\circ C$ to the resin, vortexing, spinning down and transferring the supernatant to a clean tube. This was repeated twice for maximum DNA yield.

-49-

2.15 Screening Bacterial Colonies

In the cloning process it is necessary to identify those colonies that, after transformation, contain the DNA of interest.

Initial transformants are plated onto selective medium, so only cells containing plasmid will grow. This does not distinguish those cells containing a plasmid with a cloned piece of DNA, from those containing a self-ligated plasmid with no insert.

Two methods were used, together or individually, to identify transformant colonies containing sequences of interest, α-complementation and colony hybridisation:

2.15.1 α-Complementation

This method enables colonies containing plasmids with any sort of insert to be readily identified. The plasmid must contain a lacZ or lacZ α-peptide gene. This is a short segment of *E. coli* sequence that contains the regulatory sequences and the coding information for the first 146 amino acids of the β-galactosidase gene (lacZ). This sequence contains a polylinker, so although an active gene product is normally produced, any cloned insertion inactivates the lacZ gene.

These vectors are used with host cells which code for the carboxy-terminal portion of β-galactosidase. Both parts of the enzyme are, by themselves, inactive but, when produced in the same organism, they associate to form an active β-galactosidase.

Transformants are plated on medium containing X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase), in the presence of the non-metabolisable lactose analogue IPTG (isopropylthio-β-D-galactoside),

-50-

which induces the lac operon. Those colonies that can form an active β-galactosidase through α-complementation will change colour as the induced β-galactosidase metabolises the X-Gal into a blue dye. Those colonies whose α-peptide has been insertionally inactivated by a cloned insert will not form any active β-galactosidase and so the colonies will not change colour. Colonies containing cloned insert are therefore white instead of blue and can be readily identified and picked off.

40 μl of a stock solution of X-Gal (20 μg/ml in dimethylformamide) and 4 μl of a solution of IPTG (200 μg/ml) made up to a final volume of 150 μl with LB medium, were added to a pre-made selective LB agar plate. The solution was spread over the entire surface of the plate using a sterile glass spreader and the plate incubated at 37°C until dry. The plates were inoculated with transformed bacteria of an appropriate strain, inverted and incubated at 37°C for 16 hrs.

2.16 DNA Hybridisation

2.16.1 Southern Blotting

In order to analyse DNA, it was frequently necessary to attach it to a membrane support and hybridise it with radiolabelled probe. Nylon membranes were commonly used for this purpose.

Southern blotting [Southern, 1977] of DNA from agarose gels onto nylon Hybond-N filters under standard conditions, or positively charged nylon Hybond-N+ filters under alkaline conditions, was performed both by capillary action or by using vacuum blotter (Pharmacia Vacugene apparatus).

In order to transfer DNA onto Hybond-N by capillary action, the gel was visualised in ethidium bromide and photographed. If the DNA of

interest was over 20 kb, the gel was depurinated in 0.25 M HCl for 15 mins, to facilitate the transfer of large fragments. The gel was then soaked in denaturing solution (3 M NaCl; 5 M NaOH) for 30 mins and then in neutralising solution (3 M NaCl; 0.5 M Tris-HCl pH 7.4) for 30 mins.

A blot was then set up by placing the gel on a wick soaked in transfer buffer. On this was placed the membrane followed by several layers of 3MM chromatography paper and finally a large wad of paper towels. A small weight was placed on top to ensure the layers maintained contact. This was left on the bench overnight. The transfer buffer used was 20 x SSC (3 M NaCl; 0.3 M Na citrate pH 7.0).

To blot DNA onto Hybond-N by vacuum blotting, the apparatus, gel and membrane were set up according to the manufacturers instructions. Denaturing solution was applied twice to the surface of the gel and the vacuum turned on until the denaturing solution had been drawn through the gel. Neutralising solution was applied twice to the surface of the gel and drawn through under vacuum.

The transfer buffer (20 x SSC) was applied to the surface 4 - 6 times over approximately 30 mins. If the DNA fragments to be transferred were greater than 30 kb then the transfer buffer was applied for approximately 2 hrs. The vacuum was applied at 50 mBar for all steps. For transfer onto Hybond-N$^+$, no denaturing step or neutralising step was required. 0.4 M NaOH was used both to denature the DNA and as the transfer buffer.

After transfer onto Hybond-N the filters were air dried, wrapped in Saran Wrap and the DNA fixed onto the filter by placing on a UV transilluminator for the empirically determined optimum time of 15 secs.

No UV fixation is required for Hybond-N$^+$ and therefore these filters were simply air dried before proceeding further.

2.16. 2 Colony Hybridisation

2.16.2.1 Directly from Agar Plates

When a large number of bacterial colonies were to be screened, colonies were transferred directly from the surface of the agar plate onto either 82 mm or 138 mm Hybond-N circular colony lift filters.

Colonies were grown on selective LB-agar plates overnight. The plates were then placed at $4^{O}C$ for at least 30 mins in order to harden the agar and thus improve the efficiency of the lifts. A Hybond-N filter was placed carefully onto the surface of the agar plate. The filter and underlying medium were then marked in three asymmetric positions using a 18-gauge needle dipped in ink and punched through the filter. This allowed the filter to be correctly orientated on the plate later. Using a pair of tweezers the filters were gently peeled off the plate. The filters were then either treated to lyse the colonies and fix the DNA to the filter, or used to prepare a second replica.

To make a second replica, the first filter was placed colony side up onto the surface of a fresh LB-agar plate and a second numbered filter placed on top of it. The two filters were aligned using the needle marks made earlier. The two filters were incubated at $37^{O}C$ for several hours and then kept together during the subsequent lysis and neutralisation steps, only being peeled apart before the final wash.

The original master agar plate was incubated at $37^{O}C$ for several hours in order to re-generate the colonies, sealed with parafilm and stored at $4^{O}C$ in an inverted position.

2.16.2.2 Small Scale Colony Lifts

This method was employed when a small number of colonies (between 25 and 400) were to be screened. A gridded 'master' Hybond-N filter was numbered using a pencil and placed onto the surface of a fresh selective LB-agar plate, grid-side up. The colonies to be screened were picked individually, using sterile toothpicks, from the original plates and streaked onto the 'master' filter using the grid as an alignment indicator.

A negative and a positive control were also streaked onto the 'master' filter and their positions marked using a pencil. A second filter was then placed on top of the 'master' filter and a 18-gauge needle used to mark both filters in three asymmetric positions. The second filter was peeled off and lysed immediately. The master filter on its agar plate was also incubated at 37ºC for several hours in order to regenerate the colonies.

If another replica filter was required from the master plate, transfer onto this filter was not always very effective. To remedy this, the filter was placed, colonies-side up, onto a fresh LB-agar plate and incubated at 37ºC for a few hours to allow the colonies to grow before lysis.

2.16.2.3. Colony Lysis and Binding of DNA to Filters

Filters were transferred to a plastic tray containing a sheet of 3MM paper soaked in denaturing solution (3 M NaCl; 0.5 M NaOH). Filters were left on the paper for 6 mins. They were then removed from the denaturing solution and placed on a sheet of 3MM paper soaked in neutralising solution (3 M NaCl; 0.5 M Tris-HCl pH 7.4) for 3 mins. They were then placed then on top of a second sheet of filter paper soaked in fresh neutralising solution for a further 3 mins. After this the filters were placed

in a solution of 2 x SSC (0.3 M NaCl; 0.03 M Na citrate pH 7.0) and the cell debris gently removed from the filters.

The filters were sandwiched between two pieces of 3MM paper, dried at room temperature, wrapped in Saran Wrap and UV-fixed before undergoing hybridisation. They were then washed and subjected to autoradiography. Positive colonies could be aligned to the master plates using the asymmetric needle marks.

2.16.3. Hybridisation

Radiolabelled probes were hybridised to DNA immobilised on nylon filters in specially constructed perspex hybridisation chambers with screw-down lids. During incubations, these chambers were fully immersed in a shaking water bath.

The membrane to which the DNA was bound was incubated for 1 - 6 hrs in pre-hybridisation solution [3 x SSC; 5 x Denhardt's solution; 1% (w/v) SDS; 6% (w/v) polyethylene glycol (PEG) 6000; 0.5% (w/v) dried milk powder (Marvel)]. Note, 5 x Denhardt's solution consists of 0.12% (w/v) ficoll type-400; 0.12% (w/v) polyvinylpyrrolidone 40; 0.12% (w/v) BSA.)

The temperature of the incubation was 65°C if a random oligonucleotide primed probe was used, or 45°C if a kinase end-labelled oligonucleotide probe was used. The DNA probe was added to the pre-hybridisation solution and hybridisations were carried out, for 16 - 72 hrs. At the end of the hybridisation, probe that had not bound specifically was removed by washing.

The initial wash was in 3 x SSC, 0.1% SDS (w/v) at room temperature for 10 mins. The stringency could then be increased by lowering the SSC

concentration and/or raising the temperature of the wash. Usually for a random oligonucleotide primed, labelled probe the final washing conditions were 0.1 x SSC, 0.1% (w/v) SDS at 65°C for 5 mins. For a smaller kinase end-labelled probe the final wash conditions were usually 1.0 x SSC, 0.1% (w/v) SDS at room temperature for 5 mins. Excess washing solution was removed before autoradiography but the membrane was never allowed to dry completely.

2.16.4 Construction of Random Oligonucleotide-Primed Probes

Random oligonucleotide-primed probes were synthesised from a dsDNA template of 500 bp or larger.

2.4 µl (25 µg/µl) of random sequence hexamer oligonucleotide (dNNNNNN), 2 µl (5 ng/µl) of template DNA and 8.6 µl of water were incubated in boiling water in a 1.5 ml microcentrifuge tube for 10 mins to denature the DNA. The tube was then placed on ice-water to prevent re-naturation.

1 µl of 1mg/ml BSA, 2.0 µl of 10 x oligonucleotide labelling buffer (250 mM Tris-HCl pH 8.0; 25 mM MgCl$_2$; 5 mM β-mercaptoethanol; 2 mM each of dATP, dCTP, dGTP, dTTP; 1 mM Hepes), 2.0 µl of α[^{32}P]dCTP (0.74 MBq with a specific activity of 110 TBq/mmol) and 2.0 µl of the large fragment of DNA polymerase (Klenov) were then added.

The reaction was thoroughly mixed and then incubated at room temperature for at least 1 hr. The reaction was stopped by the addition of 80 µl of Stop Buffer (50 mM Tris-HCl, pH 7.5; 50 mM NaCl; 5 mM EDTA, pH 8.0; 0.5% (w/v) SDS). The probe was incubated at 100°C for 10 mins then cooled on ice, to denature the DNA, and then added to the pre-hybridisation buffer.

2.16.5 Kinase End-Labelled Oligonucleotide Probes

Oligonucleotide probes were made by adding a radioactive phosphate group to the 5' end of an oligonucletide using T4 polynucleotide kinase. 5 µl of 10 x kinase buffer (0.5 M Tris-Cl pH 7.6; 0.1 M MgCl$_2$; 50 mM DTT; 1 mM spermidine; 1 mM EDTA pH 8.0), 33.6 µl of water, 5 µl of $\gamma[^{32}P]dATP$ (1.85 MBq with a specific activity of 110 TBq/mmol) and 1.4 µl of T4 kinase (10 units/µl) were added to 5 µl of oligonucleotide (10 µM) in a 1.5 ml microcentrifuge tube. This reaction mixture was incubated at 37°C for 45 mins before the T4 kinase was inactivated by heating to 68°C for 10 mins. The probe was then ready for use.

If a purer probe was required, the radiolabelled oligonucleotide was subjected to ethanol precipitation (See Section 2.10).

2.16.6 Quantitation of the Specific Activity of DNA Probes

The specific activity of each probe was determined by the differential absorption of the reaction products on a positively charged surface i.e. DE-81 paper. DE-81 filters are positively charged and so will strongly bind negatively charged nucleic acids, including oligonucleotides. Unincorporated radionucleotides do not bind to the filters and are removed by washing the filters extensively in sodium phosphate.

For random oligonucleotide-primed probes, 2 µl of the probe solution was spotted onto each of four labelled DE-81 filters, after the Stop Buffer had been added to the priming reaction. For kinase end-labelled

oligonucleotides, 1 μl of the probe solution was added to each of four filters after the reaction had been stopped.

Two of the filters were washed three times in 0.5 M Na_2HPO_4 for 10 mins each. These filters were rinsed in water and then in 95% (v/v) ethanol before drying in air.

All four filters were then inserted into glass scintillation vials and the amount of radioactivity on each filter counted by Cerenkov counting (in the ^3H channel of a liquid scintillation counter). The amount of radioactivity on the unwashed filter was compared to the amount on the washed filter, and the proportion of radionucleotide incorporated in probe was calculated as follows:

i)

$$\frac{\text{cpm in washed filter}}{\text{cpm in unwashed filter}} = \text{proportion cpm incorporated}$$

ii)

proportion cpm incorporated x total counts added = counts incorporated

iii)

$$\frac{\text{counts incorporated}}{\text{total mass of product}} = \text{specific activity}$$

Typical specific activities for random primed probes are around 5 MBq/μg.

2.16.7 Autoradiography

After hybridisation and washing, results were visualised by autoradiography. Filters to be autoradiographed were wrapped in Saran wrap and attached to a backing of 3MM paper. The 3MM paper was marked with radioactive ink. This allowed the position of bands or

positive colonies on the autoradiographs to be correlated back to their position on the membrane.

Membranes were exposed to Kodak XAR-5 or Fuji RX pre-flashed X-ray film at -70°C using X-ograph intensifying screens for 6 - 72 hrs. The films were then developed and studied.

2.16.8 Removal of Probes from Filters

To remove probes from DNA blots so that they could be re-probed, the membranes were immersed in a boiling solution of 0.1% (w/v) SDS which was then allowed to cool to room temperature over 30 - 60 mins. Before reprobing, the efficiency of stripping was checked by autoradiographing the membranes to ensure no signal was left. The filters were then pre-hybridised and re-probed as described above.

2.17 Polymerase Chain Reaction (PCR)

2.17.1 General PCR Amplifications

All plastic ware used for Polymerase Chain Reaction experiments was prepared wearing gloves then autoclaved to reduce DNA contamination. Aerosol resistant pipette tips were used which contained filters to reduce DNA carryover from one reaction to the next. In addition, general precautions, such as the separate storage of PCR reagents, were carried out to prevent contamination. The reagents used were molecular biology grade wherever possible. All reactions were carried out using a Perkin-Elmer Cetus PCR Thermal Cycler.

Reaction conditions for PCR reactions, unless otherwise stated in the text were: 1 x polymerase buffer (50 mM KCl; 1.5 mM $MgCl_2$; 0.1 mg/ml gelatin; 10 mM Tris-HCl, pH 8.0); 0.2 mM each of 'Ultrapure' dNTPs, dATP, dCTP, dGTP and dTTP; 0.8 µM of both primers, and 2.5 units of taq polymerase. All reactions were carried out in a final volume of 10, 50 or 100 µl which was made up with sterile H_2O.

For amplification of a DNA fragment from a plasmid, between 0.01 and 1 ng of DNA was usually used as a template, and for amplifying from mammalian genomic DNA between 100 ng and 1.0 µg of DNA was typically used as a template. The reactions were overlaid with 20 µl of PCR mineral oil to prevent evaporation.

The PCR reaction cycling conditions varied with each reaction. A denaturing step at 95°C from 30 secs up to 1.5 mins was used. This was followed by an annealing step at primer melting temperature (T_m) minus 5°C for between 1 and 2 mins.

The T_m (melting temperature) was calculated approximately on the basis of the GC content of the primers used [(Number of dGTPs and dCTPS) x 4°C + (Number of dATPs and dTTPs) x 2°C]. The annealing step was followed by an extension step at 72°C. The time allowed for this extension was at least 1 min and was lengthened on the basis:

[length of fragment to be amplified (bp)/ 1000 (bp/minute)] + 30 secs.

'Touchdown' PCR involves the use of progressively lower annealing temperatures with each round of PCR [Don, 1991]. The initial annealing temperature was about 2°C above the T_m of the primers and decreased by 1°C each round until about 5°C below the T_m. After this touchdown, 25 cycles were carried out.

A further refinement was to employ a 'Hot Start'. This involved a preliminary denaturing step of 5 mins at 95°C for all the reaction ingredients except the taq polymerase, which was added and the cycling started, before the reaction could cool down. This prevented spurious products being produced in the first round by non-specific, low temperature annealing before the cycling steps.

PCR on λ-plaques was carried out in the same way as for DNA. Since the temperature of the denaturation steps was sufficient to disrupt the phage particle, the DNA within was susceptible to PCR, so a phage suspension could be treated in much the same way as a DNA solution.

2.17.2 Preparation of Oligonucleotides and Primers for PCR Reactions

After synthesis, oligonucleotide primers were ethanol precipitated and then resuspended in sterile water. The oligonucleotides were then diluted to a stock concentration of 10 µM determined by spectrophotometry. An absorbance of 0.024 (A_{260}) with a 1 cm path length was taken to represent a concentration of 1.0 µg/ml and a concentration of 78 µg/ml was assumed to be approximately equal to a concentration of 10 µM for an oligonucleotide 24 bases long (24 mer). All handling of the oligonucleotides to be used for PCR was carried out under conditions designed to minimise contamination with DNA.

2.17.3 Preparation of DNA Probes by PCR

Many DNA probes were made by PCR using plasmid DNA as a template. The PCR was performed in the usual manner, and then 100 µl of

reaction mixture was put down a Chroma+TE spun-column to remove unincorporated dNTPs and oligonucleotides.

The concentration of the DNA was estimated by running an aliquot on an agarose gel and comparing the intensity of the ethidium bromide staining of the sample to that of a λ-Hind III standard. The probe DNA was then diluted to give a concentration of approximately 5 ng/μl. A radiolabelled probe could then be generated from this template by random priming.

2.17.4 Preparation of Template DNA for Ligation-Mediated-Anchor PCR

Template DNA for Ligation-Mediated-Anchor PCR was prepared following the method of Loh [Loh, 1991].

5 μg of C6 DNA was digested with the restriction enzyme Bgl II. Then an equal volume of 1 x restriction buffer containing 400 mM dGTP (to fill in one base of the GATC overhang created by Bgl II), and 10 units of Klenow polymerase were added and the reaction incubated at 37°C for 60 mins. The reaction was then heated to 70°C for 10 mins (to denature the Klenow fragment), phenol-chloroform extracted and ethanol precipitated. The DNA was dissolved in 50 μl of TE-buffer and a 10 μl (1 μg) aliquot was used in the ligation.

8 μl of anchor primer and coupling primer (10 μM; see Appendix II.4) and 14 μl of H$_2$O were added to 1 μg of the Bgl II digested, filled-in C6 DNA. This was heated to 65°C for 5 mins to disrupt primer/DNA interactions, then cooled to room temperature for 5 mins to facilitate the interaction of anchor and coupling primer and the generation of an overhang that can be ligated.

8 μl of 5 x ligation buffer and 1 unit of T4 DNA ligase were added, to bring the final reaction volume to 40 μl, and the reaction incubated at 15°C overnight.

8 μl of the ligation mix was used directly for amplification without removal of primers. For other experiments, primers were removed using column chromatography prior to use.

2.17.5 Cloning PCR Products

Cloning PCR products has proved to be less than straightforward in many cases. In this study, two methods were routinely used to clone PCR products.

PCR primers can be designed to contain restriction sites at their 5' ends. A PCR product was amplified using conventional primers initially (Section 2.17.1). This product was then purified using 'Magic' PCR preps, and subjected to a second round of PCR, of only 5 - 10 cycles, using primers containing restriction sites. These primers contained enough complementary sequence to anneal as normal. The products thus produced, were purified using 'Magic' PCR preps, and subjected to restriction endonuclease digestion to expose sticky ends. The PCR product could then be cloned as normal.

A second approach exploited the tendency of taq polymerase to add a template-independent adenine residue onto the 3' end of PCR products. A vector was used that contained 5' thymine overhangs (pGEM-T), effectively a sticky end with respect to PCR products. Cloning could then take place as normal.

2.18 λ-Library Construction

2.18.1 Preparation of DNA for Library Construction

A prime consideration in the construction of a library is the potential number of clones to be screened. This depends on the number of target sequences present, the total size of the genome and the amount of insert the plasmid can carry (Clarke and Carbon, 1976):

$$N \text{ (No. clones)} = \frac{\ln (1\text{-}p)}{\ln (1\text{-}f)}$$

p = probability of success, f = ratio of insert size : genome size

(See Section 4.2)

To minimise the potential number of clones, a partial purification of D8 genomic DNA was carried out. Typically, 150 μg of D8 DNA was digested to completion with the restriction enzyme Nco I. This, according to restriction maps already established [Glassford, 1993], left the region of interest on a 10 kb restriction fragment.

The total restriction products were run out on a 1% Low Melting Temperature agarose gel alongside marker DNA of known band sizes. The gel was stained with ethidium bromide and the region of the genomic digest corresponding to 10 kb, according to the markers, cut out with a sterile scalpel. This gel slice was then treated with gelase to liberate the DNA.

The DNA was stored in TE until shortly before the ligation was to be carried out. It was then digested with Eco RI to liberate an internal 4 kb fragment. The enzyme was then heat inactivated and the products put into a ligation reaction with commercially supplied bacteriophage λgt10 arms predigested with Eco RI.

It was discovered that storage of Eco RI fragments, even in TE at -20°C, reduced the efficiency of the ligation and the titre of the subsequent library. A similar decrease in efficiency was noted when the Eco RI fragment was subjected to agarose gel purification (See Section 5.2).

2.18.2 Ligation of DNA and λgt10 Vector Arms

For pilot ligations, a range of insert DNA concentrations was added to 0.5 μl (0.25μg) of Eco RI cut λgt10 arms pretreated with calf intestinal phosphatase, 1μl 10 x ligase buffer (300 mM Tris-HCl; 100 mM MgCl$_2$; 100 mM DTT; 10 mM ATP), and 10 - 15 Weiss units of T4 DNA ligase in a total volume of 10 μl. Alongside this, a negative control containing no genomic DNA, and a positive control containing a known amount of vector DNA were set up. Ligations were left overnight at 16°C.

For library construction, the optimal figures for the amounts of insert and vector, ascertained in the pilot reactions, were then multiplied five fold, maintaining the same ratio, and the ligation reaction carried out in the smallest convenient volume (usually 10 μl).

2.18.3 Packaging of λgt10 Ligation Products

After ligation, the DNA was in a long concatameric form, with insert sequence sandwiched between λ arms. This was then packaged into virus particles, which were used to infect host cells. Commercial packaging extracts are available and the Packagene (Promega) system was used here. This consists of a phage infected *E. coli* C-extract which contains a mixture of proteins and precursors capable of encapsidating λ DNA.

The Packagene extract was stored at -70°C. When ready for use, it was thawed on ice. For pilot studies, 10 μl of extract was added to 2 μl of

ligation products. During library construction, 10 µl of ligation product were added to the entire 50 µl of Packagene extract.

The tube was gently mixed and incubated at room temperature for 3 hours. SM buffer (2% (w/v) gelatin; 20 mM Tris-HCl; 100 mM NaCl; 10 mM MgSO4) was then added to the packaging reaction; 445 µl was added to the full-scale reaction along with 25 µl of chloroform, whilst 90 µl and 5 µl were added to the pilot reactions. The tube was then mixed by inversion and stored at 4°C for up to 3 weeks. During this time there was a slight drop in titre.

For long-term storage, DMSO was added to 7% (v/v) final concentration and the phage particles stored at -70°C.

2.18.4 Preparation of Host Cells for λgt10 Infection

A suitable host strain, in this case *E. coli* SRB, because of its recombination deficient properties, was grown up overnight in 50 ml of NZCYM nutrient medium + 0.2% (w/v) maltose. The cells were spun down at 4000 rpm for 10 mins at room temperature. They were then resuspended in 0.01 M MgSO4 and incubated at 37°C for 60 mins in an orbital shaker. The cells were then spun down again, resuspended in 0.01 M MgSO4 and stored at 4°C for up to 1 week.

2.18.5 Infection and Plating of Host Cells

Serial dilutions of the packaged λ-phage were made in SM buffer and 0.1 ml of each dilution were added to 1.5 ml microcentrifuge tubes. 0.1 ml of prepared host cells were added and the tube gently mixed. Cells were then incubated at 37°C for 20 mins. The infected cells were then added to 3 ml of molten 0.7% (w/v) agar NZCYM medium kept at 45°C. The

medium was then vortexed and poured onto an NZCYM agar plate. This was allowed to stand for 5 mins at room temperature while the top agar layer set. The plates were then incubated at 37°C for about 8 hrs, until plaques appeared.

2.18.6 Screening of a λ Library

After a library had been constructed, samples were plated out (See 2.18.5) to ascertain the titre i.e. the number of viable plaque forming units in the library. If this was deemed to be high enough for the library to have a high probability of containing the sequence of interest, the library was screened. Phage were plated out as described in Section 2.18.5 at sufficient concentration that indivdual plaques were barely discernable. 152 mm diameter plates were used, to facilitate the screening of large numbers of plaques. About 20 000 plaques were screened on each 152 mm plate.

When plaques were just visible on the plates, they were stored at 4°C for 60 mins to harden the top agar. A circular gridded filter (Hybond-N) was placed on the surface of the agar and left at 4°C for 5 mins. The filter and agar were then marked by puncturing the filter at three asymetric locations, with a needle soaked in ink. This enabled the filter and subsequent autoradiographs to be aligned correctly on the plate to identify plaques which hybridised to the probe.

The filter was then lifted carefully off the plate and placed on a piece of 3MM paper soaked in denaturing solution (3 M NaCl; 0.5 M NaOH) for 10 mins. Meanwhile a duplicate plaque lift was performed. The filter was lifted off the denaturing solution and the excess fluid gently wiped off. It was then placed on a sheet of 3MM paper soaked in 1 M Tris-HCl pH 7.5 for 1 min and then another soaked in 1 M Tris-HCl, 1.5 M NaCl for 10 mins. The filter was then washed in 2 x SSC for 2 mins and air dryed for 1

hr. The DNA was fixed by placing the filter, DNA-side down, on a UV transilluminator (312 nm) for 45 secs. It was then prehybridised and probed (Section 2.16.3).

Positive plaques that lit up on the autoradiographs of both the filter and the duplicate were identified on the original agar plate by aligning the asymmetric needle marks. These were removed from the plate by encircling the plaque in a 1 ml Gilson tip, cut off to create a wider bore, and sucking up both plaque and surrounding agar into the tip. It was then ejected into 0.5 ml SM buffer to which 25 µl of chlorform was added. This was then replated (Section 2.18.5) at a lower plaque density, and the process repeated until all the plaques on a plate were positive. At this point, either a large scale λ-preparation was performed, or the crude semi-purified phage in the SM buffer containing the agar plug, were subjected to PCR to amplify the cloned sequence in the λ-vector.

2.18.7 Plaque Purification

Having identified and purified a phage particle containing a sequence of interest, it was then necessary to grow it up in large amounts so that the DNA could be extracted and the sequence cloned into a vector more amenable to genetic manipulation. An alternative to this would be to amplify the interesting sequence by PCR and clone the PCR product.

A 100 ml culture of NZCYM was inoculated with *E.coli* SRB host and grown overnight in an orbital shaker. The OD_{600} of the culture was measured and the cell concentration calculated assuming:

$$1 \text{ unit } OD_{600} = 8 \times 10^8 \text{ cells/ml}$$

Four aliquots of cells containing 10^{10} cells were removed, spun at 4000 rpm for 10 mins and the supernatant discarded. Each cell pellet was resuspended in 3 ml of SM buffer and phage added and rapidly mixed.

About 5 x 10^7 pfu of phage were added, although the exact figure was determined empirically.

The cells were incubated at 37°C for 20 mins with gentle shaking and then each infected aliquot was added to 500 ml of prewarmed NZCYM at 37°C. The flask was incubated at 37°C in an orbital shaker. After about 8 hrs the flasks were examined for signs of lysis. When lysis was observed, 10 ml of chloroform were added to each flask and the flask incubated for a further 10 mins.

The lysed cultures were then cooled to room temperature and DNAase I and RNAase added to final concentrations of 200 μg/ml and 100 μg/ml respectively. After incubating at room temperature for 30 mins, solid NaCl was added to a final concentration of 1 M, dissolved by swirling and left on ice for 1 hr. Cell debris was then removed by centrifugation at 11 000 rpm for 10 mins at 4°C and all the samples pooled in a fresh flask. Solid PEG 8000 was added to 10% (w/v) and dissolved by slow stirring. The flask was then cooled and stood on ice for 1 hr. Precipitated phage particles were pelleted by centrifugation at 11 000 rpm for 10 mins at 4°C and as much of the supernatant removed as possible. The pellet was then gently resuspended in SM buffer.

An equal volume of chloroform was then added, the tube vortexed, centrifuged and the aqueous layer removed, to separate phage from cell debris and PEG 8000. 0.5 g of solid CsCl was added per ml of phage suspension and the solution carefully layered onto a preformed CsCl step gradient made by layering CsCl solutions of different densities onto one another in a clear polypropylene centrifuge tube. The position of the interface between the 1.5 g/ml and the 1.45 g/ml layers was marked as this was where the phage particles would be found. The tube was centrifuged at 22 000 rpm for 2 hrs at 4°C and the phage collected by puncturing the side of the tube with a hyperdermic needle at the appropriate point on the

gradient.The phage was then allowed to run out through the needle into a fresh tube. The suspension was then placed in an ultracentrifuge tube and covered with 1.5 g/ml CsCl. This was spun at 38 000 rpm for 24 hrs at 4°C. The phage band was again collected and stored at 4°C in CsCl solution.

2.19 DNA Sequencing

The sequencing of cloned DNA was achieved using the dideoxy-mediated chain-termination method first developed by Sanger [Sanger et al., 1977] using $\alpha[^{35}S]$dATP as described by Biggin [Biggin et al., 1983].

2.19.1. Preparation of Plasmid DNA for Sequencing

'Magic-Miniprep' kits were used to isolate and purify plasmid DNA for sequencing.

Bacterial cells containing the plasmid of interest were grown overnight in 5 ml of nutrient medium containing antibiotic. 4.5 ml of this culture was pelleted in a microcentrifuge and the pellets combined and resuspended in a total volume of 200 μl cell resuspension solution (50 mM Tris-HCl pH 7.5; 10 mM EDTA; 100 mg/ml RNase A). 200 μl of cell lysis solution (0.2 M NaOH; 1% (w/v) SDS) was added and the tube inverted to mix. 200 μl of neutralisation solution (2.55 M KOAc pH 4.8) was added and mixed by inverting the tube. The tube was spun at 13 000 rpm for 5 mins and the supernatant decanted into a new microcentrifuge tube.

1 ml of Magic-Miniprep DNA Purification Resin was added to the supernatant and mixed by inverting the tube. A Magic-Miniprep mini-column was attached to a 2 ml disposable syringe barrel and the Magic-Miniprep DNA Purification Resin containing the bound DNA pipetted into the syringe barrel. The syringe plunger was then used to gently push the slurry into the mini-column. The syringe was then removed from the column and filled with 2 ml of Column Wash Solution (200 mM NaCl; 20 mM Tris-HCl pH 7.5; 5 mM EDTA) which was gently pushed through the column using the syringe plunger. The mini-column was transferred to a microcentrifuge tube and centrifuged for 20 secs to remove all traces of wash solution and dry the resin. The column was transferred to a new microcentrifuge tube and the DNA eluted by applying 50 µl of TE buffer, preheated to 70°C, to the column and centrifuging for 20 secs at 14 000 rpm.

2.19.2. Dideoxy-Mediated Chain-Termination

The first stage in the sequencing reaction is to denature the DNA to be sequenced. 3 µl of 5 M NaOH and 3 µl of 5 mM EDTA were added to 50 µl of plasmid DNA (Section 2.19.1) and the tube incubated at 37°C for 30 mins. The solution was neutralised by adding 5.5 µl of 3 M NaOAc (pH 5.6) and the DNA precipitated with ethanol and resuspended in 7 µl of sterile water. The reagents used in the following sequencing reaction were from the 'Sequenase' kit.

1 µl of primer oligonucleotide (10 ng/µl) and 2 µl of 5 x annealing buffer (200 mM Tris-Cl pH 7.5; 100 mM MgCl$_2$; 250 mM NaCl) were added to the template DNA. The tube was then incubated at 65°C for 2 mins and then at 37°C for 30 mins. 1 µl of DTT (0.1 M), 2 µl of Labelling Nucleotide

Mixture (0.15 mM dGTP; 0.15 mM dCTP; 0.15 mM dTTP), 1 μl of α[35S] dATP (0.37 MBq with a specific activity of 110 TBq/mmol) and 2 μl of Sequenase enzyme were added and the reaction incubated at room temperature for 5 mins.

3.5 μl of the above reaction was then added to each of four pre-warmed (37°C) tubes containing 2.5 μl of one of each termination mix (with the relevant ddNTP at 8 mM and the dNTPs at 80 mM). These reactions were incubated at 37°C for 5 mins. The reactions were terminated by the addition of 4 μl of Stop Solution (95% (v/v) formamide; 20 mM EDTA; 0.05% (w/v) bromophenol blue; 0.05% (w/v) xylene cyanol). 4 μl of each sample was heated to 80°C for 5 mins and loaded onto the sequencing gel.

2.19.3. Electrophoresis of Sequencing Products on Polyacrylamide Gels

Dideoxy-sequencing reaction products were resolved by vertical electrophoresis through a 6% (w/v) acrylamide gel. A 40% (w/v) acrylamide gel mixture was made with 215 g urea, 25 ml of 10 x TBE buffer and 75 ml of acrylamide (40% (w/v) acrylamide; 19 acrylamide : 1 bis-acrylamide) in a final volume of 500 ml. To make a gel, 60 ml of this gel mixture was mixed with 360 μl of 10% (w/v) TEMED (N, N, N', N'-tetramethylethylenediamine) and 15 μl of ammonium persulphate. After the catalysts were added, the solution was pipetted between two glass plates (20 x 35 cm).

Sequencing was performed using Bio-rad sequencing apparatus. The plates, one of which was siliconised with dichloromethylsilane, were separated with 0.4 mm spacers and clamped together. The solution was added along one side of the plates (which had been inclined at an angle of approximately 30° to the horizontal) with care not to introduce air

bubbles. Sharks-teeth combs were inserted (smooth edge first) and the gel allowed to set for at least 1 hr. The plates were then washed and moved to a vertical electrophoresis tank containing 0.5 x TBE buffer. The comb was removed from the gel and re-inserted with the teeth touching the gel, to form the loading wells.

Immediately before loading, the wells were washed with buffer to remove any concentrated urea (which may have leached from the gel) or any unpolymerised acrylamide. The samples were loaded into the wells in the order G, A, T, C and the gel run at 2000 v, 25 mA.

If only the first 100 - 200 bp of sequence was to be read then the gel was run until the bromophenol blue in the dye had run to about 4 cm from the bottom of the gel. If more sequence was to be read then the bromophenol blue was run off the bottom of the gel and after a further 30 mins a second set of the same sequencing reactions were added to four more wells and the gel run until the bromophenol blue in these lanes was about 4 cm from the bottom.

Following electrophoresis the gel plates were removed from the tank, the clamps removed and the plates prised apart, leaving the gel adhered to the non-siliconised plate. The gel was 'fixed' by immersing it in fixing solution (10% (v/v) glacial acetic acid, 10% (v/v) methanol) for 10 mins. The gel was then removed from the fixing solution, excess solution drained off and the gel transferred to a sheet of 3MM paper by laying the paper on top of the gel and gently peeling the gel from the plate. The gel was covered then with a sheet of Saran Wrap and dried for 2 hrs at 80°C in a Bio-Rad model 583 slab gel-dryer and autoradiographed for 16 - 76 hrs at room temperature.

2.19.4 Cycle Sequencing

Cycle sequencing was used to sequence directly from λ-plaques, using a PCR based amplification procedure. Several commercial kits were tried: fmol, CircumVent, and Sequitherm. The latter was found to be most effective although the resolution was not as good as with conventional sequencing procedures.

Sequencing primers were end labelled as described previously using γ-[^{32}P] ATP. Protocols were similar for all kits used, with that of Sequitherm described below.

The following were mixed in a 0.5 ml microcentrifuge tube: 3μl radiolabelled primer (0.5 pmol/μl); 2.5 μl 10 x sequencing buffer; 5 - 50 fmol DNA template; 1 μl Sequitherm polymerase. The volume was made up to 16 μl with water. Four tubes containing 2 μl of termination mix (G, A, T and C respectively) were cooled on ice and 4 μl of reaction mix added to each. Each tube was overlaid with 10 μl of mineral oil then placed in a thermocycler preheated to 95°C. After 5 mins, the reactions were cycled 30 times at 95°C for 30 secs and then 70°C for 1 min. After cycling, 3 μl of stop solution was added to each reaction and the tubes stored at -20°C. Electrophoresis and autoradiography were then carried out as before (Section 2.19.3).

Chapter III

Cloning Strategies
Based on PCR

3.1 Introduction

This project began as a continuation of a well established investigation of gene expression in myeloma cells [Harrison et al., 1995; Glassford, 1993; Hudson, 1989; Hudson and Harrison, 1987].

Two lines of myeloma cell transfectants had been isolated that expressed high levels of a heterologous lysozyme gene. These lines were designated C6 and D8 (Figs 3.1 and 3.2). Genetic analysis established that the vector had integrated at a single site in both cell lines and that a single copy of the plasmid was present in C6 and two copies in D8 (Fig 3.3a).

The next objective of the project was the cloning of DNA sequences flanking the plasmid insert. This was desirable in order to facilitate gene walking and hypersensitivity studies in the wild-type cell line from which the transformants were derived.

The first step towards cloning involved the mapping of restriction sites in the region surrounding each integrated plasmid. The sites of four restriction enzymes were mapped by Southern blotting of genomic restriction digests and analysis of the resulting fragment sizes [Glassford, 1993]. Various cloning strategies were then employed with the aim of cloning a piece of flanking sequence from one or other of these lines. Straightforward plasmid cloning, plasmid rescue [Perucho et al., 1980] and inverse PCR [Ochman et al., 1988] were all attempted, unsuccessfully. More encouraging results were obtained from ligation-mediated anchor PCR [Loh, 1991]. Experiments showed the presence of plasmid sequence in PCR products of approximately the expected size although as part of a smear of products of varying size [Glassford, 1993]. It was from this point that the current work was undertaken.
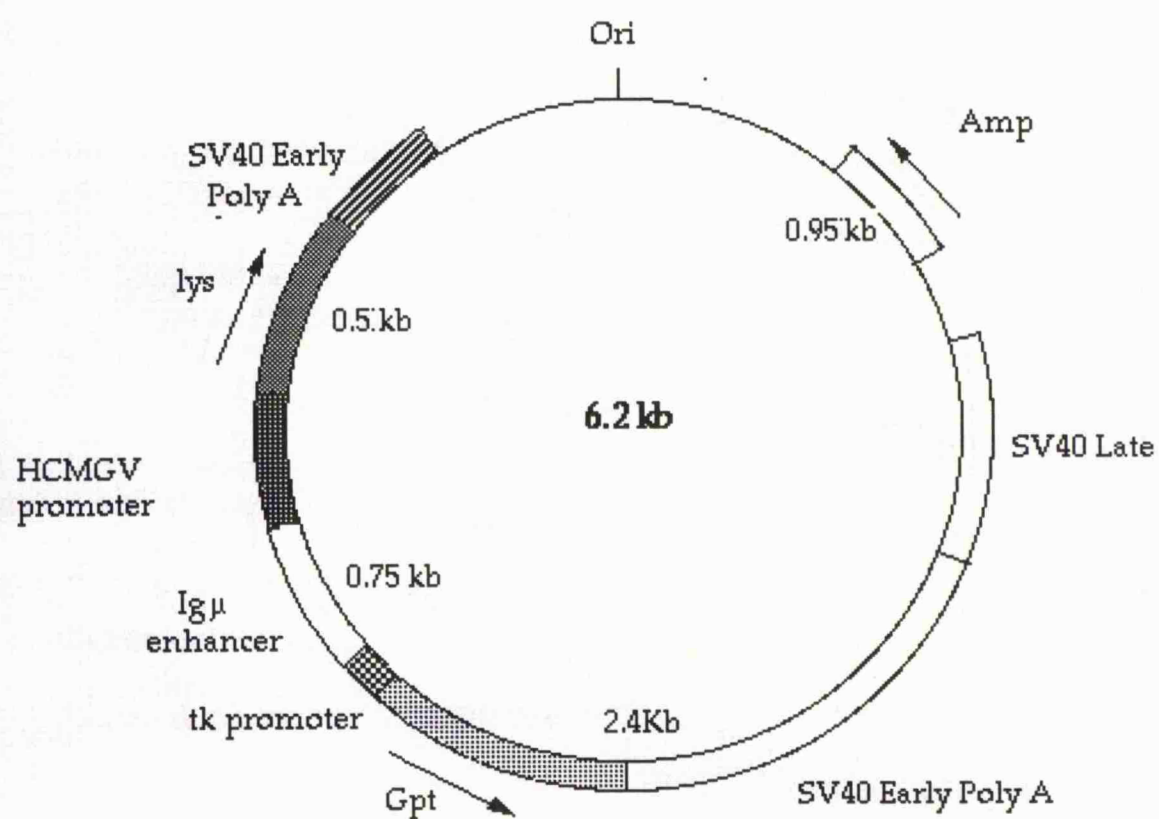
Fig 3.1 pLysSV40.gpt : The plasmid used to generate C6
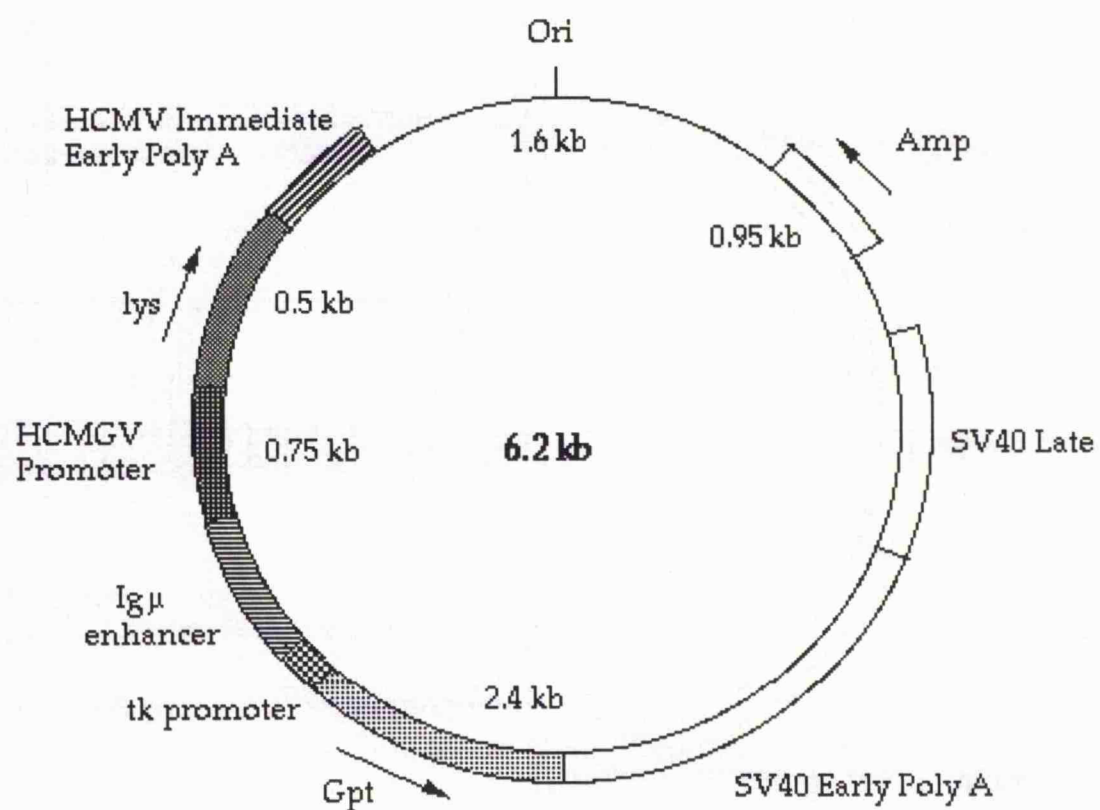
Fig 3.2 pLysCMV.gpt : Plasmid used to generate D8 line

D8

C6

B: Bgl II
E: EcoR 1
N: Nco I
P: Pvu Ii
T: Pst I

5kb

= gpt

= lys
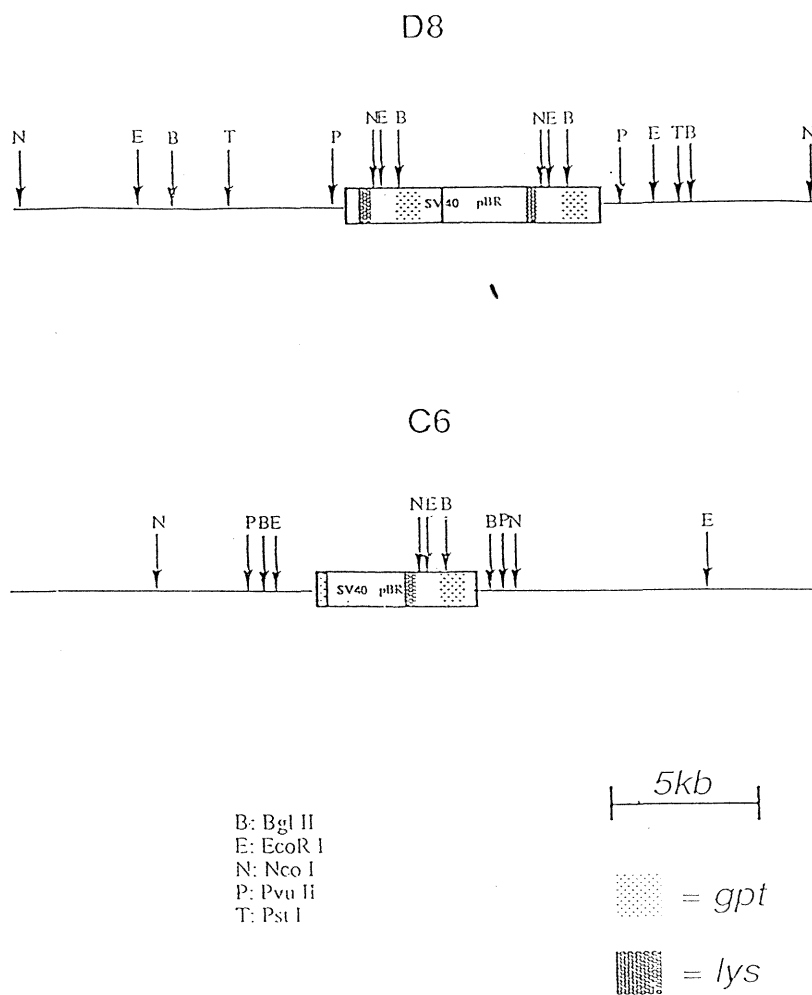
NB. Only the cleavage sites in the flanking DNA are
shown for the restriction enzyme Pst I.

Fig 3.3a Restriction maps of C6 and D8 regions of plasmid integration
From [Glassford, 1993]

Ligation-mediated anchor PCR is a technique for amplifying a section of DNA that contains known sequence at only one end [Rolfs et al., 1992; Loh, 1991; Innis et al., 1990; Erlich, 1989]. An oligonucleotide is ligated to the ends of the DNA fragment and this sequence can then be used as a 'non specific' PCR primer.

A restriction site is located in the unknown region that contains a four base pair overhang (Figs 3.4a and 3.4b). After a genomic digest with this enzyme, one base of the overhang is filled in using T4 polymerase. This prevents re-ligation of the restricted sites. The anchor oligonucleotide is then ligated onto the end of the 5' overhang.

A coupling oligonucleotide is used to facilitate this ligation. This has sufficient complementarity to the anchor to anneal to it with a three base 5′ overhang that is complementary to the three bases of the filled-in restriction site overhang. The coupling oligonucleotide cannot itself ligate to the target DNA because it lacks a 5' phosphate group. Because the 3′ end of the coupling oligonucleotide is not complementary to the anchor, it cannot itself prime a round of PCR. The coupling primer is present throughout the PCR cycles, but because it only has limited complementarity, it cannot anneal at the high temperatures used in the amplification cycles.

The first step in ligation mediated anchor PCR is to identify suitable restriction sites in the unknown flanking region. The restriction maps obtained prior to the start of this project (Fig 3.3) reveal sites in both C6 and D8 that might prove suitable.

The restriction map of the C6 integration site (Fig 3.3a) shows two Bgl II sites, 1.6 kb apart. This 1.6 kb Bgl II fragment spans the Gpt gene, SV40 sequence and into unknown flanking sequence. The exact end-points of the plasmid insert were unknown. Bgl II has a six bp recognition sequence and generates a four base overhang, which makes it ideal for an anchor
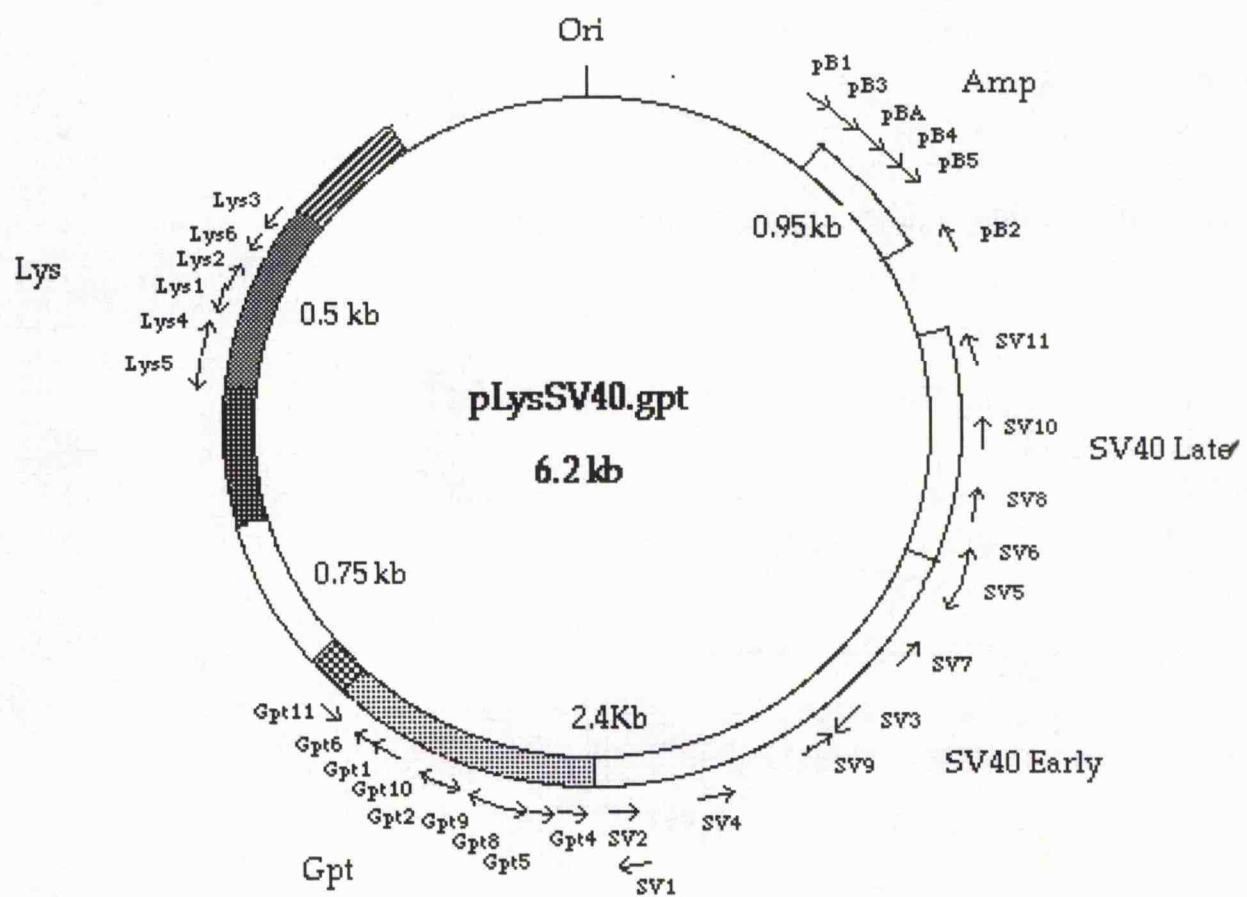
Fig 3.3b: pLysSV40.gpt plasmid showing primer locations
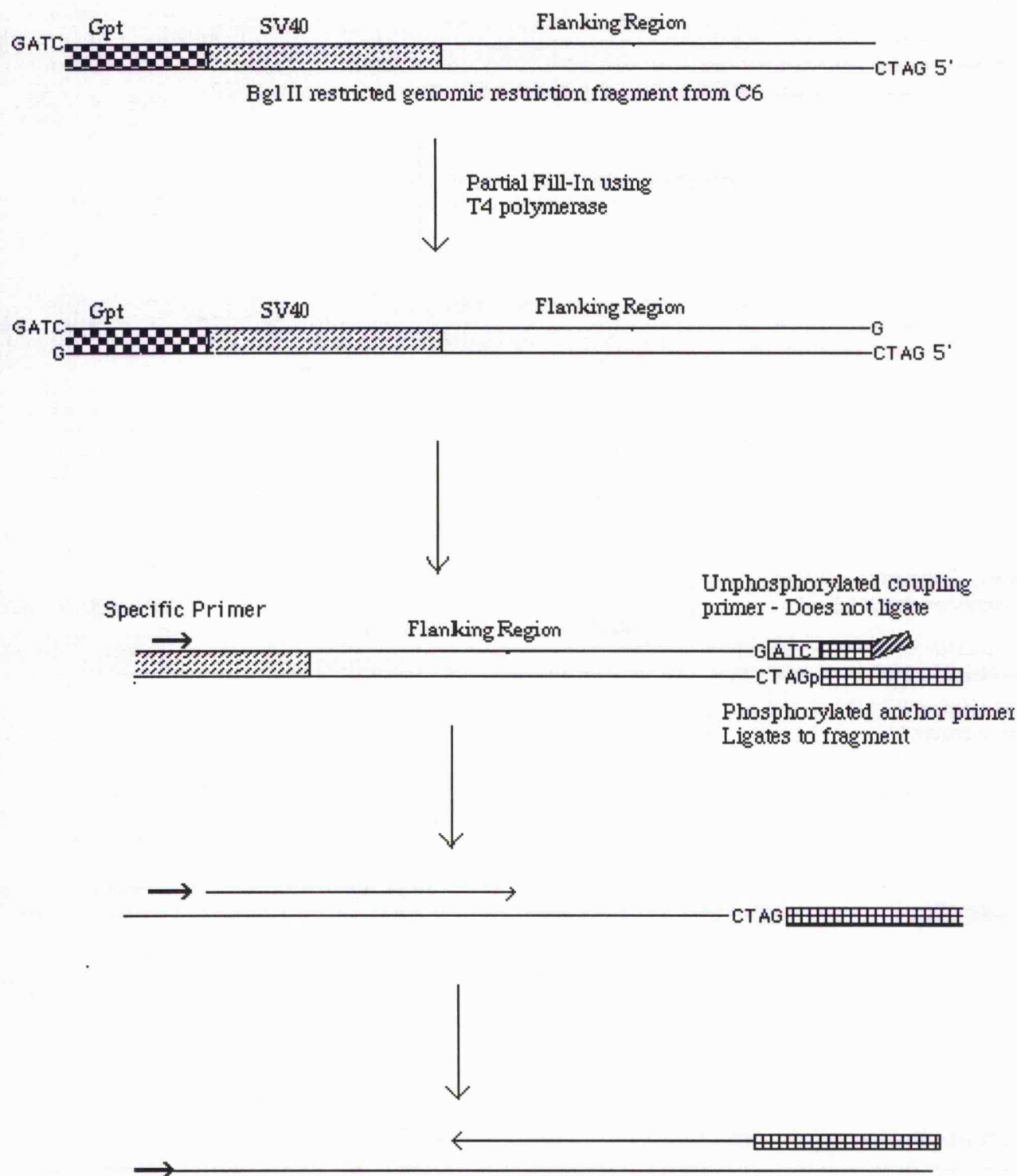
Fig 3.4 a Ligation mediated anchor PCR on C6
C6 genomic DNA is restricted using Bgl II to generate the 1,6 kb template fragment.
The overhanging ends are partially filled in using T4 polymerase to prevent self ligation.
Anchor primer is ligated onto the termini with the aid of a coupling primer, which is partly
complementary to the anchor. This oligonucleotide does not ligate because it lacks a 5' phosphate.
It does not interfere with the PCR because it can only anneal at lower temperatures. The 3'
non-complementary region prevents priming. PCR amplification then occurs between anchor and
specific primers.

Fig 3.4b Ligation Mediated Anchor PCR on D8
D8 genomic DNA is digested with PvuII restriction enzyme. The blunt ends are digested
with T4 polymerase to generate 2 base 5' overhangs. The anchor is then ligated onto
the termini with the help of a coupling oligonucleotide. A PCR product is then generated
using a sequence specific and an anchor specific primer.

PCR. The Bgl II fragment is itself situated on a 4 kb Nco I fragment. This allows for an enrichment strategy to be adopted whereby DNA of 4 kb is purified by agarose gel electrophoresis from an Nco I genomic digest and then digested with Bgl II for the anchor PCR strategy.

Assuming that a haploid mammalian genome (3×10^9 bp) weighs approximately 3.4×10^{-12} g, then 1 μg of template contains approximately 3×10^5 copies of a single-copy target sequence. Since the number of effective targets with ligated anchor was likely to be much lower than this, an enrichment strategy aiming to increase the amount of target template in the PCR reaction was devised. A second, nested PCR on the purified products of the first was also planned.

D8 genomic DNA is less amenable to an anchor PCR strategy than C6 because there are no suitable restriction sites close to the ends of the plasmid sequence (Fig 3.3a). The Pvu II sites that flank the insertion could be used instead. Pvu II is a blunt ended cutter, so a modified strategy would have to be employed (Fig 3.4b). After digestion, two 3' bases at the restriction site were chewed back using the 3' to 5' exonuclease activity of T4 polymerase. The addition of two deoxynucleotides (dCTP and dTTP), combined with the 5' to 3' polymerase activity of the enzyme, prevented any more than two bases being removed. 5' CT overhangs were left at the Pvu II site, to which the anchor primer could then be ligated. Because anchor should ligate to both ends of the restriction fragment, amplification should be possible from both ends of the fragment.

In this project, D8 was the first cell line to be developed as a target for the anchor PCR technique. Similar work was being done concurrently on C6 by other workers in the laboratory [Glassford, 1993]. Later, work on D8 was dropped and the focus of this project moved to C6 as a target for ligation-mediated anchor PCR, as it appeared to be the most suitable target.

Fig 3.5 D8 Genomic Digest
1% agarose gel showing 500ng of D8 DNA after digestion with Pvu II (Section 2.13).
Marker - 1 kb ladder

3.2 Anchor PCR on D8

10 µg of D8 genomic DNA were digested with 10 U of Pvu II in a total reaction volume of 20 µl of enzyme buffer containing 1 mg/ml BSA. After 3 hrs at 37°C, the DNA was ethanol precipitated and the pellet dried in a vacuum dessicator and resuspended in 10 µl of water.

The 3' ends were chewed back by adding 2 µl each of 2 mM dTTP, dCTP, T4 polymerase buffer (10 x) and water. 10 U of T4 polymerase was added to this, and the reaction incubated for 30 mins at 37°C. After this, the DNA was ethanol precipitated and resuspended in 40 µl of TE. A small aliquot was run on a gel to check if DNA was still present (Fig 3.5).

To ligate the anchor primer to this DNA, 8 µl of DNA (about 1 µg) was mixed with 10 µl each of 10 µM And 2 and CFB 2 oligonucleotide primers (Appendix II.4). The solution was heated to 65°C for 5 mins then cooled to room temperature. 8 µl of 5 x T4 ligase buffer and 10 U of T4 ligase were then added and the volume made up to 40 µl with water. This was incubated overnight at 15°C.

A PCR reaction was set up in a total volume of 100 µl containing 4 µl of the template prepared above (100 ng), 1 x taq polymerase buffer, 0.2 mM dNTPs, 0.8 µM And 2, 0.8 µM pB3 (Appendix II.5) 2.5 U Taq polymerase and 1 µl perfect match PCR enhancer. A touchdown PCR was performed, after a hot-start addition of enzyme, that contained two rounds of amplification at each annealing temperature from 65°C to 58°C, for 90 secs each, and then 25 cycles at 58°C. The PCR cycle also contained a 90 sec denaturation step at 95°C and a 3 mins extension at 72°C.

1 µl of PCR product from the first reaction was then diluted in 1 ml of water. 10 µl of this dilution was then added to a second PCR reaction that was identical to the first, except the specific primer pB3 had been

replaced by a nested primer pBA. The same thermocycling programme was used.

Products were then run on an agarose gel (Fig 3.6). Although the amplification product appeared as a smear of indeterminate size, the gel was Southern blotted and probed with end labelled oligonucleotides from downstream of the nested primer.

The autoradiograph indicated that, although rather heterologous in size, PCR products containing plasmid sequence had been amplified (Fig 3.7). The smear of generated product covered a size range from 0 - 2 kb, while the expected size for correctly amplified product was about 1.5 kb. The next objective was to establish whether these products were genuine or artifactual by cloning and sequencing.

3.3. Cloning the D8 PCR Product

The preferred strategy for cloning the 'smeary' PCR product was to use the Sal I restriction site in the And 2 primer to make one end of the product 'sticky' and to clone into pUC 19 using a semi-blunt ended ligation. However, this approach was discounted for two reasons. Firstly, the efficiency of this ligation was found to be very low. Secondly, a problem arose about how to identify positive clones. The only known sequence in the PCR product from which to obtain a probe is the plasmid sequence from which the pB primer series was designed (Appendix II.4). Since this sequence is also found in pUC 19, this vector, or any derived from pBR322, is not suitable to use in this context.

A different approach was adopted that employed a third round of PCR to introduce a second restriction site. A plasmid, pACYC184, was identified that did not contain any sequence derived from pBR322.

1 μl of the nested PCR product was diluted 1000 fold and 10 μl employed in a PCR reaction identical to those described above, except that

1.6 Kb
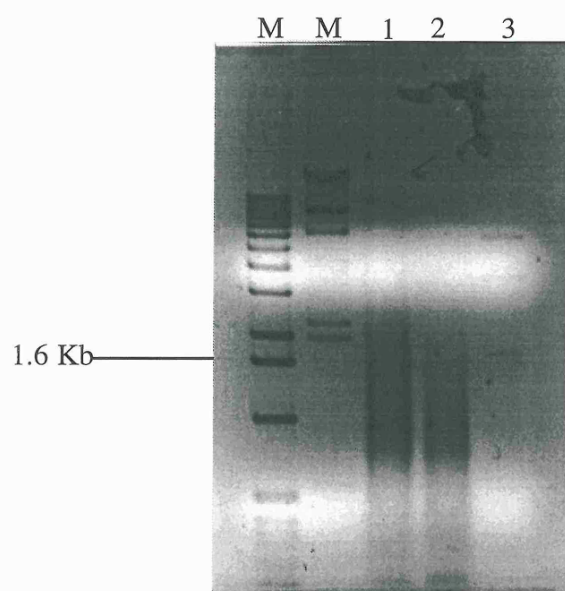
Fig 3.6 D8 Anchor PCR
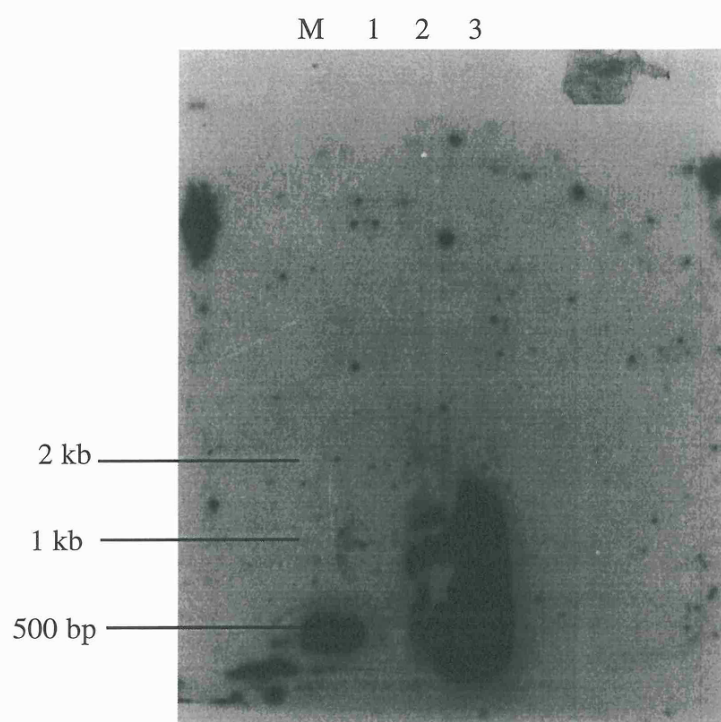1% agarose gel showing D8 anchor PCR products (See Section 2.17)
Track 1 - Initial amplification
Track 2 - Nested amplification
Track 3 - 10 ng pYAC4 digested with Eco RI and Bam HI
Markers ; 1 kb ladder and λ digested with Hind III

Fig 3.7 Southern analysis of D8 nested PCR

Autoradiograph of D8 nested PCR product pB4Bam / And2 probed with end labelled pB5 as described in Section 2.17. Membrane washed at 45°C in 0.1 x SSC for 30 mins. Autoradiograph developed for 15 hours at -70°C with intensifying screen.

Track 1 - D8 template DNA
Track 2 - Initial PCR
Track 3 - Nested PCR
Marker is 1 kb ladder

| Volume ligation mix transformed	10% transformed cells	90% transformed cells
1µl	0	1 6
8µl	1 6	9 9
10µl Control	8	5 6

Table 3.1 Numbers of transformant colonies obtained after ligation

This experiment was designed to show the presence of vector containing cloned sequence. This is indicated by the presence of more transformant colonies on the experimental plates than would be expected from background self ligation, as observed in the control.

Ligation reactions were performed in a total volume of 10µl. Approximately 50ng of pACYC184 was used in each. The amount of PCR product in the experimental reactions was difficult to quantify but was assumed to be about 10ng.

Aliquots of ligation, as indicated, were added to 100µl of competent TG2 cells. After transformation, 10% and 90% of the transformed cells were plated onto LB Chloramphenicol agar plates. The number of colonies on each plate were counted and tabulated.

For the control, only linearised plasmid was used in the ligation. The whole 10µl of ligation was used in the transformation.
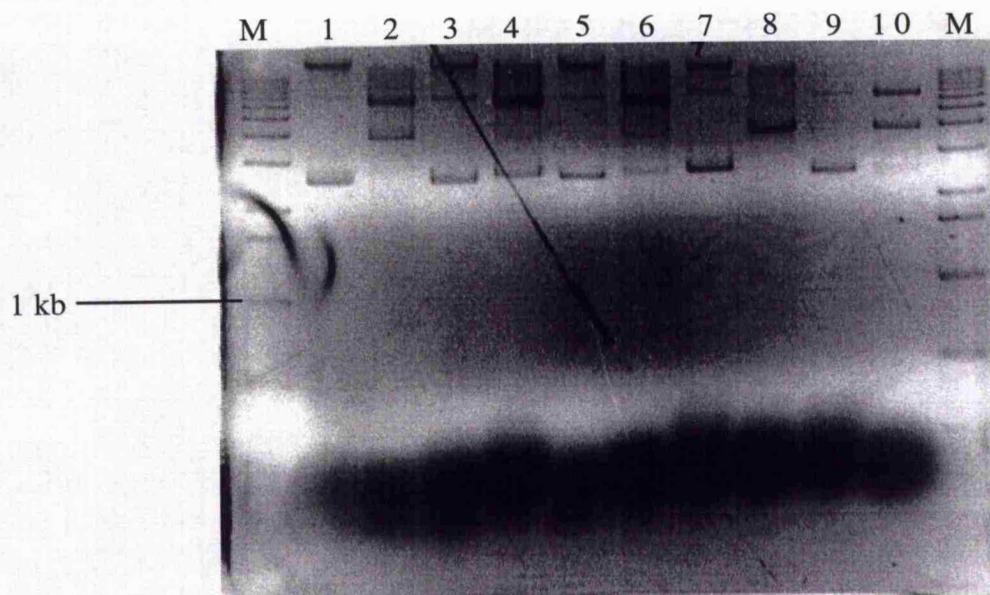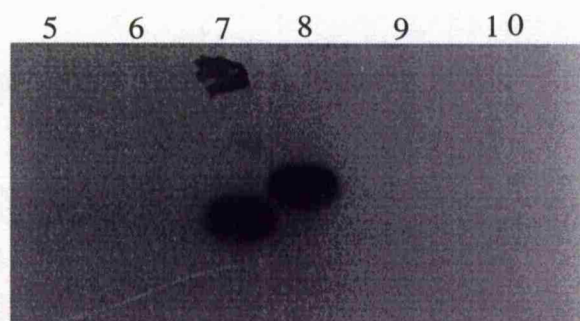
Fig 3.8a Preparations of plasmids containing putative cloned PCR product
1% agarose gel of plasmid minipreps containing putative cloned PCR products. Digests
in tracks 4 and 6 appear to have been unsuccessful.
Tracks 1, 3, 5, 7, 9 - Uncut plasmid
Tracks 2, 4, 6, 8, 10 - Sal I digests of plasmid
Marker - 1 kb ladder

Fig 3.8b Southern analysis of putative cloned D8 inserts
Autoradiograph of Southern blot of gel shown in fig 3.8a showing tracks 5 - 10, probed with end labelled pB5 (Section 2.16). Plasmid number 4 (tracks 7 and 8) contains sequence that hybridises to the probe indicating that it is derived from D8.
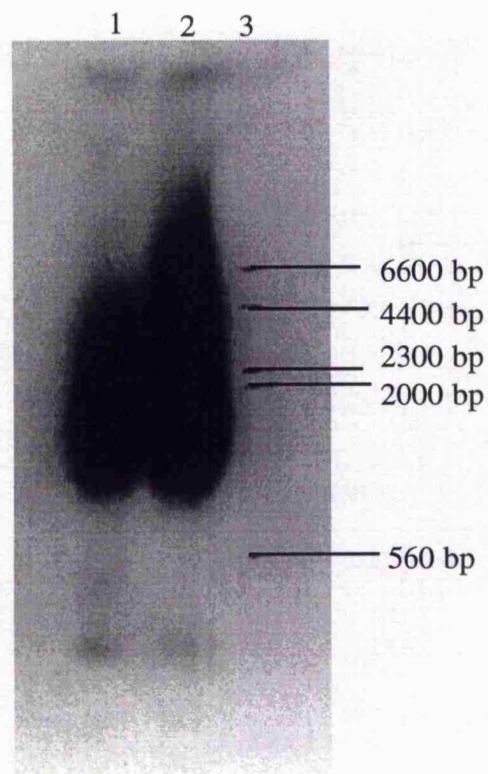
Fig 3.9 Southern analysis of restricted nested C6 PCR product
Autoradiograph of a Southern blot probed with end labelled SV40 4
oligonucleotide (Section 2.16). PCR product is shown from a nested reaction
using Gpt 4RI and And 3 on a template of product from a Gpt 7 and And 3
reaction.
Track 1 - Eco RI cut nested PCR product from the primers Gpt 4RI and Gpt 7
with the anchor primer And 3
Track 2 - Undigested PCR product
Track 3 - λ Hind III marker

the specific primer was pB4Bam (Appendix II.4), which contained a Bam HI site at its 5' end. A touchdown from 65°C to 55°C was used in a 25 cycle programme.

The products of this reaction were then digested with Bam HI and Sal I in a double digest for 3 hrs at 37°C. The digestion products were fractionated by electrophoresis in a 1% low melting point (LMP) agarose gel. A gel slice containing DNA of 1 - 2 kb was removed, treated with gelase and the DNA ethanol precipitated and cleaned in a Chroma-TE 400 column. The pACYC 184 vector was prepared by digesting with Sal I and Bam HI then phosphatasing with calf intestinal phosphatase (CIP). The linear phosphatased vector was then purified by electrophoresis in a 1% LMP agarose gel and extracted using gelase. A 4 μl aliquot of product was then ligated to the pACYC184 in a standard ligation reaction at 15°C overnight. A control containing no PCR product was also set up.

The ligation was then transformed into TG2 cells, which were plated out on LB chloramphenicol medium. An α-complementation screen could not be used because no suitable lac sequence is present in the vector. A simple initial check for inserts was carried out by comparing the numbers of transformants on control plates, from the ligation mix which did not contain PCR product, with the numbers on the experimental plates (Table 3.1).

This data provides a rough indication that vector containing insert was present in some of the colonies. More colonies were present on the experimental plates than would be expected if they were all caused by the background levels of plasmid re-ligation indicated by the control.

Colonies containing inserts were identified by hybridisation with pB5 oligonucleotide (Appendix II.4). After hybridisation with end-labelled pB5 probe, putative positives were identified and picked off the filter. The plasmids were miniprepped, Southern blotted and reprobed. At this stage,

one positive clone was identified that contained pBR sequence derived from the integrated plasmid (Fig 3.8a and b).

From Figs 3.8a and b, the linearised vector containing the cloned sequence appeared to be about 4 kb. Since the vector itself was 4 kb, the cloned insert was much smaller than expected (< 500 bp). It was therefore presumed to be artifactual. No further work was done on it and attention centered on C6, which was more amenable to PCR.

3.4 Initial C6 Cloning Experiments

At the beginning of this work, C6 anchor PCR products were inherited from previous work [Glassford, 1993]. These were generated with And 3 and Gpt 4RI oligonucleotide primers (see Appendix II) as described in Section 3.1 and appeared as smears of up to about 4 kb on both agarose gels and autoradiographs probed with plasmid-specific sequence. It was not known whether these smears represented a genuine, if heterogeneous, template based PCR product or an artifactual composite of primers produced irrespective of the presence of template

Before cloning the product into a vector, an experiment was performed to investigate whether the product was genuine or a 'composite' primer artifact. Since the primers were designed to contain Eco RI restriction sites, an Eco RI digest should in the latter case, where a large number of primer copies are present, considerably decrease the size of the observed product, whilst having no effect in the former case, where only one is present.

An Eco RI digestion of the product was performed and the products run out on a 1% agarose gel. This was Southern blotted and probed with SV40 sequence. An autoradiograph (Fig 3.9) indicated that there was no size difference between the digested and undigested PCR products. This

implies that the products are not largely or wholly composed of PCR primers. This confirms the observation that no product was present in the amplification reaction performed in the absence of template, as might be expected if it was primer derived.

In order to clone the PCR product, it was first digested with Eco RI in order to generate 'sticky ends' for ligation. This was then ligated with Eco RI cut phosphatased pUC19. The ligation mix was then transformed into DH10 cells and the transformants subjected to a blue/white screen. About 70 % of colonies were white and these were picked onto a filter, denatured, fixed and probed with radiolabelled SV40 2/3 PCR product.

Positive colonies were picked into overnight cultures (5 ml LB + ampicillin), grown for approximately 12 hrs at 37°C and then miniprepped (Section 2.7).

These preps were digested with Eco RI and run alongside uncut plasmid on a 1% agarose gel (not shown). These results showed inserts of much smaller size than the desired 800 bp. To obtain more information to explain this, four of the inserts were sequenced using the 'Sequenase' system (Sequences shown in Appendix I)

As can be seen, all four contain a mixture of plasmid sequence and PCR primers in various orientations (Fig 3.10). This seemed to be a result of failures in the PCR reaction. The presence a of downstream plasmid sequence, however, indicates that the specific primer is priming correctly. This provides more information about the boundary of the insert since SV40 sequence seems to be present. As there is only one plasmid copy in the cell line this is the only place from which SV40 sequence could have come. This allows the use of primers from the SV40 sequence downstream of the gpt gene. These primers would produce a smaller PCR product with the anchor, which might mean a more efficient amplification.
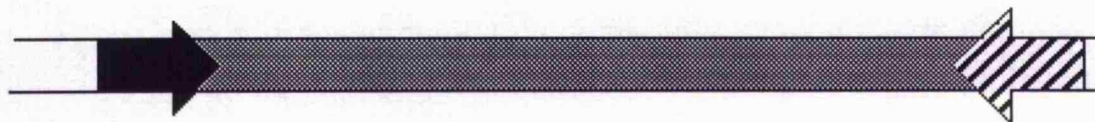
And 3/Gpt4RI Products (Appendix I)
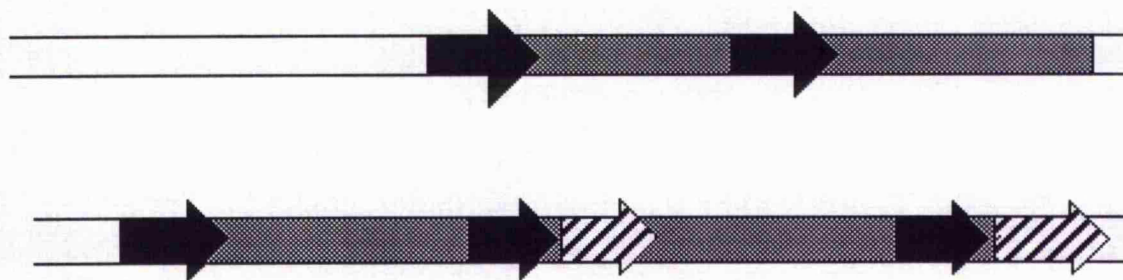
SV406/And 3 Product (Fig 3.16)
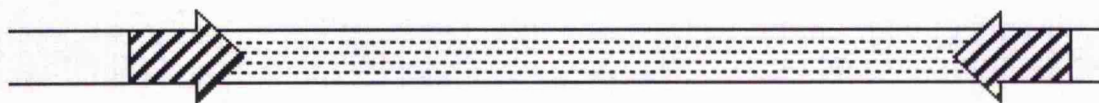
SV406 /And4N Product (Fig 3.23)

| Unknown Sequence | Integrated Plasmid | Specific Primer | Anchor Primer |

Fig 3.10 Diagrams of Cloned PCR Products

These diagrams show the configurations of cloned PCR products. It can be seen that all the clones contain unexpected combinations of primers. Where specific primers are included, plasmid sequence from the integrated vector is found, without any unknown flanking sequence. One clone contains unknown sequence but no plasmid sequence or specific primers, so it is of little use. The unshaded region represents the vector into which the products are cloned.

Alternatively, the small product size may be due to rearrangement during or after the cloning process. The high copy number of the pUC vector used may help to generate instability in the PCR sequence. In addition, a host strain that has deficient recombination systems might be required to clone this product.

3.5 Incorporation of SV40 Sequence into PCR Strategy

The sequence data obtained from the initial PCR experiments enabled specific primers to be designed much nearer to the flanking sequence containing the Bgl II site, to which the anchor is ligated. This should facilitate the PCR amplification.

The oligonucleotide primer SV40 5 was designed to be on the edge of known sequence, whilst SV40 2 and SV40 4 were confirmed as present in the insert (Appendix II.3).

SV40 5 is orientated in towards the known sequence, rather than out towards the anchor (Fig 3.3b). This allows a straightforward PCR check to be made on the presence or absence of the sequence complementary to the primer. PCR reactions were set up using three different primers (SV40 4, SV40 2 and SV40 9) with SV40 5. Ten fold dilutions of C6 DNA were used as a template and plasmid pLysSV40.gpt and myeloma line J558L DNA were used as positive and negative controls. After a 65°C to 55°C touchdown PCR over 35 cycles, bands were observed in all the experimental gel tracks (Fig 3.11a). The sizes of bands relate to the expected sizes from published SV40 sequence. This confirms that the SV40 sequence at this end of the insert stretches at least as far as the SV40 5 sequence.

Because it points (5' - 3') away from the flanking region (Fig 3.3b), SV40 5 cannot be use for an anchor PCR (Fig 3.11d). SV40 6 was designed as
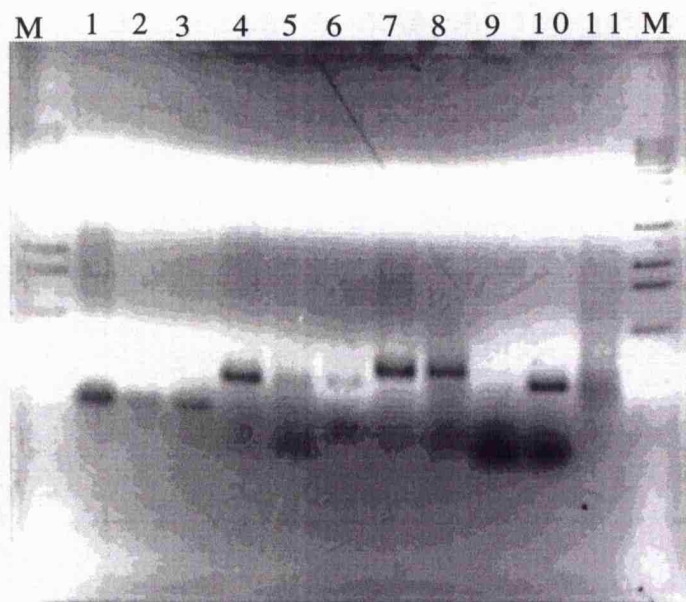
Fig 3.11a PCR from C6 genomic template
1% agarose gel showing PCR products from C6 genomic template (Section 2.17)
Tracks 1-3, SV40 4 / SV40 5; 3.8 μg, 380 pg, 38 pg template
Tracks 4-6, SV40 2 / SV40 5; 3.8 μg, 380 pg, 38 pg template
Tracks 7-9, SV40 9 / SV40 5; 3.8 μg, 380 pg, 38 pg template
Track 10, SV40 2 / SV40 5 on pLysSV40.gpt template
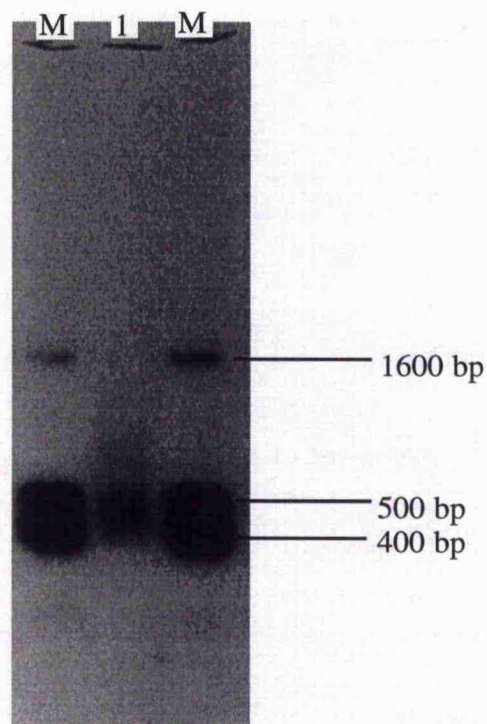Track 11, SV40 2 / SV40 5 on J558L genomic DNA
Marker, 1 kb ladder

Fig 3.11b Nested C6 product before cloning
1% agarose gel showing SV40 6 / And 3 nested PCR product from a SV40 7 / And 3
template, later cloned into pGEM-T plasmid (Section 2.17.5). Product appears to be
about 450 bp which is approximately the expected size.
Marker - 1 kb ladder

Fig 3.11c Southern analysis of C6 nested PCR product
Autoradiograph of Southern blot of gel shown in fig 3.11b, probed with randomn primed
pLysSV40.gpt sequence (Section 2.16).
Membrane washed in 0.1 x SSC at room temperature and the autoradiograph was devel-
oped for 36 hours at -70$^{\text{O}}$C with an intensifying screen

Fig 3.11d Diagram showing Locations and Orientations of Primers
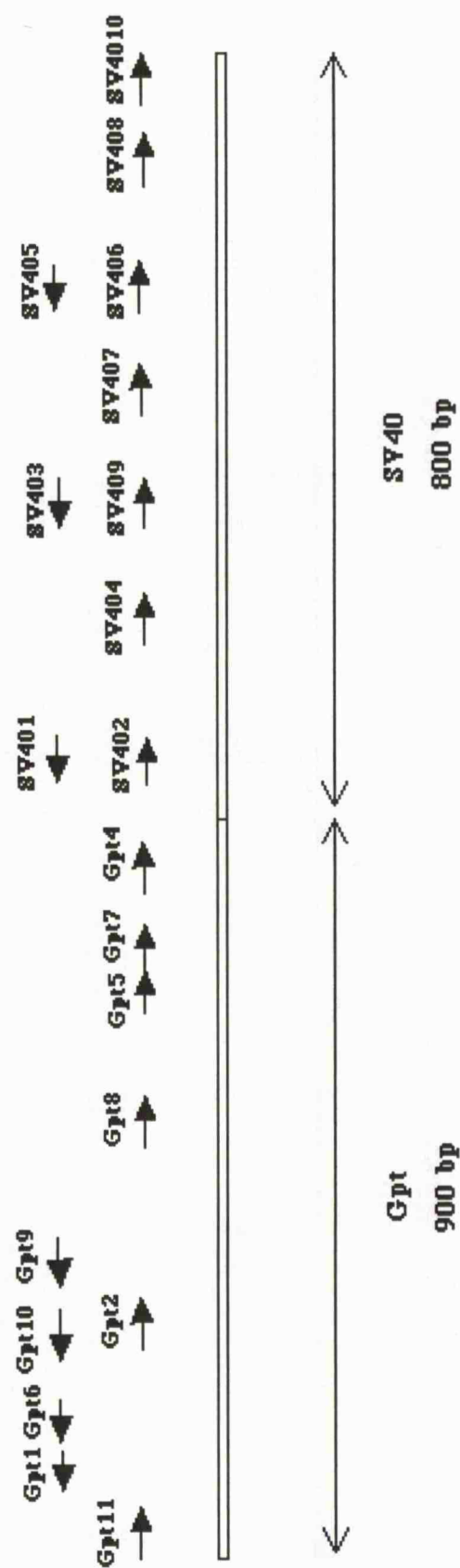
complementary to SV40 5, and pointing (5' - 3') towards the anchor. A second primer, SV40 7, was also designed, to anneal just upstream of SV40 6, so the two could be used in nested PCR experiments. This system was used to generate a smeared product of up to 500 bp. A smeared band of about 450 bp hybridised to the SV40 oligonucleotide probe (Fig 3.11b and c). This is about the size expected for this product. It can be seen that marker bands also contain SV40 sequence, the bands at 400 bp and 500 bp are clearly visible. These marker-bands produce stronger signals than the PCR product because of the large amount of DNA present.

Prior to the anchor cloning experiment, a control experiment was performed to check that the ligation was working. 20 ng of plasmid pLysSV40.gpt DNA was digested with Bgl II. The 5' overhang was filled in by one base with T4 polymerase and ligated to And 3 anchor primer in the presence of coupling primer CFB (Appendix II.4). This DNA was purified by 'Magic' PCR prep and used as a template in a PCR reaction with And 3 and SV40 1. A product should be generated if anchor oligonucleotide has succesfully ligated to the filled in Bgl II site.

As can be seen (Fig 3.12), a 900 bp band is visible for two template DNA concentrations, with intensity increasing with amount of template. This indicates that the ligation is working properly.

Having established that the new specific primers were functioning correctly and the ligation reaction was working, template DNA was prepared for an anchor PCR experiment. 100 µg of C6 DNA was digested overnight with Bgl II, filled in and ligated to And 3 anchor in the presence of CFB (Section 2.17.4). The ligation mix was purified using a Chroma-spun-400 column in a final volume of 40 µl of water. 10 µl of this DNA was used in a PCR with SV40 7 and And 3 primers in a total volume of   100 µl. The program used was a touchdown over 10 cycles from 65ºC to 55ºC followed by 25 cycles at 55ºC. The product from this reaction was
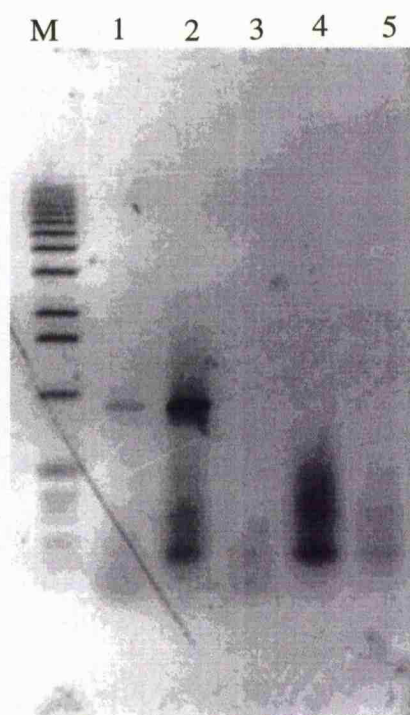
Fig 3.12 Assay of ligation efficiency with And 3 anchor primer
1% agarose gel showing PCR reactions with SV40 1 and And 3 primers on
plasmid pLysSV40.gpt DNA template to establish the efficiency of primer
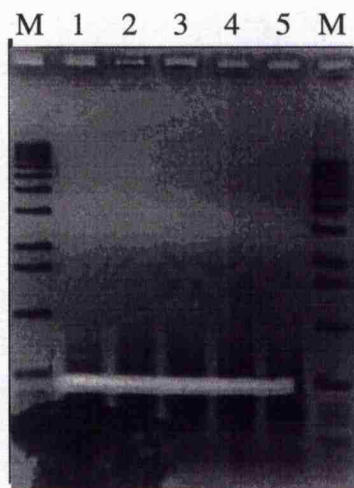ligation.
Track 1, 300 pg target plasmid
Track 2, 1.5 ng target plasmid
Track 3, No DNA control
Track 4, No ligase control
Track 5, No coupling primer control

Fig 3.13a Removal of gel slice for nested PCR
1% LMP agarose gel showing excision of gel slice prior to nested PCR (Section 2.17).
Tracks 1-5, SV40 7 / And 3 amplification product
Marker - 1 kb ladder
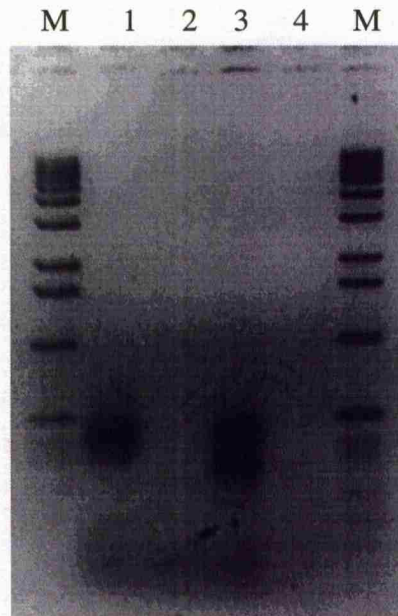
Fig 3.13b SV40 6 / And 3 nested PCR amplification
1% agarose gel showing nested PCR products from template prepared in fig 3.13a with the
following primers (as described in Section 2.17);
Track 1 - SV40 6 / And 3 product
Track 2 - Control - No DNA
Track 3 - Control - And 3 only
Track 4 - Control - SV40 6 only
Marker - 1 kb ladder

run on a 1% agarose gel and the gel slice containing products of a size corresponding to around 450 bp removed using a sterile razor blade (Fig 3.13a). The SV40 7 primer is about 450 bp downstream of the Gpt 4 primer used previously, hence the size of the expected product drops from 900 bp to 450 bp (Fig s 3.3b and 3.11d).

The DNA in the gel slice was extracted using the 'Qiaex' system and used as a template for a nested PCR using the primers And 3 and SV40 6. This was amplified as before. An aliquot of product was run out on a 1% agarose gel (Fig 3.13b). A band can be seen at about 400 bp as expected. This band is not as clean as would be expected from a double sided specificity PCR, but still appears to be too strong to be a primer-derived artifact.

It is possible that the selection of a particular size of template before the nested PCR creates a nested product of similar size, regardless of the specificity of the reaction. A series of misprimed artifacts, if selected a first time and possessing the correct 5' sequence, will also amplify in the nested reaction. Interestingly, smaller bands appear with just the anchor primer. The presence of these bands seems to be a characteristic of anchor PCR and is probably due to the misprimed amplification of small fragments with anchor sequence ligated to both ends.

The next step was to clone the 450 bp band produced by SV40 6/And 3. Because the primers used did not contain restriction sites the pGem-T system was used in the cloning reaction. The PCR product was prepared using a Chroma Spin-100 column and ligated to 0.25 µg pGEM-T vector. Upon completion, aliquots of this ligation reaction were used to transform DH10 cells. These cells were plated out in an α-complementation screen and 20 white colonies were picked off, grown in overnight LB culture and miniprepped. In order to establish the identity of the insert, minipreps were digested with Sal I and Nco I and the products analysed by agarose gel electrophoresis (Fig 3.14). These sites are located on either side of the
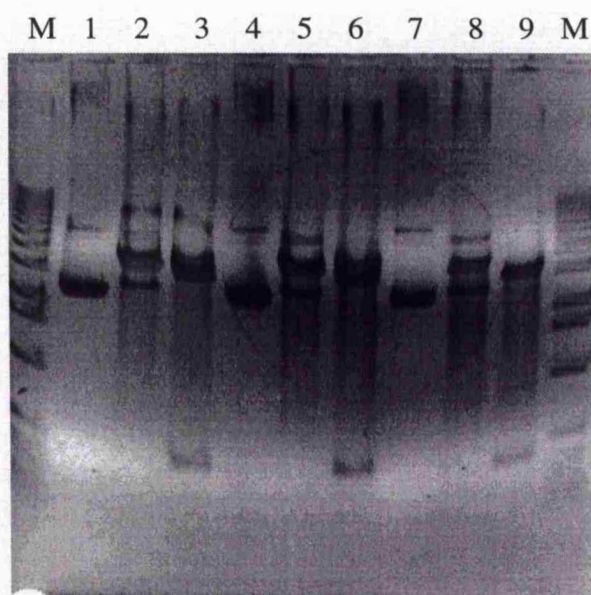
Fig 3.14 Restriction analysis of plasmids containing cloned inserts
1% agarose gel showing plasmid minipreps (Section 2.7) of pGEM-T vector
with C6 PCR product insert. These represent three of the twenty insert bear-
ing plasmids isolated. Sal I appears to have only partially digested these
plasmids, probably because of the presence of inhibitors in the DNA prep.
Tracks 1, 4, 7 - Uncut plasmid
Tracks 2, 5, 8 - Plasmid digested with Sal I
Tracks 3, 6, 9 - Plasmid digested with Sal I and Nco I
Marker - 1 kb ladder
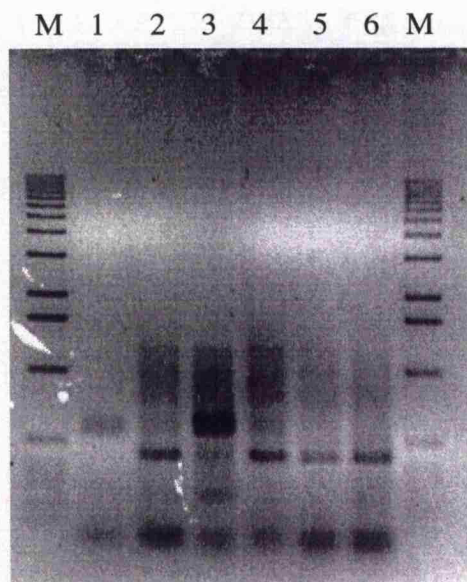
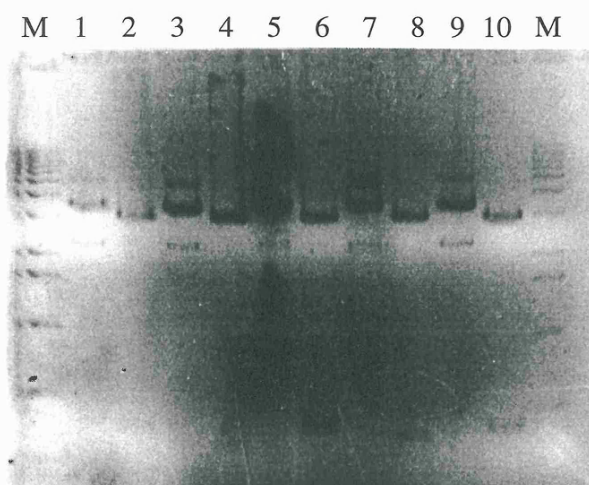**Fig 3.15 PCR amplifications across cloned inserts**
1% agarose gel showing PCR amplifications on pGEM-T plasmids containing
inserts (as described in Section 2.17) using M13 F and M13 R primers to amplify
across the insert. These represent six of the twenty clones isolated.
Tracks 1 - 6, M13 F / R amplifications on plasmid template isolated from six
                independent transformants.
Marker - 1 kb ladder

Fig 3.16 Sequence Data from Cloned C6 PCR Product

Sequence data obtained via the 'Sequenase' method using M13 forward and reverse primers to sequence the ends of the insert. This shows the presence of anchor sequence at both ends of the insert.

M13 F Primer

TGTAATACGACTCACTATAGGGCGAATTGGGCCCGACGTCGCATGCTCCC

                                And3
NNCCGCCATGNGGGATTCNCTGAATTCTGCTATGGAAATGCTGGTGTTTG

AAGAGAGAGTATTTCACAATGGCAGAAATAAATCACGGCTCTATGAGC

TCTTNGCCTTTTTTANNTTTTTGATCAGGTGGATTTCCTACCAACCTGAA

AGAATAGATTTTATATGAAGATGGAAGAGTGGGAGTAATTTTAGTTC

CCAGATAACAATATTGTCGGTCGTTCTCAATCCTGAG

M13 Reverse

ACGCCAAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAAC

CGCGTTGCGGAGCTCTCCCATATGGTCGACCTGCAGGCCGCGCCAGCAG

                         And3
ACTAGTGATTCCGCCTGAATNCTCGTATGGAGATCTCTTGACGCACGACC

TGGAGTTTACGACCAGTTT

Fig 3.17 Restriction endonuclease analysis of insert bearing plasmids
1% agarose gel showing restriction endonuclease digests of pGEM-T plasmid containing
insert (Section 2.13.1). Five of the twenty isolated clones are shown.
Sal I linearises the plasmid without cutting the insert sequence whereas Eco RI cuts out
the insert. The Sal I digests show only partial digestion in some of these tracks.
Tracks 1, 3, 5, 7, 9 - Plasmid cut with Sal I
Tracks 2, 4, 6, 8, 10 - Plasmid cut with Eco RI

cloning site, so a digest with both enzymes should excise any insert. The preps were also analysed by PCR. Aliquots of DNA were diluted 1000-fold and amplified using M13F and M13R primers (Fig 3.15).

This indicated cloned inserts of 400 bp and 600 bp. Southern analysis was inconclusive as to whether this insert was the expected PCR product, so plasmids containing the 400 bp insert were miniprepped, denatured and sequenced. Sequence analysis indicated (Fig 3.16) that the clone contained anchor sequence at both ends so it could not be a genuine PCR product. This was tested for all the clones simply by digesting the clones with Eco RI (Fig 3.17). The insert could be separated from the plasmid with a single digest, confirming the presence of Eco RI restriction sites at either end of the insert. Digests with Sal I and Eco RI were also compared. If the PCR product insert has introduced more than one Eco RI site, then a decrease in the size of the plasmid band should be visible. This is observed and confirms the presence more than one copy of And 3, which was engineered to contain an Eco RI site. The plasmid has no such site. All the PCR products cloned in this experiment are therefore artifactual.

3.6 Measurements of the Efficiency of Primer Ligation

The PCR experiments carried out, up to this point, were not successful. In order to try and establish the reasons for this failure, assays were carried out on the template construction steps to ascertain the efficiency of the fill-in, label incorporation and primer ligation reactions.

The efficiency of the fill-in step was assessed by examining the ability of a linearised plasmid to self-ligate either without treatment or after a fill-in reaction. 20 ng of pLysSV40.gpt that had been linearised with Bgl II and filled in using 4 mM GTP, was ligated at room temperature for 48 hrs. A ligation containing 20 ng of untreated linearised plasmid was also set up.

Filled In	No of Colonies
50μl	0
100μl	0
Not Filled In	
50μl	340
100μl	650

Table 3.2   Assay of fill-in effectiveness

This experiment was designed to assess the effectiveness of the fill-in reaction for Anchor PCR by observing the efficiency of prevention of self-ligation. Linear plasmid with ends that have been successfully filled in will not self-ligate, to allow transformant colonies to grow.

20 ng of pLysSV40.gpt were digested with Bgl II and filled in using 4mM GTP, as described in Section 2.17.4. This DNA was ligated for 2 days at 4^0 C in a total volume of 10μl. 1μl of ligation reaction was transformed into 100μl of competent TG2 cells.

After transformation, these cells were resuspended in 1ml of LB. Aliquots of 50μl and 100μl were plated out onto LB ampicillin plates.

As a control, similar amounts of non filled in, linear plasmid were treated in the same way.

The number of colonies on each plate were counted and tabulated.

Filled-in plasmid does not self-ligate and so does not generate any transformants. These results indicate that the fill-in reaction was successful.

Filter	Counts
Unwashed	430 000
Washed	590 000

Table 3.3 Analysis of the amount of radiolabel added to anchor primer

This experiment was designed to assess the amount of radiolabel that had been attached to the anchor primer. Anchor primer And 3 was end labelled using $\gamma[^{32}P]dATP$ and T4 polynucleotide kinase as described. The labelled primer was then precipitated using ethanol and 5M ammonium acetate. The probe was resuspended in 50μl of TE and four 2μl aliquots removed and placed onto DE-81 filters. Two of these filters were washed four times in 0.5M NaH_2PO_4 for 10 mins each. The other two were left untreated. Counts were measured by Cerenkov counting in the ^3H channel of a liquid scintillation counter.

The results from washed and unwashed filters were averaged and the results tabulated. These results indcate that most of the measured activity has been incorporated in the oligonucleotide and so cannot be washed off.

Sample	Counts
And 3- Precolumn	11 953
And 3- Postcolumn	20
Ligation-Precolumn	5 798
Ligation-Postcolumn	22

Table 3.4 Assay of radiolabelled attached during ligation

This experiment was designed to measure the amount of anchor And 3 was attached to plasmid DNA after a ligation reaction. This was used as an indicator of ligation efficiency.

5 µl of labelled And3 was added to 5 µl of 10 µM CFB and incubated at 65°C for 5 minutes, then slowly cooled to 0°C. The oligonucleotide solution was then added to a ligation reaction containing about 100 ng linearised, filled in pLysSV40.gpt. After ligation, products were separated using a Chroma-400 spin column. 2 µl aliquots of products from before and after column treatment were counted in a scintillation counter as described.

As a control, labelled oligonucleotide that had not undergone ligation was counted before and after column purification.

Counts were taken in duplicate and the results averaged and tabulated.

These results indicate that approximately twice as much radiolabelled anchor primer goes through the column after a ligation to plasmid DNA (0.4% of initial counts) than without ligation (0.2% of initial counts).

A tenth of each ligation was then transformed into TG1 cells. Aliquots of these transformed cells were plated onto LB amp$^+$ plates. Colony counts are shown in Table 3.2.

These results indicate that filling-in of the restriction site is taking place, since the filled-in plasmid is being rendered at least two orders of magnitude less efficient at self-ligation.

The next part of the protocol to undergo investigation was the ligation of the oligonucleotide anchor primer. Since this had already been partially investigated using PCR in Section 3.5 (Fig 3.12), a different, more quantitative approach was adopted. The effectiveness of end-labelling was tested first. Anchor primer was end labelled as described, then precipitated with ethanol and 5 M ammonium acetate to remove any unincorporated radioactivity from the anchor. Incorporation was tested by placing 2 µl aliquots of oligonucleotide onto DE-81 paper and then comparing unwashed filters with those washed in 0.5 M NaH$_2$PO$_4$ (Section 2.16.7). The results can be seen in Table 3.3.

The washed and unwashed filters give counts of the same order of magnitude, yet with the washed filter surprisingly being slightly higher. This discrepancy is probably due to inaccuracies in the pipetting of sample onto the filters. This error could be eradicated if the same filters were counted, washed, then counted again. This is feasible because no scintillation fluid additions are made prior to counting.

This experiment seems to indicate that all of the label present after precipitation has been incorporated onto the primer.

The ligation step itself was then examined. The Chroma-400 spin column separates DNA molecules of 400 bp and over from smaller species. Labelled oligonucleotide ligated to a large DNA molecule will pass through the column whilst unligated oligonucleotide will be retained. The fraction of label that passes through the column is, therefore, a

Sample	Counts	% Through Column
And 3: Precolumn	6800	
And 3: Postcolumn	150	2.2
Ligation-no enzyme pre-column	5400	
Ligation-no enzyme post-column	220	4.1
4°C: Pre-column	8300	
4°C: Post-column	300	3.6
16°C: Pre-column	17000	
16°C: Post-column	500	2.9
RT: Pre-column	5500	
RT: Post-column	220	4.0

Table 3.5 Radiolabel incorporation during ligation at various temperatures

This experiment was designed to follow on from that described in Table 3.4 by assessing whether the temperature of ligation had any effect on its efficiency. The experiment was undertaken in the same way as described for Table 3.4. Ligations were performed at three temperatures, 4°C, 16°C and Room temperature (ca 22°C).

Two controls were included. Labelled And 3 was counted before and after column treatment and a ligation reaction was performed at room temperature with no ligase addition.

Counts were taken in duplicate and the results averaged before being tabulated. The fraction of label incorporated was calculated by dividing the number of counts that passed through the column with the number added to the column.

These results seem to indicate that little ligation is occurring, irrespective of temperature, because no more labelled anchor is going through the column after a ligation than in the controls, when no ligation has occurred.

measure of the fraction of oligonucleotide that has ligated to larger DNA molecules.

Prior to the ligation, labelled, ethanol-precipitated anchor was mixed with an equal amount of CFB coupling primer and incubated at 65°C for 5 mins, before being cooled slowly to 0°C. This was designed to facilitate any anchor/coupling primer interaction that might be important in the ligation. The And 3 / CFB mix was then added to a ligation reaction with digested, filled-in pLysSV40.gpt plasmid.

The Chroma-400 spin column separates fragments of DNA from oligonucleotides. The ligation products were then run through a column so that unligated oligonucleotides remained in the column, whilst plasmid DNA was eluted with TE. As a control, labelled oligonucleotide alone was also run through a column. Aliquots of sample before and after column purification were counted in a scintillation counter (Table 3.4).

About 0.07% of unligated oligonucleotide counts go through the column, whilst 0.15% of ligation counts go through. This seems to indicate that about 0.1% of the labelled oligonucleotide anchor is ligated onto DNA during the anchor PCR experiment.

Another similar experiment was performed to see what effect temperature has on the efficiency (Table 3.5). The ligation reactions did not result in an increase in incorporated counts above those observed in control reactions. These results are somewhat erratic and seem to indicate that very little ligation is occurring. The temperature of the ligation appears to have little effect on its efficiency.

Each reaction contained about 100 ng of plasmid DNA. A 6 kb plasmid has a molecular weight of about 3.5×10^6 Daltons. This means that each reaction contains about 3×10^{-14} moles of plasmid. Each plasmid has two ends at which the anchor can ligate, so there are 6×10^{-14} moles of ligatable ends. Each reaction also contains 7 µl of 10 µM anchor primer. This is

about 7×10^{-11} moles. Hence there is a molar excess of primer of about 1000-fold. If all the available plasmid Bgl II ends were occupied by ligated oligonucleotide anchor, then this would only represent 0.1% incorporation of the available anchor. This would be extremely difficult to detect using this method, because the background of unligated primer passing through the column is relatively high. Other methods of separation, such as electrophoresis, might be more effective for this purpose. Extreme care would also have to be taken experimentally to establish a signal distinct from background noise. These possibilities were not pursued in this work because of time constraints.

If less labelled anchor primer were added, thereby decreasing the excess of anchor, a greater proportion of available anchor would be incorporated. This might be easier to detect, but might still be masked by the background, since the actual numbers of counts would be smaller. Alternatively, more plasmid might be used, providing more ligatable ends to which anchor can attach.

These experiments seem to indicate that the ligation efficiency is not as low as first thought. The molar excess of labelled primer tends to mask what ligation is occuring as only a very small percentage of the total anchor is involved. This makes it technically difficult to distinguish even relatively high levels of incorporation from background 'noise'.

A more effective way to assess ligation efficiency may be to use PCR, as shown in Figs 3.12 and 3.18. This is difficult to quantify but can at least confirm that some ligation is occuring. Another approach would be to run an electrophoresis gel and ascertain by autoradiography whether the labelled oligonucleotide anchor was located in the slower migrating plasmid DNA band or as a faster migrating oligonucleotide band.

3.7 Redesign of Primers for Anchor PCR

The ligation reaction is inefficient and this may contribute to the rearrangements and mispriming that can be seen in the cloned PCR products, and in the failure to produce clear PCR products. In order to try to improve the PCR reaction, a new anchor primer was designed. This primer was designated And 4N (Appendix II.5) and was intended to have certain advantages over the previous anchor primer, And 3.

As a 46 mer, And 4N might prove to be a better substrate for ligase enzyme than the smaller And 3 20mer anchor, thus improving the efficiency of the ligation reaction.

Using And 4N as an anchor allows a nested amplification to be used at the anchor. Two 20 mers, And 5 and And 6 (Appendix II.5), are complementary to And 4N and can be used sequentially as nested PCR primers. To do this, unligated And 4N must be removed after the ligation reaction so that it cannot itself prime a round of PCR.

A third feature of the And 4N primer is the presence of an Eco RI site 3' to the sequence which is complementary to And 6. No restriction site is present on the PCR primers, and only those products that have been correctly amplified from an And 4N template should contain an Eco RI site.

A 23 mer coupling primer designated CFB 3 was used with And 4N. This was to be removed after ligation, using spun column chromatography.

3.8 Assaying the Ligation Efficiency of And 4N Anchor Primer

The first experiments carried out using the redesigned anchor were intended to show that ligation was taking place and see if it was an
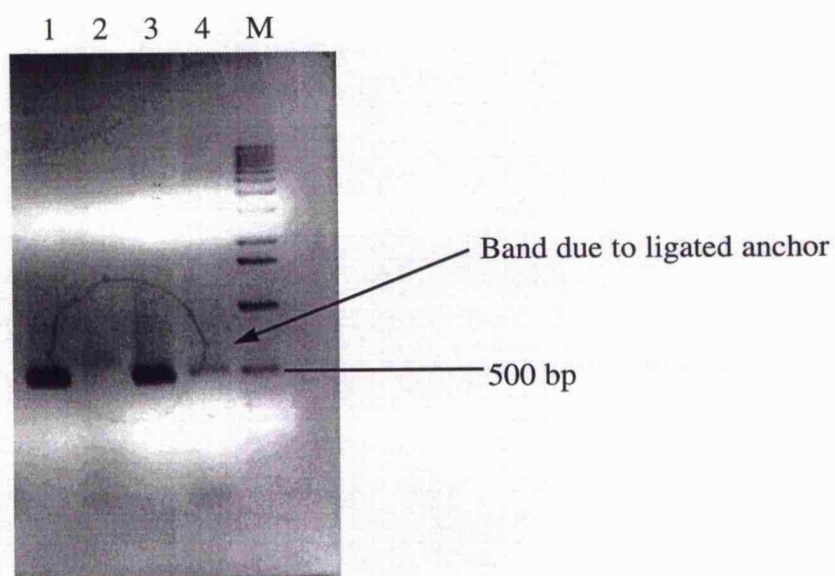
Fig 3.18 PCR assay of ligation efficiency
PCR on pLysSV40.gpt linearised with Bgl II and on linearised plasmid ligated
to And 4N anchor primer (see Section 2.17).
Track 1 - Gpt 9 / Gpt 11 amplification on pLysSV40.gpt template
Track 2 - Gpt 9 / And 5 amplification on pLysSV40.gpt template
Track 3 - Gpt 9 / Gpt 11 amplification on pLysSV40.gpt ligated with And 4N
Track 4 - Gpt 9 / And 5 amplification on pLysSV40.gpt ligated with And 4N
Marker - 1 kb ladder

improvement on And 3. This was done using the two approaches described in Sections 3.5 and 3.6

Plasmid pLysSV40.gpt DNA was prepared by Bgl II digestion and filling in. The ligation was performed by heating the reaction to 65°C prior to enzyme addition and then allowing it to slowly cool. Ligase was then added and the reaction left at 16°C overnight. The ligation was assayed by comparing the efficiencies of PCR using a specific primer, Gpt 5 and anchor primer, And 5, with PCR using Gpt 5 and a primer specific to the plasmid sequence just before the Bgl II site where the anchor is ligated, designated Gpt 11 (Appendix II.2). The amount of PCR product should correspond to the amount of template, and so the proportion of plasmid to which anchor is ligated can be assessed. The results of this experiment (Fig 3.18) indicate that ligation is taking place as expected, although it is still inefficient since the anchor primer band is much less dense than the plasmid primer band.

A second approach to the assessment of ligation efficiency was to measure the retention of end-labelled And 4N. As before, And 4N was kinase end-labelled and purified by ethanol precipitation. After the ligation, plasmid DNA was separated from unligated anchor using a Chroma-100 spun column. Counts were measured in a scintillation counter (Table 3.6).

These results are difficult to interpret because the background, as indicated by the control experiments, is high compared to the experimental results. Some incorporation may be occurring at room temperature, although pipette inaccuracies or background fluctuation could also be responsible for this result. None of the ligation reactions show significantly more label incorporation than the no-enzyme control. Any incorporation that may be occurring is lost in this high background.

Sample	Counts	% Incorporation
4°C: Pre-column	2400	
4°C: Post-column	380	15
16°C: Pre-column	3200	
16°C: Post-column	360	11
RT: Pre-column	1700	
RT: Post-column	400	24
No Enzyme: Pre-column	1700	
No Enzyme: Post-column	260	15
And4N: Pre-column	2700	
And4N: Post-column	210	9

Table 3.6 Ligation efficiency with And 4N

This experiment was essentially identical to that shown in Table 3.5, except that the long anchor primer And4N and CFB3 were used instead of And3 and CFB.

The actual experiment was undertaken in the same way as described for Table 3.4. Ligations were performed at three temperatures, 4°C, 16°C and Room temperature (c 22°C).

Two controls were included. Labelled And3 was counted before and after column treatment and a ligation reaction was performed at Room temperature with no ligase addition.

Counts were taken in duplicate and the results averaged before being tabulated. The fraction of label incorporated was calculated by dividing the number of counts that passed through the column with the number added to the column.

The higher figure for the room temperature experiment may indicate that some ligation may be occurring, or may be a statistically insignificant variation from the background, control values.

Chroma-spin columns may not be efficient enough at separating oligonucleotide from DNA for this application.

Of the two ligation assay techniques, the PCR based approach seems to be the more effective in being able to show that some ligation is occurring. This protocol could be developed into a quantitative technique. Direct densitometric comparison of the anchor and internal product bands might provide a very approximate quantitative result.

Having adjusted and experimented with the the ligation conditions to no avail, the technique was repeated. Despite the poor ligation quality, if the PCR itself is efficient, and a sufficient quantity of template is used, then the target may still be amplified.

3.9 Anchor PCR with the And 4N Anchor Primer

100 μg of C6 DNA was digested with Bgl II , filled in and ligated with And 4N. The ligation was then purified to remove unligated anchor by passing it through a Chroma-100 spin column twice. A PCR was performed using ca 10 μg of template with the primers SV40 9 and And 5. The products of this reaction were run on an LMP gel and a gel slice containing DNA of 600 - 800 bp excised. The DNA was extracted from the agarose using gelase, ethanol precipitated, and resuspended in 10 μl TE.

This SV40 9 / And 5 product was then used as a template for a nested PCR reaction using And 6 and SV40 6. The products of this PCR were visualised on an agarose gel (Fig 3.19a), which was then Southern blotted and probed with kinase end-labelled SV40 8 primer (Fig 3.19b). It can be seen that the PCR product of this reaction is a smear which hybridises to an internal primer over a wide size range.

Since this smeared product contains specific sequence from at least one end of the expected product it may contain correctly primed PCR

product. Alternatively, the smeared nature of the autoradiograph over a large size range might argue for another mixture of misprimed products containing various jumbled arrangements of primers and plasmid sequence. To resolve this issue, the product was cloned and sequenced.

Although an SV40 6 primer with an engineered Eco RI site had been designed, the cloning strategy for this product employed the pGEM-T vector, which had been used successfully in previous cloning experiments. The SV40 6 primer without the restriction site could, therefore, used in these reactions.

The SV40 6 / And 6 product was run out on an LMP gel and DNA from 400 - 600 bp excised in a gel slice. This was extracted with gelase and an aliquot containing ca 20 ng was ligated with 50 ng of pGEM-T vector. The ligation was performed overnight at 16°C. 2 µl aliquots (10 ng vector) of the product were then transformed into DH10 cells. Transformants were plated onto LB amp^{+} plates containing α-complementation reagents.

Upon incubation only about 1% of colonies on these plates were blue, while each plate contained about 500 white 'recombinant' colonies. About 120 of these colonies were picked off onto a filter on an LB amp^{+} plate, from which colony lifts were performed. After lysis, denaturation and fixing, the lifts were probed with the end-labelled oligonucleotide SV40 8. Colonies were identified, from the autoradiographs, that hybridised strongly to the probe.

In order to confirm the presence of a cloned PCR product in these transformants, minipreps were made of the plasmid DNA. These were digested with Eco RI. The presence of a single site (Fig 3.20), indicates that only one copy of the anchor primer is present.

Upon digestion with Sal 1 and Nco 1, the insert is removed from the plasmid (Fig 3.21). Although both the ethidium bromide stained gel (Fig 3.21) and subsequent Southern analysis (Fig 3.22) seem to indicate that
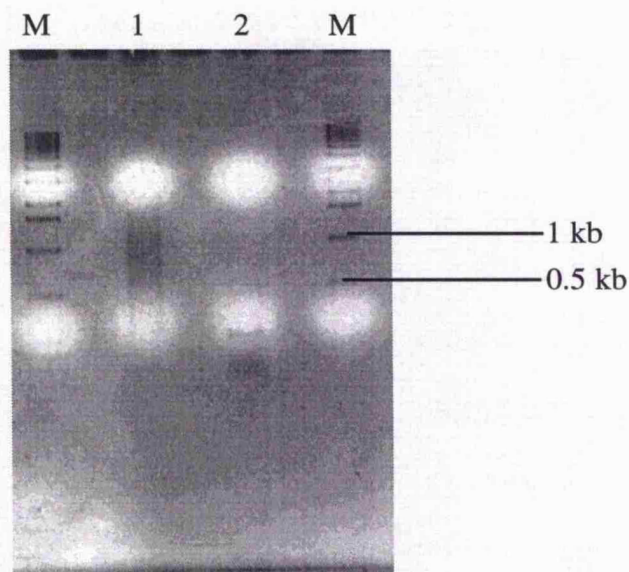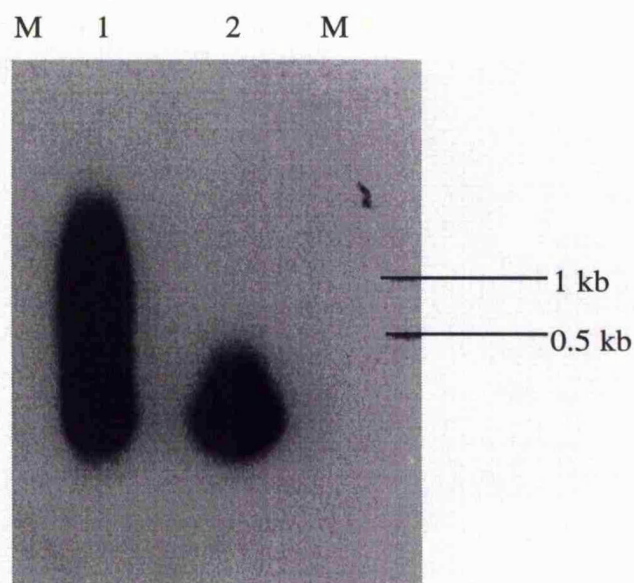
Fig 3.19a Ligation mediated anchor PCR on C6 genomic DNA
1% agarose gel showing the products of two separate nested PCR reactions (Section 2.17).
The two reactions have produced markedly different products because of the small differences
inevitable in the set up of independent reactions.
Tracks 1, 2 - SV40 6 / And 6 amplification of SV40 9 / And 5 PCR product template
Marker - 1 kb ladder

Fig 3.19b Southern analysis of ligation mediated anchor PCR on C6 genomic DNA
Autoradiograph of Southern blot of agarose gel shown in fig 3.19a hybridised with end-labelled
SV40 8 (See Section 2.16). The membrane was washed in 0.1 x SSC at 35°C for 30 mins.
Autoradiograph left at -70°C for 12 hours with an intensifying screen.

this digestion was only partial, they also indicate that SV40 8 hybridises to the insert. This, combined with the results of the digests, indicates that both ends of the PCR product are present in these clones and so the complete product should be intact. However, since the expected size of the product is about 400 bp and the insert here is about 200 bp, the cloned sequence may not be the expected product.

The insert was sequenced using the 'Sequenase' kit. The sequence is shown in Fig 3.23. It can be seen that the cloned product has the SV40 6 primer at one end and the And 6 primer at the other, as would be expected from a correctly primed PCR product. However, all the sequence in the clone between these two primers can be identified as SV40 sequence. Although this sequence presumably derives from the integrated plasmid, it is of little interest, since it is already well known. There is no sign of any other sequence in this clone that may have come from the genomic region flanking the integrated plasmid, which would be of great interest.

This clone is therefore of little use in furthering the aims of the project. It is not known how this clone has been generated, since it is not the product of a correctly functioning anchor PCR. A tantalising clue is the presence of an 8 bp repeat of And 6 sequence at the 3' end of the clone. This may be the residue of a recombination event that may have produced the truncated clone.

3.10 Conclusions

Several attempts have been made to amplify and clone flanking region sequence using ligation-mediated anchor PCR. Results have been strikingly similar for C6 and D8. On agarose gels, the products have been smears rather than bands, also evident on autoradiographs. These smears hybridise to oligonucleotide probes from sequence downstream of the PCR
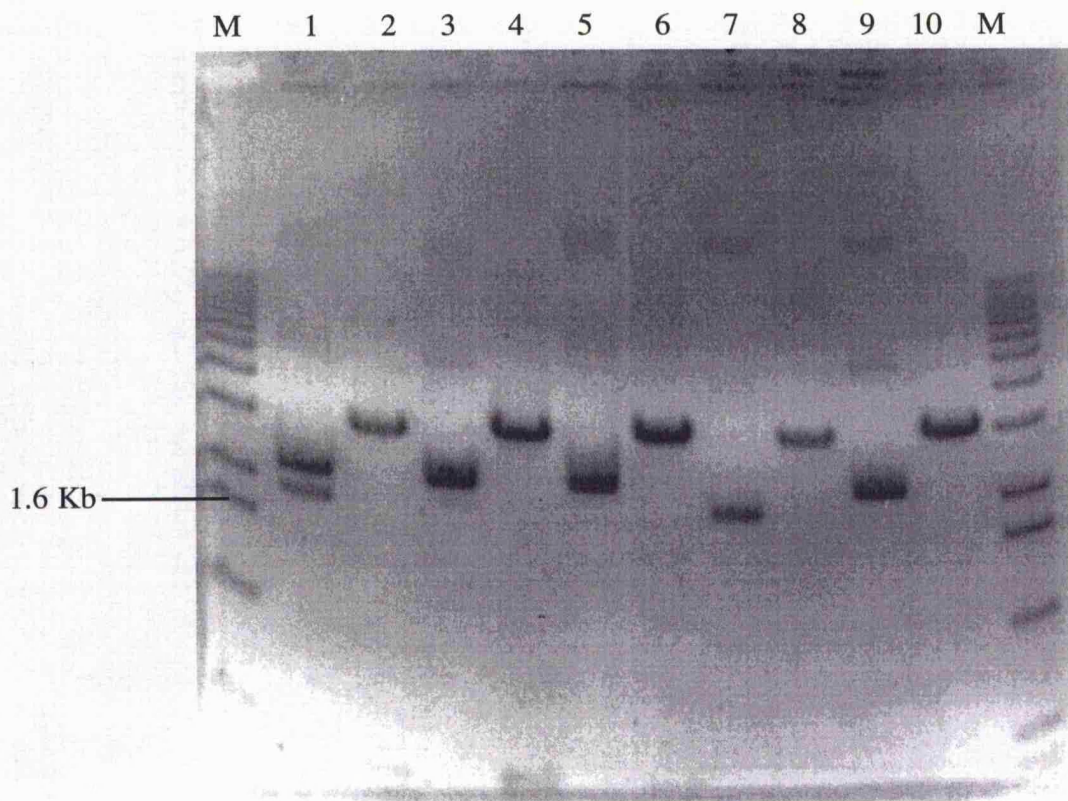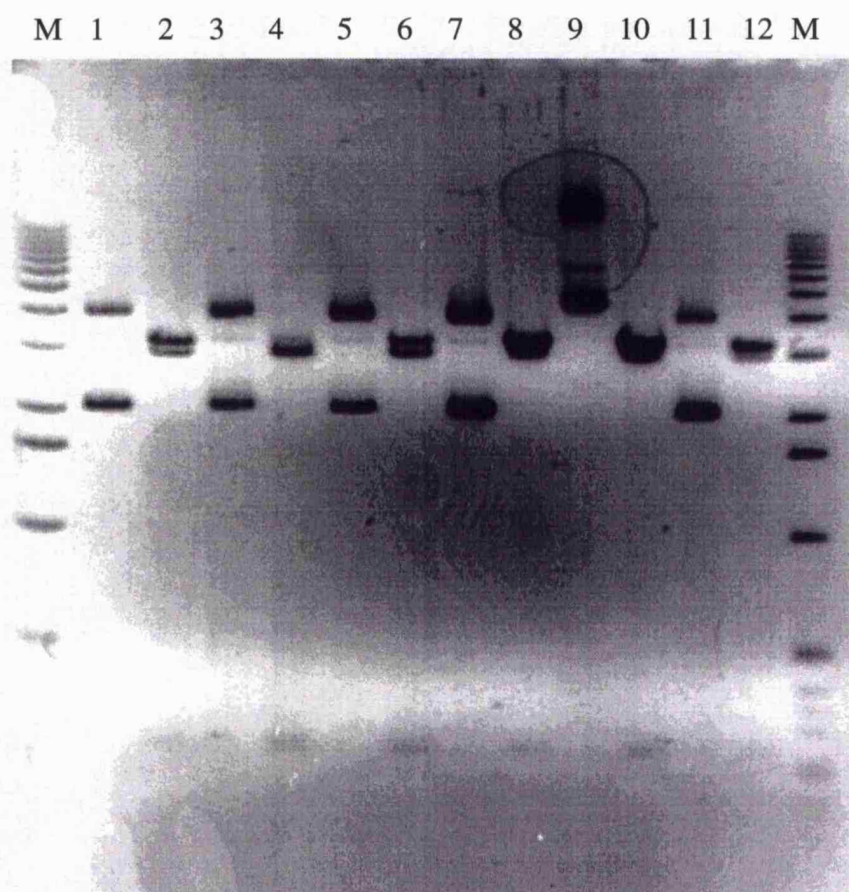
Fig 3.20 Restriction endonuclease analysis of PCR product
1% agarose gel showing five plasmid preparations from five transformants containing cloned PCR product. The uncut plasmid appears to be predominantly in the supercoiled state. Eco RI can be seen to linearise the plasmid in each case.
Tracks 1, 3, 5, 7, 9 - Uncut pGEM-T plasmid with insert
Tracks 2, 4, 6, 8, 10- Eco RI cut plasmid

Fig 3.21 Restriction endonuclease analysis of cloned PCR products
1% agarose gel showing six plasmid minipreparations of DNA from transformants
containing cloned SV40 6 / And 6 PCR product in pGEM-T. The double bands
indicate incomplete digestion. The insert sequence appears to be about 200 bp.
Tracks 1, 3, 5, 7, 9, 11 - Uncut plasmid
Tracks 2, 4, 6, 8, 10, 12 - Nco I, Sal I double digest
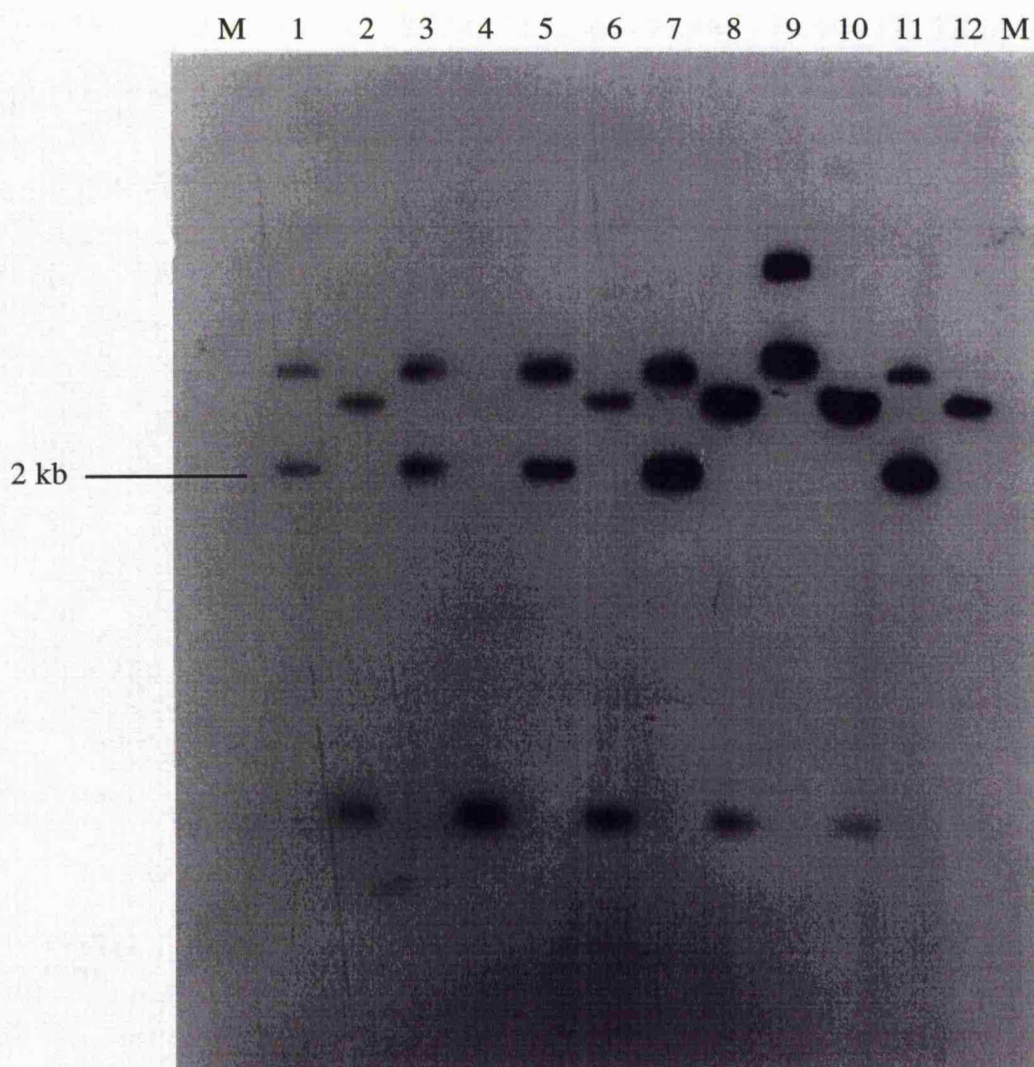
Fig 3.22 Southern analysis of cloned PCR products
The gel shown in fig 3.21 was Southern blotted and the membrane hybridised with
end-labelled SV40 8 as described in Section 2.16. This confirms the incomplete
digestion, as the probe has hybridised to the 3 kb bands as well as the 200 bp insert.

The membrane was washed in 0.1 x SSC for 30 mins at 35°C. The autoradiograph
was exposed for 30 mins at -70 ^{O}C with an intensifying screen.

Fig 3.23 Sequence from Cloned C6 SV40 6 / And 6 PCR Products

Sequence data produced from the SV40 6 / And 6 PCR product cloned into pGEM-T (Promega), using the 'Sequenase' sequencing technique with plasmid based M13 sequencing primers. Also shown is the plasmid SV40 sequence as a comparison. Discrepancies between the two (Bold) are probably due to errors incorporated by the taq polymerase.

Sequence of SV40 6 / And 6 Clone

```
                                              SV406-->
GACGTCGCATGCTCCCGCGCCGCCATGTCCGGCGGGATTCAGGCATAGAG

                  SV408-->
TGTCTGCTATTAATAACTATGCTCAAAAATTGTCTACCTTGAGGTTTTTA


ATTTGTAAAGGGGTTAATAACCAATATTTGATGACTAGTGCCTTGACT

          <-------------SV40 Sequence ------------------->
AGAGATCATATCAGCCATACCACATTGTAGAGGTTTTACTTGCTTT

          <--Rpt -->          And 4N
AAAACTCCATAGCTCCATAGCAGAATTC
```

Plasmid SV40 Sequence [van Heuverswyn et al. 1978]

CAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTA

CCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATG

TATAGTGCCTTGACTAGAGATCATAATCACCCATACCACATTTGT

AGACGTTTTACTTGCTTTAAAAAACC

primer and so contain at least some product correctly primed from the specific primer. However, after cloning, all products examined were smaller than expected and frequently contained bizarre arrangements of primer and plasmid sequence. No sequence that could not be attributed to either plasmid or primer was found.

There are two possible causes for these truncated PCR products. The region beyond the plasmid sequence, the unknown flanking region, might be inherently unstable and undergo rearrangements to leave truncated products. Cloning into a prokaryotic host-system might exacerbate this. A high molecular weight, correctly primed PCR product may have undergone either immediate rearrangement, or it may have been successfully ligated into the pGEM-T vector. Rearrangement may then have occurred after transformation to produce the truncated products.

The high molecular weight of the smeared product on the autoradiographs, which hybridises with plasmid probe, supports this hypothesis. Eco RI digestion does not appear to have an effect on the product size so there does not appear to be a large number of anchor primer copies present. The sequence data might also indicate that a deletion event has occurred. The plasmid vector sequence is intact in all the clones that were sequenced, whilst the putative 'unstable' flanking region has been lost. There is some evidence from the literature to support the idea of eukaryotic DNA sequences that are unstable during PCR [Holstege et al., 1994].

The second possibility is that the PCR reaction is not functioning correctly and no correctly amplified sequence exists to clone. The very presence of anchor primer sequence in the clones seems to indicate a misprimed PCR, rather than extensive rearrangement of a longer product. This is also suggested by the poor primer ligation efficiencies observed, and the smeared PCR products obtained on gels. No PCR product bands were observed from either agarose gel or autoradiograph. Even if the

correct product was produced a great deal of other product, of a wide range of sizes, was also produced. This, in itself, indicates problems with the PCR and may have been a result of too many cycles during the PCR [Bell and DeMarini, 1991].

This does not explain why DNA of a particular size should 'shrink' during cloning. It may simply be due to the continued presence of small amounts of shorter products, after fractionation, which are then more readily cloned than larger products.

The use of recombination deficient strains DH5 and DH10 did not improve the cloning procedure, neither did the use of a low copy number plasmid (pACYC184) for the analogous D8 clone. These still yielded truncated PCR products containing a mishmash of plasmid and primer sequence. This further suggests problems with the PCR itself.

Ligation-mediated anchor PCR is an esoteric and technically difficult approach to the problem of cloning flanking regions and was chosen largely because other, more appropriate methods had already been tried and rejected by other workers [Glassford, 1993]. More refinement of techniques such as plasmid rescue and inverse PCR might have yielded better returns for time and effort, despite the lack of initial promise, since these techniques were better documented in the literature.

Further refinement of the anchor PCR process is also possible. It has been observed that the ligaton step was inefficient. In this setting where the number of templates is limited, even with various enrichment procedures, an efficient ligation is important. A longer coupling oligonucleotide could be used such as CFB 3 (see Appendix II.4). T4 ligase may be more efficient with the larger substrate. A step to remove superfluous oligonucleotide before amplification was already in place to remove And 4N, so, as long as it was unphosphorylated and did not ligate, the coupling oligo, CFB 3, would be removed along with it.

Another refinement is to use much larger pieces of DNA as anchors. These should ligate efficiently, and can be designed to contain a region of non-complimentarity, so that the central portion of the double stranded anchor melts as the temperature increases. PCR primers can then anneal to this melted region where the DNA is effectively single stranded and amplification can occur. This system is known as 'Vectorette' PCR and has been developed commercially by Genosys Ltd.

Other commercial systems based on ligation-mediated anchor PCR have recently been developed. The PromoterFinder and Marathon amplification kits from Clontech provide libraries of genomic DNA or cDNA with adaptor or anchor molecules already ligated to the termini. Specific products can be amplified from these templates using a specific primer and an anchor-specific primer.

The length of the development time of these products, after the initial reports of anchor PCR, indicates the technical difficulties involved. This project has confirmed these difficulties.

After a large amount of work on the PCR cloning approach, it was decided that the construction and screening of a λ-phage library was likely to be more productive in isolating flanking sequence.

Chapter IV

Cloning Strategies Based on λ–Bacteriophage

4.1 Introduction

Previous cloning strategies used in this project to try to isolate flanking sequence were based on delicate and intricate variations of PCR. These proved to be inefficient and unsuccessful and attention shifted to λ-phage cloning, as a straightforward, practical alternative.

Bacteriophage λ is a dsDNA virus of *E. coli*. The 48.5 kb phage genome exists in a linear form in the head of the bacteriophage particle [Summary: Sambrook et al., 1989; Review; Old and Primrose, 1985]. An infectious phage binds to a receptor encoded by the *lam B* gene whose product transports maltose into the cell. This receptor can be induced by adding maltose to the culture medium. This increases the efficiency of phage infection. After infection, when the viral genome has entered the cell, the termini of the linear phage genome, which have 12 nucleotide 5' overhangs, ligate together to generate a *cos* site on a circular genome molecule.

At this stage, there are two life cycles that the phage can adopt. The lytic cycle is characterised by the expression of three sets of genes: early, middle and late. The cycle is established, in preference to lysogeny, by early gene transcription from the P_L and P_R promoters that flank the *c*I repressor. Middle gene transcription then establishes and controls the rolling circle replication that generates more copies of the phage genome. Genome copies are packaged into phage particles by the products of late gene transcription. The host cell then lyses and phage particles are released to infect neighbouring cells. Eventually this leads to the formation of a plaque of cell lysis in a bacterial lawn.

The phage may instead undergo lysogeny, in which the *c*I repressor blocks transcription of the early genes. The phage genome

then integrates at a specific location in the host genome. Only a low level of gene expression is then maintained and the host cell is unaffected. The presence of lysogeny results in turbid plaques on a bacterial lawn, as some cells become lysogenous whilst others lyse.

Because its genetics are well understood, λ-phage has been used extensively as a cloning vector and modified to perform particular tasks.

Simple insertion vectors, such as λgt10, contain a single cloning site into which DNA is inserted. These vectors have a small capacity (ca 7.6 kb) because a viral genome of greater than 105% of the wild-type size is packaged inefficiently.

Replacement vectors enable more DNA to be cloned and work on the principle of removing an unnecessary part of the viral genome, in the form of a stuffer fragment, and replacing it with insert. Since only about 60% of the viral genome is required for the lytic cycle, about 18 kb can be replaced by inserted DNA. Since there is also a minimum size of genome that can be packaged (>78%), re-ligated vector that contains no insert will not be packaged efficiently. This acts as a selection for recombinant phage containing insert sequence.

Markers can be incorporated into vectors to enable recombinant phage to be distinguished from wild-type. A lacZ gene is present at the λgt11 cloning site and non-recombinant plaques form a blue colour in a *lac⁻* host on a medium containing X-Gal (Section 2.15.1).

The cloning site of λgt10 is within the *c*I repressor gene. Recombinants will have no active cI gene and so will form clear plaques distinguishable from the turbid, lysogenous parental plaques. In *hfl* host-strains, the non-recombinant vector forms lysogens with high efficiency because the *c*I activator, *c*II is highly expressed. Since these lysogens are immune to further infection, low levels of lysis occur, and small, turbid plaques form. When the *c*I gene is inactivated by the insertion of foreign DNA, the

bacteriophage lytic cycle is unhindered and plaques appear. Only recombinant phage form plaques that can be screened.

Wild-type λ-phage grows slowly in lysogenic cells containing the P2 prophage. This is the Spi$^+$ phenotype. The genes *red* and *gam* have been shown to be responsible for this growth inhibition. Vectors have been designed that carry either the *red* and *gam* gene on a stuffer fragment. Recombinant phage have this stuffer replaced by an inserted sequence, are Spi$^-$ and grow well on P2 lysogen hosts. Since the wild-type phage does not grow, the recombinant is selected [Loenen and Brammar, 1980].

The *red* gene product is involved in the resolution of viral DNA, replicating in the θ-form, into closed circular molecules. These molecules can then become the substrates for rolling circle replication, prior to packaging.

The *gam* gene product inactivates exonuclease V, whch is encoded by the host cells *recB* and *recC* genes. This exonuclease degrades the linear concatamers that are preferentially packaged so that only rare DNA molecules produced by the less efficient q-replication can be packaged. Unless the infecting bacteriophage is *gam*$^+$, efficient propagation can only occur in *recA*$^+$ hosts. This may lead to recombination events in the insert if the foreign DNA contains repetitive elements. Vectors such as the Charon 32-35 contain the *gam* gene to overcome this problem.

Recognition sequences for the host *recBC* enzyme are known as *chi* sites. These are not present in the wild-type λ-phage, but when present in mutants, they overcome the small plaque size phenotype produced by *red*$^-$ *gam*$^-$ phage in wild-type *E. coli*. This is because of an increase in the efficiency of formation of closed circular dimers from q-replication. More molecules are present to be packaged and normal sized plaques are formed. Most λ-phage vectors are *red*$^-$*gam*$^-$ in the absence of stuffer fragments, and so recombinant plaques have a small phenotype. If the

-100-

inserted DNA contains a eukaryotic sequence that mimics a *chi* site, then these plaques will grow to a higher titer and become over-represented in library amplifications. The problem can be overcome by using vectors containing a *chi* site (eg λDASH or λFIX), where all plaques are of similar size.

The λgt10 vector [Huynh et al., 1985] used in this project is an insertion vector that can accommodate 6 kb of inserted sequence in an Eco RI cloning site. Selection of recombinants can be made on hosts with the *hflA* mutation, since recombinants are cI^- whilst non-recombinants are cI^+ and lysogenise efficiently. Both wild-type λ-vector and recombinant are red^+gam^+ and can be propagated on $recA^-$ hosts.

The host used in this project was not the C600*hfl* strain recommended by the supplier, which is commonly used for cDNA cloning, but the strain SRB. This was because this is a recombination deficient strain commonly used in the construction of genomic libraries. It was hoped that genomic inserts might be more stable in this host.

4.2   Commercial Library Screening

Leading on from the work on C6 using anchor PCR, it was decided to screen a commercial genomic library using the heterogenous PCR product obtained from C6 (Section 3.4). It was thought that flanking sequence from C6 was contained in the PCR products and similar sequences could be isolated from the library without generating or characterising a *bona fide* PCR product. This screen was carried out in parallel with studies on the C6 using PCR (Section 3).

The commercial library that was screened contained mouse genomic B6/CBA F1J DNA from the female spleen in λFixII. It had an average insert size of 9 - 23 kb. The probe used was the heterogeneous PCR product

from C6. Although much of the expected sequence of this product was plasmid, it was hoped that the flanking sequence also present would hybridise to sequence in the library. So long as some correct product is present and labelled in the probe, the heterogeneous nature of the product does not matter, as false-positives derived from misprimed PCR products can be identified and discarded later.

About 60 000 plaques from the library were screened, in duplicate. None hybridised to the probe.

There are two reasons why no positive clones were identified. Firstly, very few clones were screened: the number of clones that have to be screened to have a given probability of finding a clone can be calculated from the formula shown [Clarke and Carbon, 1976]:

$$V=\ln(1-p)/\ln(1-f) \qquad \text{p : probability of success, f : ratio of insert size to genome}$$

To have a 99% chance of identifying a clone, assuming that the mouse genome contains 3×10^9 bp and the average insert size is 1.6×10^4 bp, 9×10^5 clones need to be screened. Only about 7% of the required number of clones were screened. From this number, the chances of identifying a clone are only about 30%.

Subsequent sequencing of cloned PCR products indicated that the probe contained little or no flanking sequence. This made it ineffectual as a probe. The number of clones screened with this sequence is largely irrelevant, as nothing would have hybridised to a probe containing only plasmid-derived sequence.

It was decided to postpone the screening of this library until cloned flanking sequence was available. This would give a much better chance of success because cloned flanking sequence would be a more effective probe. To this end, attention switched, first to refinement of anchor PCR, then to

λ-library construction and screening, with the intention of returning to the screening of this commercial library with any suitable sequence obtained.

4.3 Strategies for Library Construction

Despite a large amount of development work, anchor PCR proved to be unsuitable for generating flanking sequence which could be used to screen libraries and so the method was abandoned (Chapter 3).

An alternative approach was devised that employed more conventional cloning techniques. A λ-library was constructed using genomic DNA from either D8 or C6. This library was then screened using plasmid sequence to identify vectors with inserts containing both plasmid and flanking sequence. Suitable clones could then be isolated and characterised.

To construct this library, a simple cloning strategy was designed using information already gleaned about the insertion sites in each cell line.

The DNA sequence to be cloned was the 4 kb Eco RI restriction fragment in D8. This particular fragment was chosen because it was an ideal size to be accommodated in the λgt10 vector, which has an Eco RI cloning site.

To directly isolate such a clone would involve the screening of a large number of plaques. If the DNA was partially purified prior to cloning then the number required could be brought down to a more manageable size. This was achieved by initially digesting the D8 genomic DNA with Nco I and then gel-eluting DNA of around 10 kb (Fig 3.3a). A second digest was performed on the size-fractionated DNA with Eco RI. In this way, only 4 kb Eco RI fragments within 10 kb Nco I fragments were cloned. This reduced the number of clones to be screened.

Three λ-libraries, designated A - C, were constructed using this strategy. These are described in Sections 4.4 to 4.6

4.4 Library Construction A

This section describes the initial construction of a library, using the strategy outlined in Section 4.3.

D8 genomic DNA was prepared for cloning by digesting 100 μg of DNA with Nco I in the presence of 1 mg/ml BSA. After incubation for 5 hrs at 37°C, the digested DNA was precipitated with ethanol and resuspended in a volume of TE sufficient to enable it to be loaded onto a 1% agarose gel. The sample was resolved by electrophoresis at 100 v for 3 hrs alongside DNA size-markers. Using these markers, a gel slice corresponding to 10 kb was excised from the gel and the DNA extracted using the Qiaex kit. This DNA was digested with Eco RI and fractionated a second time on a 1% agarose gel. This DNA was extracted a second time using Qiaex and then phenol extracted and ethanol precipitated, prior to ligation with 0.25 μg λgt10 vector arms. After incubation overnight at room temperature, half of the ligation was packaged using commercial λ-packaging extract. The phage particles thus formed were suspended in 0.47 ml SM buffer with 10 μl chloroform and stored at 4°C.

The efficiency of packaging was established using positive control DNA (λgt11), with host strain C600*hfl*. This was shown to contain about 10^7 pfu/μg, which is the lower limit recommended by the suppliers.

Although the control DNA was packaged efficiently, a measurement of the actual library titre was required. The library needed to contain sufficient clones to have a reasonable chance of containing the clone of interest, otherwise it would not be worth screening. Hence, a 1 μl aliquot of library was diluted into 99 μl SM buffer and a 10 μl aliquot of this

-104-

diluent was shown to contain 16 plaques. From this, it can be readily calculated that the library contains 1.6×10^5 pfu/ml, or 8×10^4 clones in a total volume of 500 µl. This can be expressed as 6.4×10^5 recombinants/µg vector.

From the Clarke and Carbon equation:

$$N=\ln(1-p)/\ln(1-f) \qquad \text{p=probability of success, f=ratio of insert size to genome}$$

the number of clones to be screened can be calculated.

About 3×10^6 clones would have to be screened if no enrichment were carried out. Enrichment of about ten-fold occurs on each gel fractionation. This is an estimate based on the amount of material excised and the amount left. Over two fractionations, this amounts to a 100-fold enrichment. This means that 3×10^4 clones need to be screened. The present library, therefore, appears to have only about a quarter of the members required for a 99% probability of isolating a clone.

The library was screened as indicated in Section 2.18.6 using C600*hfl* host and LB medium. Initially, a PCR product from the Gpt 11/ Gpt 9 primers was used as a template for the probe. This showed a large number of positives which were presumed to be false - since a small number of plaques are being screened for a single sequence, obtaining more than one positive is very improbable.

The PCR product Lys 5/Gpt 6 was then used as a template for radiolabelled probe. This proved to be a more reliable probe. Plaques that had apparently hybridised to the Gpt 9/Gpt 11 probe did not hybridise to the new probe, apparently confirming that they were not true positives.

From this library, 2520 plaques were screened. None of these plaques hybridised to the Lys 5/Gpt 6 probe. The number of plaques screened was very low, well below that suggested by the initial titre. The reason for the disparity between the initial titre measurement and the actual titre is unclear, but may have been due to the the conditions under which the

library was stored prior to screening. The absence of a particular sequence was unsurprising in view of the calculated number of clones required to be screened compared to that actually screened.

The success of the approach is heavily dependent on the number of clones screened. The next stage of the project was to construct a second library containing sufficient clones to make screening for single-copy sequence worthwhile.

4.5 Library Construction B

A second library was constructed after the failure the identify any clones in the first.

180 µg of D8 genomic DNA were prepared for cloning by digestion with Nco I, fractionation, digestion with Eco RI and further fractionation. The DNA was prepared for ligation using the Qiaex kit and ligated to 0.25 µg λgt10 vector arms. Ligated vector DNA was packaged using a commercial kit and the titre of the packaged phage assessed using SRB host cells.

A titre of about 100 pfu/µl was measured for this library. Since a total of 500 µl was present, the library contained about 5×10^4 clones. For a 99% chance of successfully identifying a single copy sequence, 3×10^6 clones would have to be screened (Clarke and Carbon equation; Sections 2.18.1, 4.2 and 4.4). If the enrichment process carried out here purifies the target sequence 100-fold, then 3×10^4 clones need to be screened ($3 \times 10^6/100$). The actual amount of enrichment that occurred was difficult to quantify and this enrichment value was only an estimate.

This library was apparently sufficiently large to have a high probability of containing the sequence of interest. However, although

4×10^4 plaques were screened using the SV40 2/5 probe, no positives were identified.

In order to identify a clone containing the sequence of interest more clones needed to be screened. This would require improvements to be made in the library construction process to increase the titre, so that there would be sufficient clones present for there to be a strong probability of the desired clone being found. The more clones that are screened, the higher the probability of success.

The ligation and packaging were checked using a positive control DNA insert (pSP64 plasmid). This yielded about 1.2×10^7 pfu/µg which is within the manufacturers guidelines. This indicated that these steps in the protocol were not contributing to the low titre and the problem lay with the insert DNA.

4.6 Library Construction C

Previous libraries had titres which were too low to offer a reasonable chance of finding a particular single sequence. The reason for these low titres was unclear but since the control experiments indicated that the packaging was working efficiently, it seemed reasonable to conclude that the DNA preparation steps were responsible and so the protocol was simplified. A series of pilot experiments were carried out to see if the titre could be raised in this way.

D8 genomic DNA was digested with Eco RI restriction endonuclease and then fractionated on a gel prior to ligation with vector. When the titre was being assessed using SRB host, after packaging, it was observed that the negative control, containing no insert, produced more plaques than the ligation that did contain insert DNA (ca 6×10^4 pfu/µg). This insert

DNA did not appear to ligate at all to the vector arms, making it useless for library construction.

The simplest possible method of producing insert DNA is to digest it with Eco RI and clone without fractionation. This was tried with an initial 12 µg of DNA, of which aliquots from about 0.25 µg to about 1 µg were ligated with 0.25 µg of vector arms. This proved to be much more successful, with an average of 5×10^5 pfu/µg. Further experiments raised this to 1.5×10^6 pfu/µg. Despite this great improvement, this library was not screenable because this DNA was unfractionated, and a large number of clones would have to be screened to have a realistic chance of finding a particular sequence.

From the Clarke and Carbon equation in Section 4.2 and 4.4, the number of clones that have to be screened can be calculated. This calculation indicated that there would be ca 3×10^6 plaques to screen, for a 99% chance of successfully isolating a single copy sequence. This was too many to be technically feasible, given available facilities.

An enrichment step must, therefore, be included in the protocol. A 20-fold enrichment would mean screening only 10^5 clones - a feasible number.

From the initial work, it seemed apparent that the fractionation process using agarose gels and Qiaex kits played a significant part in making the insert DNA 'unclonable' and decreasing the titre of the resultant library.

A new strategy was adopted, whereby the D8 DNA was digested with Nco I and fractionated, as before, but then eluted from the gel using the gelase enzyme. The DNA was then digested with Eco RI, prepared using a 'Magic' Clean Up column and ligated in as short a time as possible. This would prevent any possible degradation or damage to the ends of the DNA.

Using this insert DNA, a library was made with a titre of 5×10^5 pfu/µg in the SRB host strain. In the same experiment, positive control DNA yielded a titre of 5×10^6 pfu/µg whilst a negative control, containing no insert DNA had a titre of 5×10^4 pfu/µg. This titre was considered high enough to be worth screening.

As a rough empirical guide to the proportion of plaques due to wild-type λgt10, an aliquot of the library was plated out onto BBL medium in 0.65% agar BBL, using C600*hfl* as host. 55 plaques were visible, none of which had the distinctive small, turbid character of a wild type λgt10 plaque - all were large, clear recombinant plaques.

The library was plated out using SRB host at a density of ca 20 000 plaques per 152 mm plate. These plaques were lifted, fixed and hybridised in duplicate, alongside positive controls consisting of dot blots of pLysSV40.gpt plasmid DNA.

A total of 1.9×10^5 plaques were screened with a radiolabelled pLysSV40.gpt probe. Four putative positives were identified from the first round screening. These were picked off, suspended in SM buffer and used to reinfect more host SRB in a second round screen.

Only one of the putative positives identified produced positive plaques in the second round screen. Between 3% (9/300) and 6% (17/300) of plaques from this suspension of first round plaque were positive.

Plaques identified from the autoradiograph as hybridising to the probe were picked off for a third round screen. The third screen was carried out at low plaque density and, as 25% - 33% of the plaques were positive, single positive plaques could be identified and picked off. These were stored at $4^{\circ}C$ in SM buffer containing chloroform. A final fourth round screen was subsequently performed to further purify these plaques. The results (Table 4.1) indicated that almost all the plaques present hybridised to the probe. Thus the plaque suspensions in SM buffer are nearly homogeneous

Plate	No of plaques	No of +ves	% +ve plaques
1	202	201	100
2	52	41	79
3	50	49	98
4	18	17	94
5	16	14	87

Table 4.1 Proportion of plaques hybridising to probe in the fourth round screen

These results indicate that a large majority of plaques on each of the fourth round screen plates contain the cloned sequnce of interest.

and therefore suitable for either large scale purification (Section 2.18.7) or insert amplification by PCR.

The cloned sequence was partially characterised by the use of different radiolabelled probes. The filters used in the screening were stripped and reprobed with short sequence elements derived from the pLysSV40.gpt plasmid. It was found that plaques that hybridised to probes derived from the whole plasmid sequence did not hybridise to a probe from the lysozyme gene (PCR product: Lys 2/3) or to a probe containing lysozyme and Gpt sequence (PCR product: Lys 5/Gpt 6). The plaques did, however, hybridise to a probe from the Gpt region (PCR product: Gpt9/11). This indicated that the only plasmid sequence in the clone was Gpt sequence. No other sequence from the plasmid was present.

4.7 Investigation and Analysis of Cloned Sequence

Initially, it was necessary to determine that the 4 kb restriction fragment that was the cloning target was actually present in the D8 DNA preparation used to make the library. An aliquot of D8 genomic DNA was assessed by Southern blotting and hybridisation to plasmid sequence (Fig 4.1). This autoradiograph indicates a band of lower molecular weight than expected from previous restriction analysis.

Further analysis of the clone was required in order to establish the significance of this result. Attempts to make a large scale preparation of the λgt10 clone were unsuccessful and produced yields that were much too low to extract DNA. An alternative approach was to investigate the clone by PCR.

The clone was analysed extensively by PCR. Oligonucleotide primers designed to amplify across any λgt10 insert were used along with the Gpt 11 primer. A product was observed (Fig 4.2) only with the two λgt10
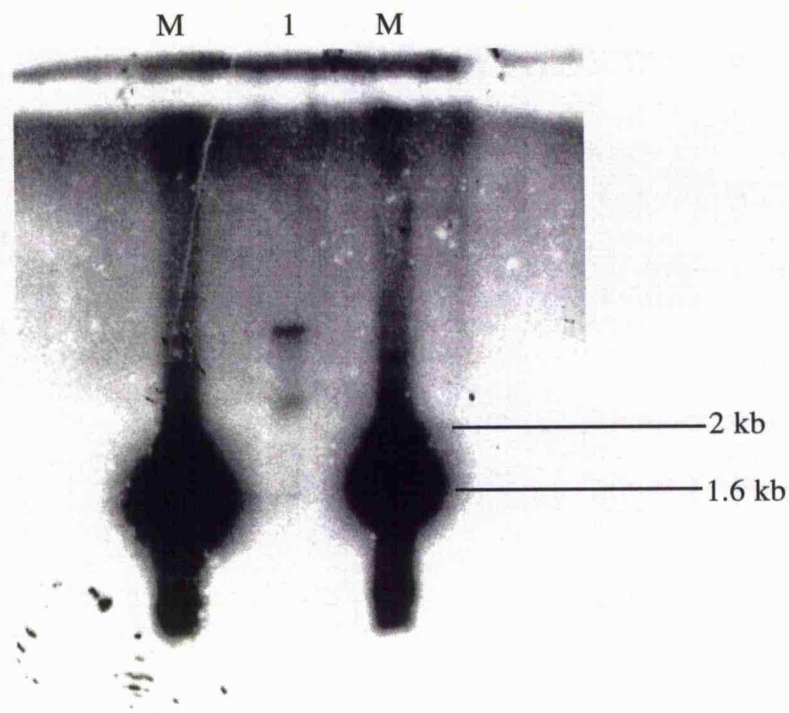
Fig 4.1 Southern analysis of endonuclease treated D8 genomic DNA
Autoradiograph of Southern blot of D8 genomic DNA digested with Nco I
and Eco RI. The probe used was randomly primed ^{32}PdCTP labelled
pLysCMW.gpt plasmid sequence. Autoradiograph was developed for 72
hours at -70°C with an intensifying screen.
M - 1 kb marker
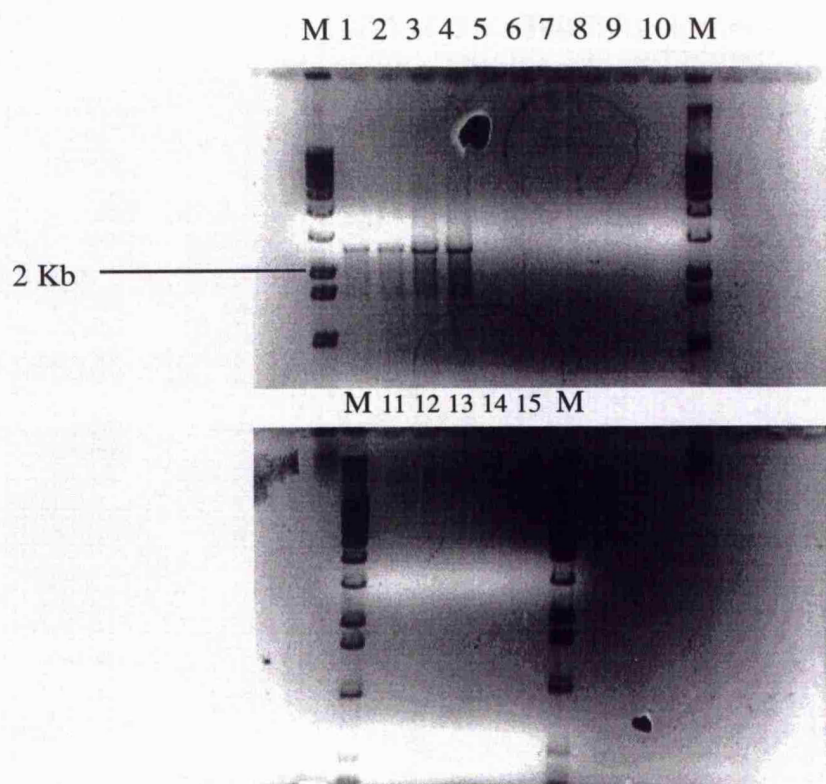
Fig 4.2 PCR on four λ plaques with putative inserts
PCR was carried out on the supernatant of four agar plugs containing plaques
suspended in SM buffer, that appeared to hybridise to the probe in plaque
hybridisation experiments (Section 2.16). The cloned sequence in each of these
plaques was assumed to be identical. The following amplification reactions are
shown;
Tracks 1 - 4, λgt10F / λgt10 R
Track 5, λgt10F / λgt10 R : No template
Tracks 6 - 9, λgt10F / Gpt 11
Track 10, λgt10F / Gpt 11 : No template
Track 11 - 14, λgt10 R / Gpt 11
Track 15, λgt10 R / Gpt 11 : No template
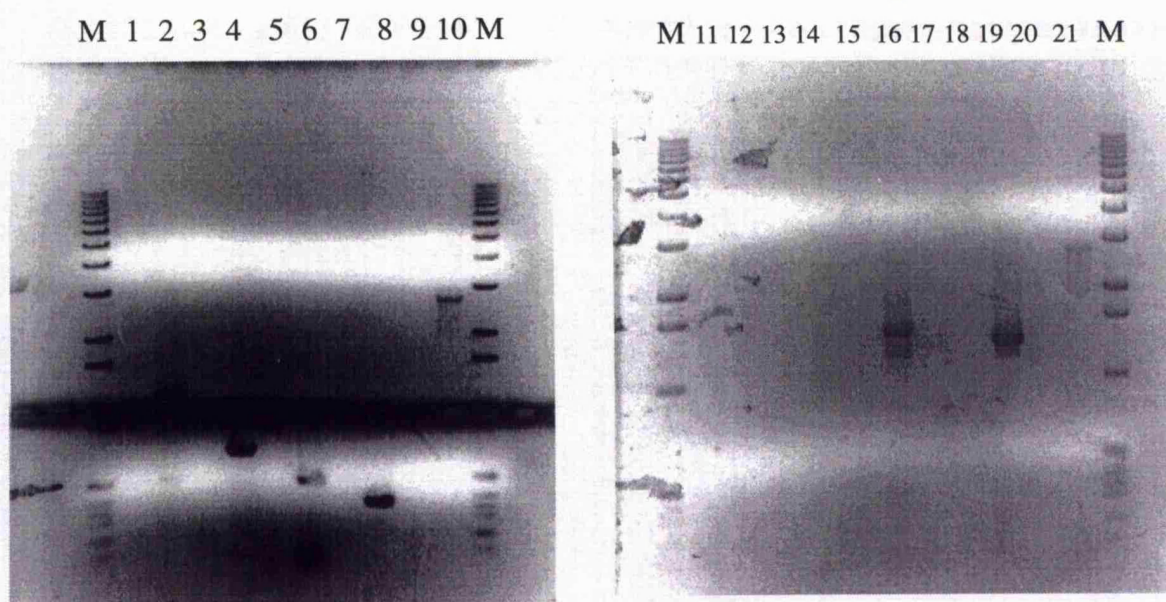
Fig 4.3 PCR analysis of λ phage containing insert

PCR was carried out on one of the plaque suspensions shown to contain an insert (See Fig 4.2),
using a range of primers as described in Section 2.17. Primer locations and orientations are
shown in fig 3.3b and in Appendix II.

Track 1, λgt10F / Gpt 8

Track 2, λgt10R / Gpt 8

Track 3, λgt10F / SV40 9

Track 4, λgt10R / SV40 9

Track 5, λgt10F / SV40 6

Track 6, λgt10R / SV40 6

Track 7, λgt10F / SV40 10

Track 8, λgt10R / SV40 10

Track 9, λgt10F / λgt10F

Track 10, λgt10R / λgt10F

Track 11, λgt10F / Lys 5

Track 12, λgt10R / Lys 5

Track 13, λgt10F / Lys 4

Track 14, λgt10R / Lys 4

Track 15, λgt10F / Gpt 2

Track 16, λgt10R / Gpt 2

Track 17, λgt10F / Gpt 1

Track 18, λgt10R / Gpt 1

Track 19, λgt10F / Gpt 9

Track 20, λgt10R / Gpt 9

Track 21, λgt10F / λgt10R

primers indicating an insert size of about 2.5 kb. This is smaller than the restriction fragment that was the target of the cloning, although it should be noted that there was no size fractionation of the final insert DNA used. Further PCR analysis indicated the presence of Gpt and SV40 sequence in the clone (Fig 4.3).

Having ascertained that the clones did indeed contain the plasmid sequence, the λgt10 F/R primer PCR product was cloned into pGEM-T vector. During this cloning procedure sequence information was obtained using the 'fmol' and 'Sequitherm' cycle sequencing kits. The quality of this sequence data was poor compared to 'Sequenase' data which could be obtained from cloned sequence.

Problems were encountered in the PCR cloning as a high number of false-positives were obtained in the α-complementation screen. Putative clones were finally obtained using DH10B host cells. These were picked off onto filters, lifted and hybridised with radiolabelled SV40 2/5 PCR product. Clones that hybridised to the probe and, therefore, contained the PCR product, were replated for further manipulation.

These clones were examined by restriction mapping. No sites in the insert were found for Kpn I, Nco I, Bgl II, or Pvu II (Fig 4.4). This indicated that the Gpt gene in the clone does not extend as far as the documented Kpn I site. There did, however, appear to be a Pst I site in the clone. This was situated about 800 bp from the plasmid site, about 750 bp from the end of the insert. A simple restriction map of the clone could thus be drawn.

From the cloned insert, high quality sequence data was obtained using the 'Sequenase' system, using the λgt10 R and F PCR primers as sequencing primers (Fig 4.5). Using this sequence, and the PCR and restriction data, a map of the insert and its orientation in the vector was determined (Fig 4.6)
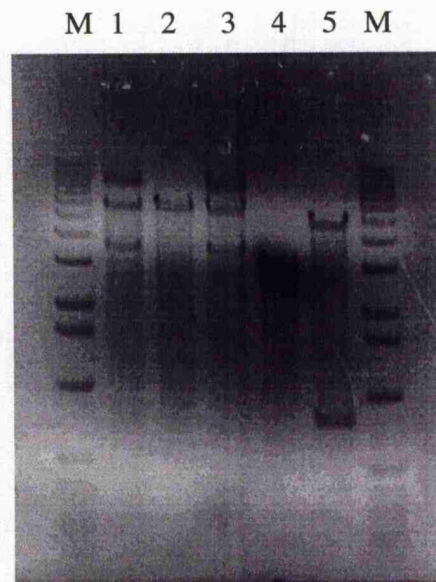
Fig 4.4 Restriction analysis of insert cloned into pGEM-T plasmid
Plasmid minipreps of pGEM-T containing insert amplified from λ phage and cloned as a
PCR product, were analysed by digestion with various restriction endonucleases as
described in Section 2.13.
Track 1 - Kpn I
Track 2 - Nco I
Track 3 - Bgl II
Track 4 - Pvu II
Track 5 - Pst I

```
ns2   1             tcttnatttc tgggatctga cacctcacac agagatacat caggtcagaa

      51            caccaatgta tgtaagacaa aaatacatat attttaaaaa gaaagcaaag

      101           agaaatatgt tggtttagaa taaaatgaaa aaccgaagcc agttgtcaga

      151           atcctattaa tgctttcttt gcatttgcat ttgccaggac tatatttcag

      201           ttttgtttta tttattattt tttattaaga tttcattgtg tgtgtatgat

      251           gtcttacctg catgtatttc catatacagg gtgaatatct ggtatctgt


                              NS5
ns4  1            gcctacccac tctgcctttt tgtctcttat tcaactatat ttagaaaaat

      51            gcaagataac aagtatttag aaatatcaca tgatttcata gatgatatgt

      101           tatgcatatt aaagagagtg gataca
```

Fig 4.5 Sequence from unknown region of D8 sequence cloned into λ phage

The positions of these stretches of sequence is shown in Fig 4.6.
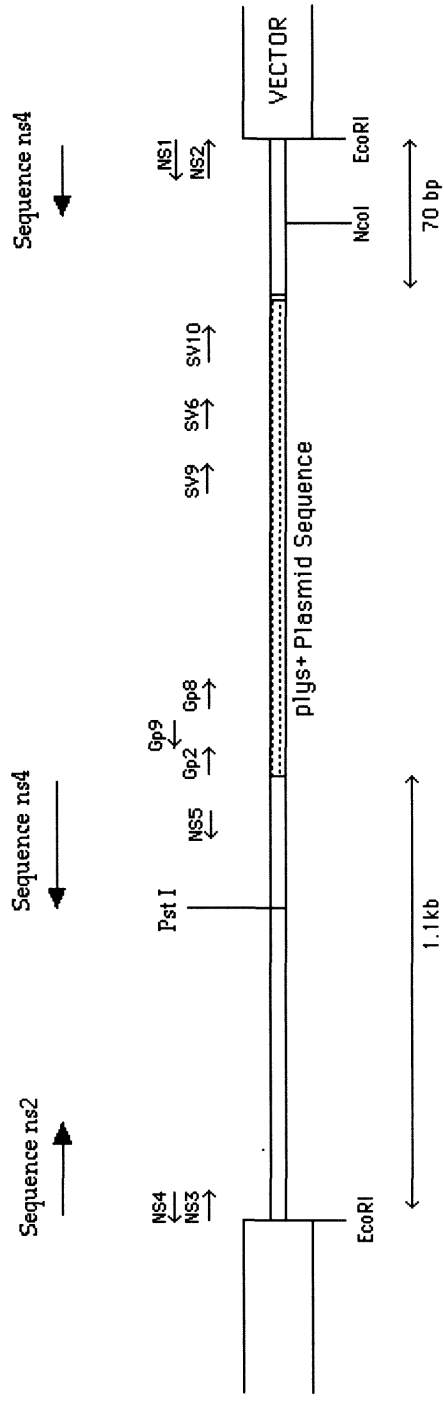
Fig 4.6 Diagram of Insert
Diagram of the cloned genomic fragment showing plasmid and unknown regions and postions of primer

4.8 Analysis of D8 using Sequence from the Clone

A clone was isolated from a genomic library that contained some sequence that did not originate from the integrated plasmid. The clone containing this sequence was not of the size and arrangement indicated by previous knowledge of the insert site. It was necessary to find out whether this unknown sequence originated in the same locus as the plasmid integration or was completely unconnected.

From the sequence data obtained, oligonucleotide primers were designed. These were complementary to the ends of the unknown, non-plasmid sequence in the clone (Fig 4.6) and designated NS1 - 5. These were then used in PCR experiments on genomic DNA preparations to try to establish the location of the cloned sequence in the cell line ie whether it was in normal PCR range of the integrated plasmid.

A PCR using NS3 and NS5 on the D8 genomic DNA seemed to indicate that this was a contigious sequence in D8 (Fig 4.7). Further experiments seemed to indicate a distance of some 4 kb between NS2 and NS5 in the J558L cell line from which D8 was derived (Fig 4.8).

A number of experiments were then carried out to establish whether a PCR product could be amplified from D8 DNA using a plasmid primer and a primer from the identified clone sequence. If a product could be successfully amplified, then the unknown sequence from the clone must be close to the integrated plasmid in the D8 genome. Results (Fig 4.9a, b and 4.10) indicated that products could be amplified from these reactions. Although the relationship between the arrangement of sequence in the clone and that in the D8 genome proved to be more complicated than expected, a simple map of the insertion was then drawn. This map allowed the positions of the unknown sequence primers to be orientated with respect to the position of the plasmid insert (Fig 4.11).

-112-

M 1 2 3 4 5 6 7 M

1 kb

0.5 kb

Fig 4.7 PCR on plasmid and D8 genomic DNA with NS 3 / NS 5 primers from un identified sequence

PCR was carried out using NS 3 and NS 5 primers, from the unidentified region, on the pGEM-T plasmid containing the insert and on D8 genomic DNA as described in Section 2.17.

Track 1 - 1 µg genomic template

Track 2 - 100 ng genomic template

Track 3 - 10 ng genomic template

Track 4 - 100 ng plasmid template

Track 5 - 10 ng plasmid template

Track 6 - 1 ng plasmid template

Track 7 - 100 pg plasmid template
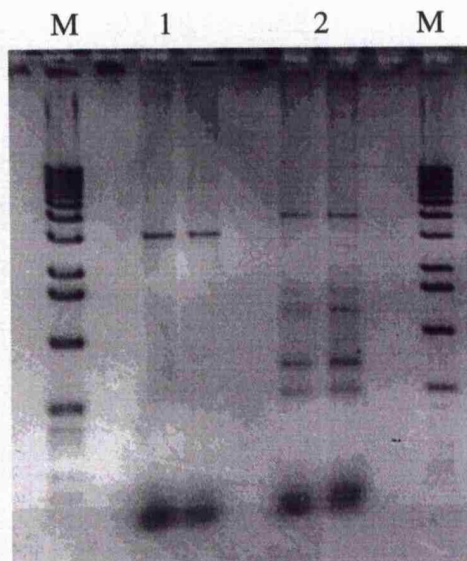
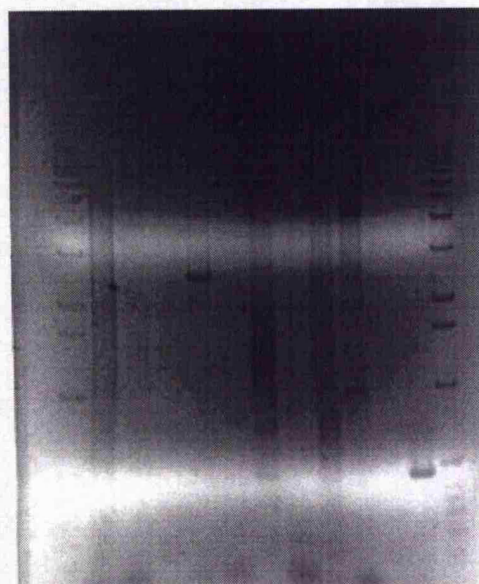**Fig 4.8 PCR amplification with NS 2 / NS 5 primers on J558L DNA**
PCR amplification was carried out on 500 ng of J558L genomic DNA template
using two sets of primers as described in Section 2.17.
Track 1 - NS 2 / NS 4
Track 2 - NS 2 / NS 5
Marker - 1 kb ladder

Fig 4.9a PCR amplification of D8 template

1% agarose gel showing PCR amplification products from 500 ng of D8 genomic
DNA template as described in Section 2.17. The following primers were used in
each reaction ;

Track 1 - NS 1 / SV40 8   Track 7 - NS 4 / SV40 8

Track 2 - NS 1 / pB5   Track 8 - NS 4 / pB5

Track 3 - NS 2 / SV40 8   Track 9 - NS 5 / SV40 8

Track 4 - NS 2 / pB5   Track 10 - NS 5 / pB5

Track 5 - NS 3 / SV40 8   Track 11 - Gpt 9 / Gpt 11

Track 6 - NS 3 / pB5

1 2 3 4 5 6 7 8 9 10 11 M
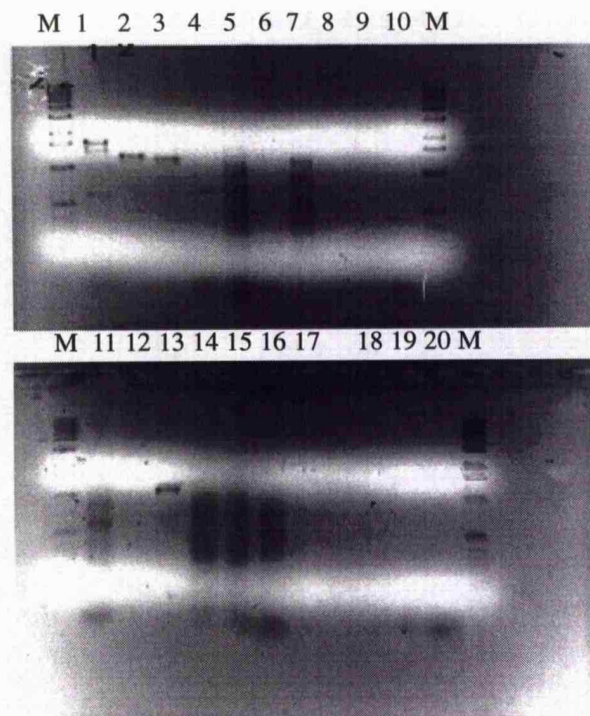
4 kb
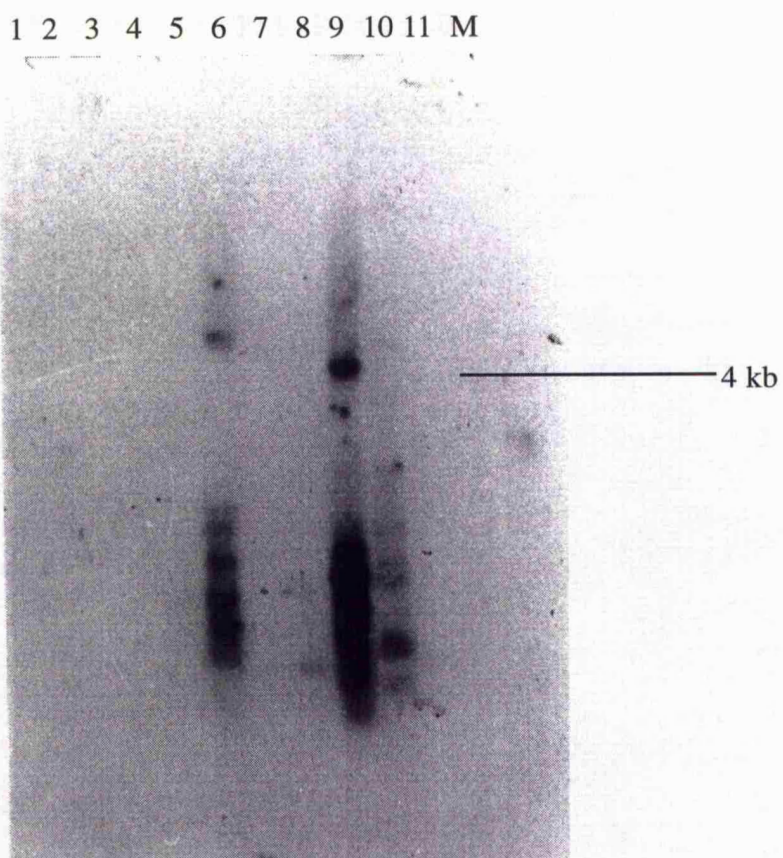
Fig 4.9b Southern analysis of PCR products from fig 4.9a
Autoradiograph of Southern blot of gel shown in fig 4.9a, probed with radiolabelled NS 3 / NS 5
PCR product. Membrane was washed in 0.1 x SSC at 37°C for 30 mins. Autoradiograph was
developed for 48 hours at -70°C with an intensifying screen (See Section 2.16) ;
Tracks - As for fig 4.9a.

Fig 4.10 Assessment of various templates for PCR amplification

1% agarose gel showing PCR products from amplifications of genomic preparations of D8, a mutant line of D8 and the J558L line. These were carried out as described in Section 2.17.

Tracks 1 - 5, D8

Tracks 6 - 10, mutant D8

Tracks 11 - 15, J558L

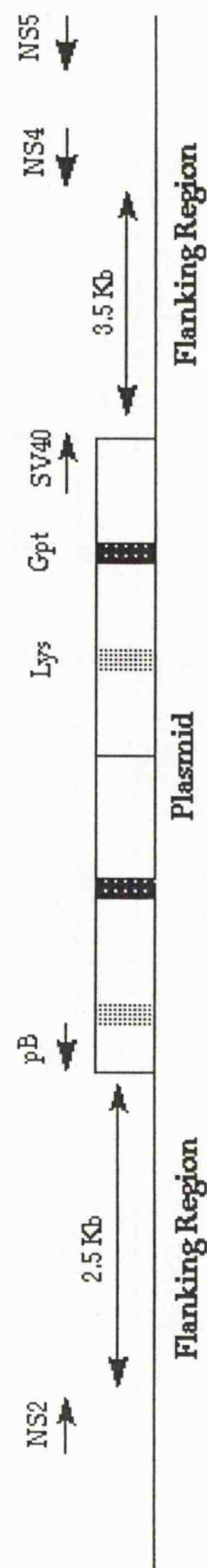Tracks 16 - 20, No DNA

Primers ; 1 - NS 3 / Gpt 3

          2 - NS 3 / Gpt 9

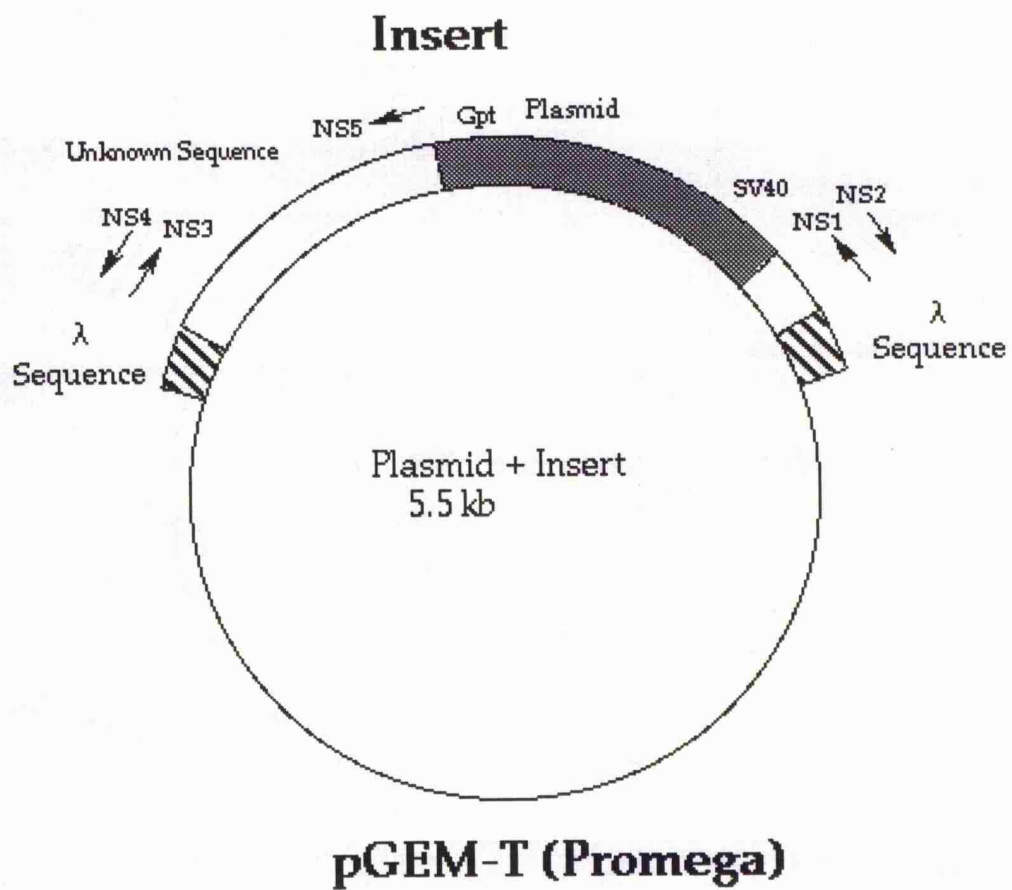   3 - NS 3 / Gpt 10

          4 - NS 1 / SV40 4

          5 - NS 1 / SV40 6

Fig 4.11 Model of D8 Insertion Site According to PCR Data

This map is derived from the data shown in Figs 4.9 a and 4.10 and shows the possible locations of the unknown sequence primers, with respect to specific sequence primers, in the D8 integration site.

Insert

Unknown Sequence

NS5

Gpt

Plasmid

SV40

NS2

NS1

λ Sequence

NS4 NS3

λ Sequence

Plasmid + Insert
5.5 kb

pGEM-T (Promega)

Fig 4.12 pGEM-T plasmid containing insert
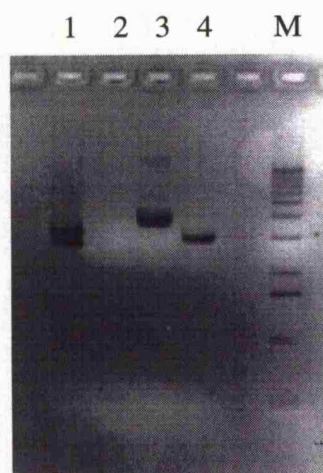
Fig 4.13a Southern analysis of plasmid DNA
1% agarose gel showing various plasmid preparations. This gel was Southern blotted and hybridised with PCR product.
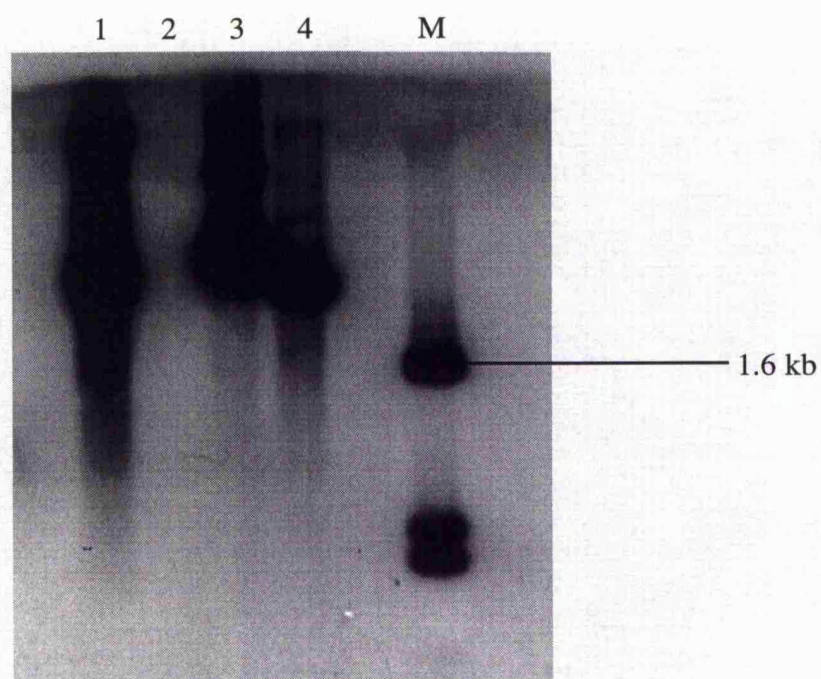Track 1 - Eco RI cut plasmid with insert
Track 2 - Blank control
Track 3 - pGEM-T with insert
Track 4 - Linear pGEM-T

Fig 4.13b Autoradiograph of Southern blot of gel in fig 4.13a
Membrane probed with random primed radiolabelled NS 4 / SV40 11 product.
Washes were performed in 0.1 x SSC for 30 mins at 37°C. Autoradiograph was
developed for 6 hours at -70°C with an intensifying screen (See Section 2.16).
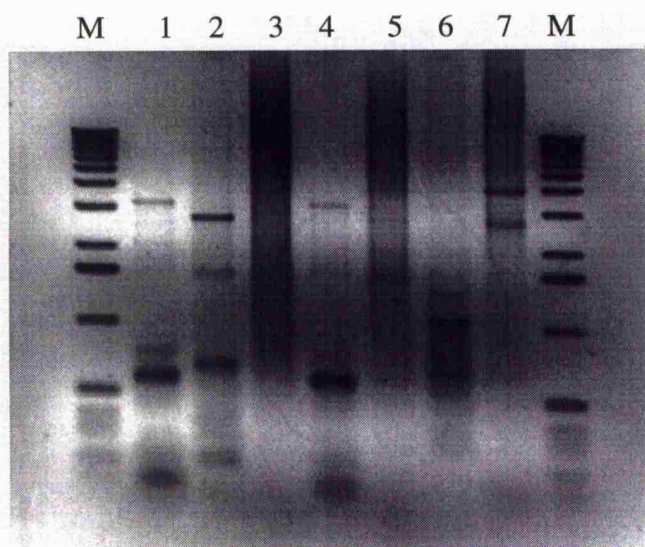
Fig 4.14 Nested PCR on SV40 11 / NS 5 template
 PCR was carried out on two independently amplified SV40 11 / NS 5 PCR prod-
uct templates (1-3 and 4-6) amplified from D8, as described in Section 2.17. The
products were run out on a 1% agarose gel. The following primers were used ;
Tracks 1, 4 - SV40 11 / NS 4
Tracks 2, 5 - pBA / NS 2
Tracks 3, 6 - M13F / NS 5
Track 7 - NS 5 / SV40 11 on uncut pGEM-T with insert
Marker - 1 kb ladder

It proved difficult to produce clean, negative control reactions in these amplifications. This indicated that there was an alternative interpretation of the data. All the PCR product sizes were those expected of amplifications of the pGEM-T plasmid containing the original cloned insert (Fig 4.12). Such contamination would also explain the bands appearing in negative control tracks. If this was indeed the case, then it would not be possible to link the unknown cloned sequence with the locus of the plasmid integration, because the nature of the rearrangement that had produced this cloned sequence, from native D8, was unknown. Southern analysis would provide strong, but not definitive evidence. Two approaches were adopted to establish whether the PCR results are genuine or a result of contamination.

If the PCR bands were due to plasmid contamination then the pGEM-T sequence would be present in the PCR product, instead of D8 genomic sequence. This was investigated by Southern analysis. PCR products on a gel were blotted and probed with radiolabelled pGEM-T sequence. These blots showed specific hybridisation to the probe, indicating that contamination is responsible for the PCR products observed.

To confirm this result, a Southern blot of the plasmid containing original cloned sequence was performed. This was then probed with the radiolabelled PCR product (Fig 4.13a and b). The autoradiograph indicated that the probe hybridised both to plasmid backbone and to insert, confirming the initial finding.

Further confirmation was obtained through nested PCR experiments on the purified PCR products of the initial PCR amplification of D8. Using SV 11/NS 5 template DNA, a product was obtained using NS 2 and pB A primers (Fig 4.14). This amplification would be expected if the PCR product contained the plasmid sequence.

From these results, it seemed reasonable to conclude that the amplification bands observed are caused by contamination with plasmid template. The oligonucleotide primer sequences from the cloned unknown region were, therefore, not likely to be close enough to the plasmid sequence to be able to interact with plasmid primers to generate an amplification product. Another possible alternative was that a genuine amplification product was being masked by the contaminant signal. This is difficult to prove or disprove. Despite taking precautions against contamination and throwing out all PCR reagents when it occurred, it proved difficult to eliminate the contaminant signal.

Southern analysis was used to examine whether the unknown sequence in the clone might be close to the plasmid sequence in the D8 cell line. Digests of D8 genomic DNA were electrophoresed, Southern blotted and hybridised with radiolabelled NS 3/NS 5 PCR product. On the subsequent autoradiographs (data not shown), it proved to be impossible to separate a specific signal from the non-specific background signal. No information could be derived from these experiments because of this technical deficiency.

4.9 Conclusion

A D8 genomic library was constructed in λgt10 vector. Although problems were encountered in obtaining sufficiently high titres, a clone containing plasmid sequence was identified and isolated. The DNA enrichment process was carried out prior to library construction with a particular target fragment in mind. Despite this, the clone obtained was very different from that expected; it was smaller and contained two regions of unidentified DNA flanking a truncated piece of plasmid sequence. It could not be demonstrated that the unknown regions in the

clone correlated with the flanking regions in the native D8 and so plans for further work on the plasmid flanking region were shelved. It would be interesting if future studies could establish the relationship of this unknown sequence with the D8 locus.

Analysis of the clone by PCR, cycle sequencing and further cloning failed to reveal any indication of the nature of the rearrangement that had occurred. Furthermore, there were several possible stages at which this may have occurred.

It is possible that this clone represented a separate locus to that mapped previously, through a second, undetected integration event during transfection. If this were the case, the unknown sequence may be totally unrelated to the flanking region at the high activity locus. It would be difficult to explain, however, why such a secondary integration was not identified during previous work on D8 [Glassford 1993].

It is possible that the D8 line may have undergone a recombination event within the locus of integration, generating the cloned sequence. Although D8 genomic Southern blots performed before cloning do indicate bands of unexpected sizes, it is difficult to imagine how such a rearrangement could still produce intact and functioning genes for both lysozyme and selectable marker. Any rearrangement must have been localised to the particular culture from which the DNA was prepared.

For the clone to be an artifact of the enrichment and cloning process, unidentified restriction sites or enzyme non-specific 'star' activity must have generated a plasmid fragment which then ligated to other fragments prior to ligating to the vector arms. This seems improbable, but the presence of an Nco I site close to one end of the plasmid sequence might suggest such a mechanism, at least for this part of the clone.

The final possibility is that the clone had undergone rearrangement during the process of λ-infection and replication. Although the host cell

line SRB is recombination deficient, the *red* gene product from λgt10 is involved in recombination and may have played a part in rearranging the clone.

It is difficult to establish which of these theories, if any, is correct and the nature of the clone remains a mystery. Without further work to confirm the relevance of the sequence obtained, further progress towards investigating the locus of the vector insertion is impossible. However, the results confirm that the λ cloning approach is the most effective thus far. Despite the use of commercial reagents, it was still difficult to produce a library of sufficiently high titre for screening. Procedures for enrichment of the insert DNA prior to cloning seem to have a deleterious effect on the library titre, presumably affecting the efficiency of ligation.

Interestingly, the technical difficulties that were initially highlighted in this work have subsequently been confirmed and then circumvented by changes in the methods used to enrich the genomic DNA. A partial genomic digest followed by enrichment on a sucrose gradient, rather than an agarose gel, has proved successful in constructing libraries in λ_{DASH} of sufficiently high titre to isolate specific sequence. This has allowed flanking sequence to be cloned from C6 and possibly also D8 [Harrison, Personal communication, 1995]. This confirms the conclusions of this work regarding the feasibility of using λ–cloning techniques to isolate unknown flanking region DNA. Although ultimately unsuccessful, the value of the general approach was demonstrated by the isolation of a potentially interesting clone.

Chapter V

Discussion

5.1 PCR Based Cloning Strategies

The rationale behind this approach lay in previous work performed in
this laboratory [Glassford, 1993]. A number of different cloning techniques
had been tried, none were successful. The most promising approach
seemed to involve the somewhat esoteric procedure of ligation-mediated
anchor PCR. The basis for this conclusion was the generation of smeared
autoradiographs that hybridised with plasmid DNA from downstream of
the PCR primer. None of this 'hybridisable' material had been cloned. The
method offered a means of generating and cloning sequence relatively
quickly, compared to other approaches.

This project initially reproduced this PCR result and proceeded to
clone and sequence products. Despite the size range of the smear in the
autoradiograph, only relatively small products were cloned - far smaller
than the size of the expected PCR product. These proved to be a mixture of
primers and plasmid sequence.

Were these products caused by rearrangements occurring during the
cloning of a genuine PCR product, or were they sequences faithfully
cloned from a mixture of misprimed PCR artifacts? In favour of the
former, the products that entered the cloning process were much larger
than the cloned products. In favour of the latter, while the host strains eg
DH5, DH10 are recombination deficient, the products contained primer
and plasmid sequence that was repeated in different arrangements with no
unknown sequence present. This does not appear to be the product of an
obvious recombination mechanism.

Investigation was carried out into the PCR protocol itself via a
number of control experiments. These seemed to indicate that while the
ligation step was inefficient, the method did work on plasmid DNA. To be

successful on a genomic template, greater efficiency was needed. It is highly doubtful whether this was ever obtained.

The autoradiographs were similar in appearance to those reported elsewhere and attributed to an excess of PCR cycles [Bell and DeMarini, 1991]. However, it is unlikely that this is the cause of these problems since relatively low cycle numbers (25 - 30) were employed.

There are several possible mechanisms by which PCR can produce unexpected artifacts. If the single stranded templates from early PCR cycles possessed secondary structure, a number of PCR problems may develop. An internal loop that does not hinder access to the 3' end may either slow down the polymerase or be digested through its 5'-3' exonuclease activity. If the 3' end of the product anneals within the product itself, it can act as a primer. This creates a primer annealing site on the wrong strand and allows the amplification of products longer than desired with only a single primer.

If there was a high concentration of products, this 3' annealing may have occured between different product molecules. Intra- and inter-molecular annealing can generate both larger and smaller products than expected, which contain specific internal sequence. The presence of product from reactions containing only the anchor primer provides evidence for this mechanism. However, this may have been an artifact of the ligation process. It would have been informative to assess the effect of denaturant agents such as DMSO that diminish secondary structure in the PCR process.

Computer analysis reveals a certain amount of self complementarity amongst the primers, particularly And 4N (Appendix III), but not sufficient to disrupt an amplification.

High product concentrations and long products may also contribute to the preferential reannealing of products in regions close to one end. This

would interfere with the extension of product from the opposite primer. The product might then be degraded, or the polymerase might slow down or jump to the other strand. Such products could also be amplified with a single primer but would leave a smear on an autoradiograph because the point at which the polymerase switched strands would be highly variable.

At high product concentrations it is also possible that product and primer compete for annealing sites. This would result in numerous gap and nick sites that would lead to degradation by the 5' - 3' exonuclease of taq. Smears would be observed when the products were visualised.

The suppression PCR effect is based on some of these effects. Anchor primers ligated to each end of each DNA fragment anneal preferentially in an intra-molecular manner, forming hairpin structures and preventing amplification by anchor derived primers. Only the presence of primers specific for sequence within the fragment of DNA will cause the template to be amplified. In this way spurious, background amplification is minimised, as only the correct template can be amplified [Siebert et al., 1995].

Whether any of these mechanisms, or more prosaic technical deficiences accounted for the failure of this approach is unknown. Any initial promise shown by this technique has been shown to be illusory. To be successful, considerable technical challenges have to be overcome. These challenges appear to have recently been met by several commercial kits based on this approach, such as Vectorette PCR (Genosys) and PromoterFinder (Clontech). It would be interesting to apply these kits to the problems presented by this project.

5.2 λ–Bacteriophage Approaches

The use of λ-phage as a cloning vector is well established. A strategy was devised that centered around cloning a 4 kb Eco RI fragment from D8 into λgt10, a simple insertion vector usually reserved for cDNA work. Partial purification of the genomic fragment, based on restriction maps of the insert site, decreased the number of clones to be screened.

The effectiveness of the approach, particularly when compared to the PCR work, was confirmed by the isolation of a clone that contained both the plasmid and an unknown sequence.

The cloned sequence was different to that expected. Instead of 4 kb, it was 2.8 kb and consisted of ca 1.3 kb plasmid sequence containing both SV40 and Gpt sequence flanked by about 70 bp of unknown sequence at one end and about 1.1 kb at the other.

The unknown regions were partially sequenced but could not be identified through database searches. Attempts to establish the location of these regions in the native D8 genome through PCR and Southern blotting were inconclusive.

An explanation for the presence of this clone has been elusive. There are several possibilities. The cloned sequence may have existed in the cell line before the genomic DNA was extracted. It may represent a second vector integration event independent of that previously mapped which genomic analysis failed to reveal. Alternatively, rearrangment may have occurred in the mapped vector locus. Since cells are grown under selective conditions, it is difficult to see how such a rearrangement could be sub-lethal. There is Southern hybridisation evidence to suggest that sub 4 kb Eco RI fragments containing plasmid sequence were present in the genomic DNA preparation prior to cloning.

The cloned fragment may be an artifact of the cloning process. The presence of an Nco I site close to the end of the SV40 plasmid sequence suggests that at least one end of the clone had been generated in this manner. It is also possible that rearrangment of a larger fragment occurred during λ-phage propagation in the host bacterial strain. The host cells were recombinase deficient and designed for use with eukaryotic genomic DNA, but the product of the *red* gene from λgt10 can catalyse recombination.

It is unclear how the cloned fragment was generated and so it is difficult to establish how useful the unidentified flanking sequences obtained will be in future work. They may be derived from regions unrelated to the plasmid insertion site and therefore of no use in trying to isolate nearby regulatory elements.

5.3 Conclusion

This work contrasts two different cloning approaches. The PCR approach was fiddly, difficult and ultimately, never close to realising its potential. The second approach, based on more traditional cloning methods, succeeded in isolating a fragment composed of plasmid and an unknown sequence. Although the unexpected arrangement of the fragment casts doubt on its usefulness, this project shows the way forward for future work in this area.

More sophisticated vectors capable of carrying larger inserts enable different fragments to be targetted without the need for potentially distorting enrichment procedures. More recent experiments with the λDASH vector have confirmed the potential of this approach. Putative flanking regions from both C6 and D8 have now been isolated from genomic libraries in λDASH. The work described here facilitated this success

by showing that other potential approaches could be discounted and allowing resources to be focussed on λ–cloning stratagies.

Finding an explanation of the high expression of C6 and D8 remains a useful objective. Besides being the subject of curiosity, it has tremendous potential in the development of a useful expression system based on myeloma cells. These cells have many useful attributes for an expression hosts have the potential to generate high levels of recombinant protein. With an 'expression cassette' engineered into the genome, they may represent a considerable improvement on current cell-based systems. However, as the technology develops, cell-based system might be appropriate only for certain proteins that cannot be expressed in a functional form in potentially cheaper or more convenient systems such as bacterial hosts or transgenic animals.

The aim of this project was to isolate and examine regulatory elements. This was never achieved because of technical difficulties. However, considerable progress was made in developing the anchor PCR technique and it is unclear why useful products could not be amplified. λ-cloning was conclusively shown to be a feasible approach to the problem. A clone containing potentially interesting sequence was isolated. Although the authenticity of this clone could not be verified, the groundwork was firmly laid for future investigations using λ-cloning as the method of choice.

Appendix I

Sequence of Clones from C6 PCR

Sequence of four clones of PCR product from C6 Anchor PCR, as described in Section 3.4.

I.1 Clone 1

```
           <---------- Polylinker ----------->>---------Gpt4R1---------->
0          GGGTACCGAGCTCGAGCTCGAATTCGGTGATTGGCTACGGTGGCGGCAAC


                                        >-------SV40  2--->
50         TGGATTTATGAGTGGGCCCCGGATCTTTGTGAAGGAACCTTACTTCTGTG


100        GTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTA


150        AATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGT


200        TTGT------------------------ 80 bp ------------------------GCTAGAAGAATGCCAT

           -------SV40  4------------->                  <-----And3----
300        CTAGTGATGATGAGCTACTGCTGACTCTCAACATCTACTCCTCCATAGCA

           ------Gpt4R1----------->
350        GAATTCGGTGATTGGCTACCGTGGCGGCAACTGGATTTATGAGTGGGCCCG

                         >-------------SV40  2---------->
400        CGGATCTTTGTAAGGAACCTTACTTCTGTGGTGTGACCATAATTGAACAA

                      ---And3---------<>----Gpt4R1-------->
450        AACTACTCCATAGCAGAATTCGGTAATTGGCTACCAT
```

I.2 Clones 2,3 From the M13F Primer

```
                          <-XbaI-><-BamHI-SmaI->     <-SacI---EcoRI-----
0          ATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTCG

           -----Gpt4RI----------->
50         GTGATTGGCTACCGTGGCTCCATAGCATGACGTCGACTTGACCAATTGAA


100        CGATGCCGAACGTACCGAGGATGATTTATAC-----------------------------------

           <-----Gpt4RI----------------->         Plasmid  Sequence ------>
131        GATACGGTGATTGGCCTACCGTGGCGGCAACTGGATTTATGAGTGGGCCC
```

CGGATCTTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAA

ACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGTGT

ATAATGTGTTAAACTACTGATCTGTTGTTGTG

I.3 Clone 4

```
                                            <-SacI-> EcoRI  Bottum
0        ACGTCCAGCCTTGCATGATCTCCTAGGGGCCCATGGCTCGAGCTTAAGAT


         Gpt4RI  ??
50       TGAACTAACCGATGTCTAGCCGAGTGCTATCGATACTGCTAGCTGACGGT


100      TACTTGCTACTGGCTTG---------------------------------------------------------------


                              <--SV40  ?
0        ACTAATGCTTAAGACGATACTCTAGCCCCGGGTGAGTATTTAGTCAATGCGG


         Gpt4RI  ??              And 3        <------SV40  Sequence
50       AGCCATCGGTTAGTGGCTTAAGAGCATACCTCATCAAACAGGTTAATACA


100      GTTGTGGTGTCTTACTTCCAAGGAAGTGTTTCTAGGCCCCGGGTGAGTAT


                      Gpt4 RI                  And 3
146      TTAGGTCAACGGCGGTGCCATCGGTTACCGGCTTAAGACGATACCT
```

Appendix II

PCR Primer Sequences

II.1 *Lys* cDNA Sequence Primers:

[Jung et al., 1980, Glassford 1993]

(* start and end of coding sequence)

```
                        HCMV   Promoter   <-----   5'

                    *       <--- LYS5
  1 TCCCGCTGTG TGTACGACAC TGGCAACATG AGGTCTTTGC TAATCTTGGT

 51 GCTTTGCTTC CTGCCCCTGG CTGCTCTGGG GAAAGTCTTT GGACGATGTG

          LYS4 --->                              <--- LYS1
101 AGCTGGCAGC GGCTATGAAG CGTCACGGAC TTGATAACTA TCGGGGATAC

151 AGCCTGGGAA ACTGGGTGTG TGTTGCAAAA TTCGAGAGTA ACTTCAACAC

201 CCAGGCTACA AACCGTAACA CCGATGGGAG TACCGACTAC GGAATCCTAC

251 AGATCAACAG CCGCTGGTGG TGCAACGATG GCAGGACCCC AGGCTCCAGG

          LYS2 --->                              <--- LYS6
301 AACCTGTGCA ACATCCCGTG CTCAGCCCTG CTGAGCTCAG ACATAACAGC

351 GAGCGTGAAC TGCGCGAAGA AGATCGTCAG CGATGGAAAC GGCATGAGCG

401 CGTGGGTCGC CTGGCGCAAC CGCTGCAAGG GTACCGACGT CCAGGCGTGG

       <--- LYS3   *
451 ATCAGAGGCT GCCGGCTGTG AGGAGCTGCC GCACCCGGCC CGCCCGCTGC

501 ACAGCCGGCC GCTTTGCGAG CGCGACGCTA CCCGCTTGGC AGTTTTAAAC

551 GCATCCCTCA TTAAAACGAC TATACGCAAA CGCC ----->  5' Poly A signal
```

II.2 *gpt* Sequence Primers:

[Richardson et al. 1983]

(* start and end of coding sequence in *E. coli*)

(NB. This sequence in in the reverse orientation to the SV40 Early Poly A sequence.)

```
                         Bgl  II              Gpt11 --->
        TK Promoter <---- 5' AGATCTCTAT AATCTCGCGC AACCTATTTT

                                                          *
   151 CCCCTCGAAC ACTTTTTAAG CCGTAGATAA ACAGGCTGGG ACACTTCACA

                              <--- GPT6
   201 TGAGCGAAAA ATACATCGTC ACCTGGGACA TGTTGCAGAT CCATGCACGT

                              <--- GPT1
   251 AAACTCGCAA GCCGACTGAT GCCTTCTGAA CAATGGAAAG GCATTATTGC

   301 CGTAAGCCGT GGCGGTCTGG TACCGGGTGC GTTACTGGCG CGTGAACTGG

   351 GTATTCGTCA TGTCGATACC GTTTGTATTT CCAGCTACGA TCACGACAAC

   401 CAGCGCGAGC TTAAAGTGCT GAAACGCGCA GAAGGCGATG GCGAAGGCTT

   451 CATCGTTATT GATGACCTGG TGGATACCGG TGGTACTGCG GTTGCGATTC

   501 GTGAAATGTA TCCAAAAGCG CACTTTGTCA CCATCTTCGC AAAACCGGCT

                                  (<--- Gpt10) GPT2 --->
   551 GGTCGTCCGC TGGTTGATGA CTATGTTGTT(GATATCCCGC AAGATACCTG)

                         <--- Gpt9
   601 GATTGAACAG CCGTGGGATA TGGGCGTCGT ATTCGTCCCG CCAATCTCCG

       *
   651 GTCGCTAATC TTTTCAACGC CTGGCACTGC CGGGCGTTGT TCTTTTTAAC

   701 TTCAGGCGGG TTACAATAGT TTCCAGTAAG TATTCTGGAG GCTGCATCCA

   751 TGACACAGGC AAACCTGAGC GAAACCCTGT TCAAACCCCG CTTTAAACAT

                         Gpt8 --->
   801 CCTGAAACCT CGACGCTAGT CCGCCGCTTT AATCACGGCG CACAACCGCC
```

851 TGTGCAGTCG GCCCTTGATG GTAAAACCAT CCCTCACTGG TATCGCATGA

901 TTAACCGTCT GATGTGGATC TGGCGCGGCA TTGACCCACG CGAAATCCTC

                              GPT5 --->   **GPT7 --->**(overlaps GPT5 sequence)
951 GACGTCCAGG CACGTATTGT <u>GATGAGCGAT GCCGAACGTA CCGA</u>CGATGA
                                        |------------------------

                        <--- GPT3
                        GPT4/Gpt4EcoRI --->(reverse orientation to GPT3)
1001 TTTATACGAT AC<u>GGTGATTG GCTACCGTGG CGGCAA</u>CTGG ATTTATGAGT
        ---|

1051 GGGCCCCGGA TCT **3' ------> SV40 sequence**

II.3 SV40 Early Sequence PCR Primers:

[van Heuverswyn et al., 1978]

(NB. This sequence is in the opposite orientation to the *gpt* sequence).

3851 GAGCTTCCTG GGGATCCAGA CATGATAAGA TACATTGAT<u>G AGTTTGGACA</u>

    <--- SV4011
3901 <u>AACCACAAC</u>T AGAATGCAGT GAAAAAAATG CTTTATTTGT GAAATTTGTG

3951 ATGCTATTGC TTTATTTGTA ACCATTATAA GCTGCAATAA ACAAGTTAAC

4001 AACAACAATT GCATTCATTT TATGTTTCAG GTTCAGGGGG AGGTGTGGGA

                 <--- SV4010
4051 GGTTTTTTAA AGCAAGTAAA ACGT<u>CTACAA ATGTGGTATG GCTG</u>ATTATG

4101 ATCTCTAGTC AAGGCACTAT ACATCAAATA TTCCTTATTA ACCCCTTTAC

            <--- SV408
4151 AAATTAAAAA GCTAAA<u>GGTA CACAATTTTT GAGCATAG</u>TT ATTAA<u>TAGCA</u>

   <--- SV406(SV405 --->)
4201 <u>GACACTCTAT GCCTGTGTG</u>G AGTAAGAAAA AACAGTATGT TATGATTATA

        <--- SV407
4251 ACTGTTAT<u>GC CTACTTATAA AGGTTACA</u>GA ATATTTTCC ATAATTTCT

4301 TGTATAGCAG TGCAGCTTTT TCCTTTGTGG TGTAAATAGC AAAGCAAGCA

4351 AGAGTTCTAT TACTAAACAC AGCATGACTC AAAAAACTTA GCAA<u>TTCTGA</u>

    SV403 --->(<--- SV409)
4401 <u>AGGAAA(GTCC TTGGGGTC)</u>TT CTACCTTTCT CTTCTTTTTT GGAGGAGTAG

         <--- SV404
4451 AATGTT<u>GAGA GTCAGCAGTA GCCTCATCAT</u> CACTAGATGG CATTTCTTCT

4501 GAGCAAAACA GGTTTTCCTC ATTAAAGGCA TTCCACCACT GCTCCCATTC

4551 ATCAGTTCCA TAGGTTGGAA TCTAAAATAC ACAAACAATT AGAATCAGTA

4601 GTTTAACACA TTATACACTT AAAAATTTTA TATTTACCTT AGAGCTTTAA

4651 ATCTCTGTAG GTAGTTTGTC CAATTAT<u>GTC ACACCACAGA AGTAAGGTTC</u>

4701 <u>C</u>TTCACAAAG ATC **3'-----> *gpt* sequence**

II.4 pBr plasmid sequence primers

530 CTGGCTTAAC TATGCGGCAT CAGAGCAGAT TGTACTGAGA GTGCACCATA

PB1 ---->
580 <u>TGCGGTGTGA AATACCGCAC CGATGC</u>GTAA GGAGAAAATA C<u>CGCATCAGG</u>

PB3 ---->
630 <u>CGCTCTTCCG CTT</u>CCTCGCT CACTGACTCG CTGCGCTCGG TCGTTCGGCT

PBA ---->
680 GCGGCG<u>AGCG GTATCAGCTC ACTCAAAGGC</u> GGTAATACGG TTATCCACAG

PB4 ---->
730 AATCAGGGGA TAACGCAGGA AA<u>GAACATGT GAGCAAAAGG CCAGCA</u>AAAG

780 GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC ATAGGCTCCG

PB5 ---->
830 CCCCCCT<u>GAC GAGCATCACA AAATCGACG</u> CTCAAGTCAG AGGTGGCGAA

<---- PB2
2540 CACATTTCCC CGAAAA<u>GTGC CACCTGACGT CTAAGAAACC</u> ATTATTATCA

II.5 Anchor PCR Primers:

CFB:

ATCTCCATACGT

CFB2:

AGTCCATACGT

AND*:

AGGATCCTACTCGAGTGCTATGGA

AND2:

GATCAGTCGACGTCATGCTATGGA

AND3:

Eco R1
CGCCTGAATTCTGCTATGGA

GPT4EcoRI:

| **Eco RI** | **Homology to GPT4 sequence** |

CATGTAGAATTCGGTGATTGGCTACCGTGG

AND4N:

GCTATAGCTTTCAGGCATCGTGAGCTCATCCTGA
ATTCTGCTATGGA

AND5:

GCTATAGCTTCAGGCATCGT

AND6:

CATCGTGAGCTCATCCTGAA

CFB3:

ATCTCCATAGCAGAATTCAGCTA

II.6 Unknown Region Primers

NS1   CTTCTGGAATAGCTCAGAGG

N S 2   CCTCTGAGCTATTCCAGAAG

NS3   GATACATCAGGTCAGAACAC

N S 4   GGTGTTCTGACCTGATGTAT

N S 5   CCACTCTGCCTTTTTGTCTC

Appendix III

Analysis of Primer Sequences for Possible Self Interactions

Oligonucleotides analysed using the Primerselect program. The strength of the interaction is indicated by the free energy ΔG.

III.1 Potential Hairpin Structures

And2

```
And2 , 3 bp (Loop=3), ΔG = 1.8 kc/m
  5' GATCAGTCGACGTCATG─┐
                 |||   C
          3' AGGTAT──┘
```
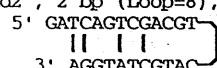
```
And2 , 2 bp (Loop=10), ΔG = 2.3 kc/m
      5' GATCAGTCGA─┐
           | ||     │
   3' AGGTATCGTACTGC─┘
```

```
And2 , 2 bp (Loop=5), ΔG = 2.4 kc/m
  5' GATCAGTCGACG─┐
         ||   |   T
      3' AGGTATCGTAC─┘
```

```
And2 , 2 bp (Loop=8), ΔG = 2.5 kc/m
  5' GATCAGTCGACGT─┐
       ||  | |      │
    3' AGGTATCGTAC──┘
```

```
And2 , 2 bp (Loop=8), ΔG = 2.5 kc/m
   5' GATCAGTCGACGTCATGC─┐
                 ||       │
          3' AGGTAT──┘
```

```
And2 , 2 bp (Loop=7), ΔG = 2.6 kc/m
     5' GATCAGTC─┐
          || |    G
   3' AGGTATCGTACTGCA─┘
```

```
And2 , 2 bp (Loop=10), ΔG = 2.7 kc/m
   5' GATCAGTCGACGTCA─┐
          || |         │
         3' AGGTATCGT──┘
```

```
And2 , 2 bp (Loop=10), ΔG = 2.7 kc/m
       5' GATCAGT─┐
            ||     │
   3' AGGTATCGTACTGCAGC─┘
```

```
And2 , 2 bp (Loop=11), ΔG = 2.9 kc/m
   5' GATCAGTCGAC─┐
          ||       G
     3' AGGTATCGTACT─┘
```

```
And2 , 2 bp (Loop=4), ΔG = 2.9 kc/m
            5' GATC─┐
                 ||  │
   3'. AGGTATCGTACTGCAGCTGA─┘
```

And 3

```
And3, 2 bp (Loop=7), ΔG = 1.0 kc/m
5' CGCCTGAATT┐
       || |    C
3' AGGTATCGT┘

And3, 2 bp (Loop=9), ΔG = 1.1 kc/m
    5' CGCCTGA┐
          ||    A
3' AGGTATCGTCTT┘

And3, 2 bp (Loop=7), ΔG = 2.5 kc/m
5' CGCCTGAATTCTGC┐
           ||     T
      3' AGGTA┘

And3, 2 bp (Loop=6), ΔG = 2.8 kc/m
5' CGCCTGAATTCT┐
       |  ||    ┐
   3' AGGTATCG┘
```

And 5

```
And5, 2 bp (Loop=5), ΔG = 1.3 kc/m
5' GCTATAGCTT┐
        ||    C
3' TGCTACGGA┘

And5, 2 bp (Loop=4), ΔG = 1.4 kc/m
        5' GCTA┐
              ||  ┐
3' TGCTACGGACTTCGAT┘

And5, 2 bp (Loop=3), ΔG = 2.1 kc/m
    5' GCTATAG┐
          ||  |  C
3' TGCTACGGACTT┘

And5, 2 bp (Loop=8), ΔG = 2.5 kc/m
        5' GCTATAG┐
              ||    ┐
3' TGCTACGGACTTC┘

And5, 2 bp (Loop=4), ΔG = 3.0 kc/m
5' GCTATAGCTT┐
    |  ||  |  ┐
3' TGCTACGGAC┘
```

And 6

And6, 3 bp (Loop=3), ΔG = 1.7 kc/m
```
5' CATCGTGAGCTCAT─┐
              │││  C
        3' AAGTC──┘
```

And6, 2 bp (Loop=4), ΔG = 2.5 kc/m
```
   5' CATCGTGAG─┐
 ·    ││   ││   │
3' AAGTCCTACTC─┘
```

And6, 2 bp (Loop=6), ΔG = 2.7 kc/m
```
5' CATCGTGAGCTC─┐
        │ ││    │
   3' AAGTCCTA─┘
```

And6, 2 bp (Loop=9), ΔG = 2.7 kc/m
```
   5' CATCGTG─┐
        ││     A
3' AAGTCCTACTCG─┘
```

And6, 2 bp (Loop=5), ΔG = 2.8 kc/m
```
5' CATCGTGAGC─┐
     ││   ││   T
   3' AAGTCCTAC─┘
```

And6, 2 bp (Loop=3), ΔG = 3.2 kc/m
```
              5' CAT─┐
                 ││   C
3' AAGTCCTACTCGAGTG─┘
```

And 4N

And4N, 6 bp (Loop=4), ΔG = -5.6 kc/m (bad!)
```
   5' GCTATAGCTTTCAGGCATCGTG─┐
      ││ ││││││        │      │
3' AGGTATCGTCTTAAGTCCTACTCGA─┘
```

And4N, 4 bp (Loop=7), ΔG = -2.2 kc/m (bad!)
```
            5' GCTATAGCTTTCAGG─┐
               │    ││││ │      C
3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTA─┘
```

And4N, 3 bp (Loop=7), ΔG = -0.6 kc/m
```
                 5' GCTATAGCTTTC─┐
                    │││  │ │       A
3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGG─┘
```

And4N, 3 bp (Loop=8), ΔG = -0.6 kc/m
```
5' GCTATAGCTTTCAGGCATCGTGAGCTCATCCT─┐
              ││   │││  │            │
        3' AGGTATCGTCTTAAG─┘
```

And4N, 3 bp (Loop=7), ΔG = -0.5 kc/m
```
5' GCTATAGCTTTCAGGCATCGTGAGCTCATCCTGAATT─┐
                      │││  │    │           C
              3' AGGTATCGT─┘
```

And4N, 3 bp (Loop=13), ΔG = 0.1 kc/m
```
5' GCTATAGCTTTCAGGCATCGTGAGCTC⌐
         |    |||    |          A
      3' AGGTATCGTCTTAAGTCCT⌐┘
```

And4N, 3 bp (Loop=8), ΔG = 0.7 kc/m
```
5' GCTATAGCTTTCAGGCATCGTGAGCTCATC⌐
              |||  |      |        ⌐
         3' AGGTATCGTCTTAAGTC⌐┘
```

And4N, 3 bp (Loop=5), ΔG = 0.8 kc/m
```
5' GCTATAGCTTTCAGGCATCGTGAGC⌐
       |   |  |||   ||   ||   T
      3' AGGTATCGTCTTAAGTCCTAC⌐┘
```

And4N, 2 bp (Loop=7), ΔG = 1.0 kc/m
```
5' GCTATAGCTTTCAGGCATC⌐
        ||          ||   G
   3' AGGTATCGTCTTAAGTCCTACTCGAGT⌐┘
```

And4N, 2 bp (Loop=6), ΔG = 1.2 kc/m
```
                        5' GCTATAGCTTT⌐
                                     ||  ⌐
   3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGAC⌐┘
```

And4N, 2 bp (Loop=11), ΔG = 1.4 kc/m
```
      5' GCTATAGCTTTCAGGCATCG⌐
         ||   |   ·| |        T
   3' AGGTATCGTCTTAAGTCCTACTCGAG⌐┘
```

And4N, 2 bp (Loop=4), ΔG = 1.4 kc/m
```
                              5' GCTA⌐
                                   ||  ⌐
   3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGACTTTCGAT⌐┘
```

And4N, 2 bp (Loop=4), ΔG = 1.4 kc/m
```
                              5' GCTATAGC⌐
                                   ||  |   ⌐
   3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGACTTT⌐┘
```

III.2 Primer Dimer Interactions

Stability indicated by ΔG values. Also shown is the worst possible ΔG value for a perfectly matched dimer.

And 2

```
And2 , 3 bp,  ΔG = -3.4 kc/m (worst= -43.0)
   5' GATCAGTCGACGTCATGCTATGGA 3'
             III   III
              3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -3.1 kc/m (worst= -43.0)
   5' GATCAGTCGACGTCATGCTATGGA 3'
                    II
              3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -2.0 kc/m (worst= -43.0)
   5' GATCAGTCGACGTCATGCTATGGA 3'
        II I   IIII   I II
         3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -2.0 kc/m (worst= -43.0)
     5' GATCAGTCGACGTCATGCTATGGA 3'
        I II          II I
   3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -2.0 kc/m (worst= -43.0)
   5' GATCAGTCGACGTCATGCTATGGA 3'
       II    I I I I    II
      3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -1.6 kc/m (worst= -43.0)
    5' GATCAGTCGACGTCATGCTATGGA 3'
          II           II
   3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -1.6 kc/m (worst= -43.0)
   5' GATCAGTCGACGTCATGCTATGGA 3'
          II         II
              3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -1.6 kc/m (worst= -43.0)
   5' GATCAGTCGACGTCATGCTATGGA 3'
       II  I I        I I  II
      3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -1.5 kc/m (worst= -43.0)
    5' GATCAGTCGACGTCATGCTATGGA 3'
       II     IIIIII      II
   3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -1.5 kc/m (worst= -43.0)
   5' GATCAGTCGACGTCATGCTATGGA 3'
                  I II I
              3' AGGTATCGTACTGCAGCTGACTAG 5'


And2 , 2 bp,  ΔG = -1.0 kc/m (worst= -43.0)
    5' GATCAGTCGACGTCATGCTATGGA 3'
              II
              3' AGGTATCGTACTGCAGCTGACTAG 5'
```

And 3

```
And3, 2 bp, ΔG = -3.1 kc/m (worst= -38.9)
       5' CGCCTGAATTCTGCTATGGA 3'
            ||   | |    ||
       3' AGGTATCGTCTTAAGTCCGC 5'


And3, 2 bp, ΔG = -3.1 kc/m (worst= -38.9)
       5' CGCCTGAATTCTGCTATGGA 3'
                    ||
          3' AGGTATCGTCTTAAGTCCGC 5'


And3, 2 bp, ΔG = -3.1 kc/m (worst= -38.9)
      5' CGCCTGAATTCTGCTATGGA 3'
           ||  |       |  ||
         3' AGGTATCGTCTTAAGTCCGC 5'


And3, 2 bp, ΔG = -1.6 kc/m (worst= -38.9)
       5' CGCCTGAATTCTGCTATGGA 3'
                 ||        ||
             3' AGGTATCGTCTTAAGTCCGC 5'


And3, 2 bp, ΔG = -1.5 kc/m (worst= -38.9)
       5' CGCCTGAATTCTGCTATGGA 3'
                      |  || |
                 3' AGGTATCGTCTTAAGTCCGC 5'
```

And 5

```
And5, 2 bp, ΔG = -3.6 kc/m (worst= -37.4)
       5' GCTATAGCTTCAGGCATCGT 3'
                       ||
                    3' TGCTACGGACTTCGATATCG 5'


And5, 2 bp, ΔG = -3.1 kc/m (worst= -37.4)
       5' GCTATAGCTTCAGGCATCGT 3'
                 ||  |  |  ||
          3' TGCTACGGACTTCGATATCG 5'


And5, 2 bp, ΔG = -3.1 kc/m (worst= -37.4)
         5' GCTATAGCTTCAGGCATCGT 3'
              ||   |    |   ||
          3' TGCTACGGACTTCGATATCG 5'


And5, 2 bp, ΔG = -3.1 kc/m (worst= -37.4)
       5' GCTATAGCTTCAGGCATCGT 3'
                  |  ||  |
                 3' TGCTACGGACTTCGATATCG 5'


And5, 2 bp, ΔG = -1.6 kc/m (worst= -37.4)
          5' GCTATAGCTTCAGGCATCGT 3'
               ||   ||||    ||
          3' TGCTACGGACTTCGATATCG 5'


And5, 2 bp, ΔG = -1.6 kc/m (worst= -37.4)
       5' GCTATAGCTTCAGGCATCGT 3'
          | ||   ||   ||   ||| |
          3' TGCTACGGACTTCGATATCG 5'


And5, 2 bp, ΔG = -1.5 kc/m (worst= -37.4)
       5' GCTATAGCTTCAGGCATCGT 3'
                     ||
                 3' TGCTACGGACTTCGATATCG 5'
```

And 6

```
And6, 3 bp, ΔG = -3.5 kc/m (worst= -36.3)
  5' CATCGTGAGCTCATCCTGAA 3'
            |||    |||
          3' AAGTCCTACTCGAGTGCTAC 5'

And6, 2 bp, ΔG = -2.0 kc/m (worst= -36.3)
   5' CATCGTGAGCTCATCCTGAA 3'
      ||   |||||||   ||
    3' AAGTCCTACTCGAGTGCTAC 5'

And6, 2 bp, ΔG = -1.6 kc/m (worst= -36.3)
  5' CATCGTGAGCTCATCCTGAA 3'
        |  ||      ||  |
        3' AAGTCCTACTCGAGTGCTAC 5'

And6, 2 bp, ΔG = -1.6 kc/m (worst= -36.3)
  5' CATCGTGAGCTCATCCTGAA 3'
        ||  ||     ||  ||
       3' AAGTCCTACTCGAGTGCTAC 5'

And6, 2 bp, ΔG = -1.6 kc/m (worst= -36.3)
  5' CATCGTGAGCTCATCCTGAA 3'
              ||   ||
          3' AAGTCCTACTCGAGTGCTAC 5'
```

And 4N

```
And4N, 5 bp, ΔG = -7.2 kc/m (bad!) (worst= -88.3)
  5' GCTATAGCTTTCAGGCATCGTGAGCTCATCCTGAATTCTGCTATGGA 3'
     | |||||          ||||| |
   3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGACTTTCGATATCG 5'

And4N, 3 bp, ΔG = -5.1 kc/m (bad!) (worst= -88.3)
  5' GCTATAGCTTTCAGGCATCGTGAGCTCATCCTGAATTCTGCTATGGA 3'
           |    |||  |           |   |||    |
         3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGACTTTCGATATCG 5'

And4N, 3 bp, ΔG = -4.7 kc/m (bad!) (worst= -88.3)
  5' GCTATAGCTTTCAGGCATCGTGAGCTCATCCTGAATTCTGCTATGGA 3'
              ||   |||  |          |  |||   ||
            3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGACTTTCGATATCG 5'

And4N, 3 bp, ΔG = -4.6 kc/m (bad!) (worst= -88.3)
    5' GCTATAGCTTTCAGGCATCGTGAGCTCATCCTGAATTCTGCTATGGA 3'
                        |||  |       |  |||
                  3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGACTTTCGATATCG 5'

And4N, 3 bp, ΔG = -3.4 kc/m (worst= -88.3)
  5' GCTATAGCTTTCAGGCATCGTGAGCTCATCCTGAATTCTGCTATGGA 3'
                |||    | |      |||
            3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGACTTTCGATATCG 5'

And4N, 3 bp, ΔG = -3.4 kc/m (worst= -88.3)
  5' GCTATAGCTTTCAGGCATCGTGAGCTCATCCTGAATTCTGCTATGGA 3'
       ||| |    |         |    | |||
            3' AGGTATCGTCTTAAGTCCTACTCGAGTGCTACGGACTTTCGATATCG 5'
```

III.3 Anchor interactions with coupling primer

For both And2 and And3, the strongest coupling primer/anchor interaction
was the one around which they were designed. In the case of And2, other
interactions are almost as strong. The And4N/CFB3 interaction could not be
analysed by the program in this way.

And2/CFB2

```
And2  vs. CFB2, 5 bp, ΔG = -7.5 kc/m (bad!) (worst= -43.0)
  5' GATCAGTCGACGTCATGCTATGGA 3'
                      |||||
              3' TGCATACCCTGA 5'


And2  vs. CFB2, 4 bp, ΔG = -6.3 kc/m (bad!) (worst= -43.0)
  5' GATCAGTCGACGTCATGCTATGGA 3'
          ||||   |   |
              3' TGCATACCCTGA 5'


And2  vs. CFB2, 3 bp, ΔG = -4.6 kc/m (bad!) (worst= -43.0)
  5' GATCAGTCGACGTCATGCTATGGA 3'
              | |    |||
                3' TGCATACCCTGA 5'


And2  vs. CFB2, 2 bp, ΔG = -3.6 kc/m (worst= -43.0)
 . 5' GATCAGTCGACGTCATGCTATGGA 3'
              ||
          3' TGCATACCCTGA 5'


And2  vs. CFB2, 3 bp, ΔG = -3.4 kc/m (worst= -43.0)
  5' GATCAGTCGACGTCATGCTATGGA 3'
              ||| |
            3' TGCATACCCTGA 5'


And2  vs. CFB2, 3 bp, ΔG = -2.9 kc/m (worst= -43.0)
  5'. GATCAGTCGACGTCATGCTATGGA 3'
         |   |||
  3' TGCATACCCTGA 5'


And2  vs. CFB2, 2 bp, ΔG = -1.6 kc/m (worst= -43.0)
  5' GATCAGTCGACGTCATGCTATGGA 3'
                  |    ||
          3' TGCATACCCTGA 5'


And2  vs. CFB2, 2 bp, ΔG = -1.3 kc/m (worst= -43.0)
  5' GATCAGTCGACGTCATGCTATGGA 3'
          ||    | |
        3' TGCATACCCTGA 5'
```

And3/CFB

```
And3 vs. CFB, 6 bp, ΔG = -9.0 kc/m (bad!) (worst= -38.9)
   5' CGCCTGAATTCTGCTATGGA 3'
                  ||||||
            3' TGCATACCTCTA 5'


And3 vs. CFB, 2 bp, ΔG = -2.0 kc/m (worst= -38.9)
   5' CGCCTGAATTCTGCTATGGA 3'
            |  | ||
         3' TGCATACCTCTA 5'


And3 vs. CFB, 2 bp, ΔG = -2.0 kc/m (worst= -38.9)
     5' CGCCTGAATTCTGCTATGGA 3'
         ||  || |
     3' TGCATACCTCTA 5'


And3 vs. CFB, 2 bp, ΔG = -1.6 kc/m (worst= -38.9)
       5' CGCCTGAATTCTGCTATGGA 3'
       .           || |
     3' TGCATACCTCTA 5'


And3 vs. CFB, 2 bp, ΔG = -1.6 kc/m (worst= -38.9)
        5' CGCCTGAATTCTGCTATGGA 3'
                   ||
     3' TGCATACCTCTA 5'


And3 vs. CFB, 2 bp, ΔG = -1.6 kc/m (worst= -38.9)
     5' CGCCTGAATTCTGCTATGGA 3'
            |   |   ||
            3' TGCATACCTCTA 5'


And3 vs. CFB, 2 bp, ΔG = -1.5 kc/m (worst= -38.9)
     5' CGCCTGAATTCTGCTATGGA 3'
               |   |  ||
         3' TGCATACCTCTA 5'


And3 vs. CFB, 2 bp, ΔG = -1.5 kc/m (worst= -38.9)
     5' CGCCTGAATTCTGCTATGGA 3
            | ||   | |
         3' TGCATACCTCTA 5'


And3 vs. CFB, 2 bp, ΔG = -1.5 kc/m (worst= -38.9)
     . 5' CGCCTGAATTCTGCTATGGA 3'
                   ||
     3' TGCATACCTCTA 5'
```

Appendix IV

Abbreviations

A	Adenine
Ac	Acetate
ATP	Adenosine 5'-triphosphate
bis-Tris	bis[2-Hydroxylethyl]iminotris[hydroxymethyl]methane;2-bis[2-Hydroxylethyl]amino-2-[hydroxymethyl]-1,3-propanediol
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
CIP	Calf-intestinal alkaline phosphatase
C	Cytosine
HCMV	Human cytomegalovirus strain-A169
dAMP	2'-deoxyadenosine 5'-monophosphate
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
ddATP	2'3'-dideoxyadenosine 5'-triphosphate
ddCTP	2'3'-dideoxycytidine 5'-triphosphate
ddGTP	2'-3'dideoxyguanosine 5'-triphosphate
ddTTP	2'-3'dideoxythymidine 5'-triphosphate
DEAE	Diethylaminoethyl
DMEM	Dulbecco's modified Eagles-medium
DMSO	Dimethyl sulphoxide

DNA	Deoxyribonucleic acid
DNAase	Deoxyribonuclease
dNTP	Deoxynucleotide triphosphate
DHFR	Dihydrofolate reductase
DTT	Dithiothreitol
E-buffer	Tri-acetate-EDTA electrophoresis buffer
EBV	Epstein Barr virus
EDTA	Diaminoethanotetra-acetic acid disodium salt
EtBr	Ethidium bromide
G	Guanine
GCE	Glucocorticoid response element
gpt	Xanthine-guanine phosphoribosyl transferase gene of *E. coli*
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
hnRNA	Heterogeneous nuclear RNA
HS	DNAase 1 hypersensitive site
Ig	Immunoglobulin
IgH	Immunoglobulin heavy-chain
IgL	Immunoglobulin light chain
IPTG	Isopropylthio-β D-galactoside
kb	Kilobase pairs
Klenov	Large fragment of DNA-polymerase I
l	Lambda IgL-chain
LB	Luria-Bertani medium
LCR	Locus control region
MAR	Matrix attachment region

MoMLV	Moloney murine leukaemia virus
MOPS	3-(N-morpholino) propanesulphonic acid
mRNA	Messenger RNA
OD	Optical density
PAGE	Polyacrylamide-gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
RNP	Ribonucleoprotein particles
RSV	Rous sarcoma virus
SAR	Scaffold attachment region
snRNP	Small nuclear ribonucleoprotein particles
SV40	Simian virus 40
T	Thymine
TF	Transcription factor
Tris	Tris[hydroxymethyl]aminomethane

Bibliography

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

Al-Shawi, R., Kinnaird, J., Burke, J. and Bishop, J.O.*Molecular and Cellular Biology* **1990**, 1192-1198 (1989).

Almer, A. and Hørz, W. *EMBO Journal* **5**, 2681-2687 (1986).

Almouzni, G., Mechali, M. and Wolffe, A. *Molecular and Cellular Biology* **11**, 655-665 (1991).

Aronow, B., Silbiger, R., Dusing, M., Stock, J., Yager, K.L., Potter, S.S., Hutton, J.J. and Wiginton, D.A. *Molecular and Cellular Biology* **12**, 4170-4185 (1992).

Atchison, M.L. *Annual Review Of Cell Biology* **4**, 127-153 (1988).

Ausio, J. *Trends in Biochemical Sciences* **17**, 187-191 (1992).

Banerji, J., Rusconi, S. and Schaffner, W. *Cell* **27**, 299-308 (1981).

Bartolome, S., Bermudez, A. and Daban, J. *Journal of Cell Science* **107**, 2983-2992 (1994).

Baumann, G., Raschke, E.,Bevan, M. and Schoffl, F. *EMBO Journal* **6**, 1161- 1166 (1987).

Becker, P. *Bioessays* **16**, 541-547 (1994).

Bell, D.A. and DeMarini, D.M. *Nucleic Acids Research* **19**, 5079 (1991)

Biggin, M.D., Gibson, T.J. and Hong, E.F. *Proceedings of the National Academy Sciences USA* **80**, 3963-3965 (1983).

Birchmeier, C., Folk, W. and Birnstiel, M.L. *Cell* **35**, 433-440 (1983).

Birchmeier, C., Grosschedl, R. and Birnstiel, M.L. *Cell* **28**, 739-745 (1982).

Bode, J. and Maass, K. *Biochemistry* **27**, 4706-4711 (1988).

Bonifer, C., Yannoutsos, N., Kruger, G., Grosveld, F. and Sippel, A.E. *Nucleic Acids Research* **22**, 4202-4210 (1994).

Bonifer, C., Vidal, M., Grosveld, F. and Sippel, A.E. *EMBO Journal* **9**, 2843-2848 (1990).

Boulikas, T. *International Journal of Oncology* **6**, 1313-1318 (1995).

Boulikas, T. *Journal of Cellular Biochemistry* **55**, 513-529 (1994).

Bowen, B. *Nucleic Acids Research* **9** 5093-5108 (1981)

Brake, A.J., Merryweather, J.P., Loit, D.G., Heberlei,U.A., Masriarz, F.R., Mullenbach, G.T., Urdea, M.S., Valenzuela, P. and Barr, P.J. *Proceedings of the National Academy of Sciences USA* **81**, 4642-4646 (1984).

Bresnick, E.H. and Felsenfield, G. *Proceedings of the National Academy of Sciences USA* **91**, 1314-1317 (1994).

Burke, D.T., Carle, G.F. and Olson, M.V. *Science* **236**, 806-812 (1987).

Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R. and Townes, T.M. *Proceedings of the National Academy of Sciences USA* **88**, 1626-1630 (1991).

Caterina, J.J., Ciavatta, D.J., Donze, D. and Behringer, R.R., Townes, T.M. *Nucleic Acids Research* **22**, 1006-1011 (1994).

Chamberlain, J.W., Vasavada, H.A., Ganguly, S. and Weissman, S.M. *Molecular and Cellular Biology* **11**, 3564-3572 (1991).

Chang, N.C., Matteucci, M., Perry, J., Wulf,J.-J. Chen, C.-Y. and Hitzeman, R.A. *Molecular and Cellular Biology* **6**, 1812-1819 (1986).

Chasman, D.I., Lue, N.F., Buchman, A.R., LaPointe, J.W., Lorch, Y. and Kornberg, R.P. *Genes and Development* **4**, 503-514 (1990).

Cheyette, T.E., Ip, T., Faber, S., Matsui, Y. and Chalkley, R. *Nucleic Acids Research* **20**, 3427-3433 (1992).

Churchill, M. and Travers, A. *Trends in Biochemical Sciences* **16**, 92-97 (1991).

Clarke, L. and Carbon, J. *Cell* **9**, 91-97 (1976).

Cockerill, P.N. *Nucleic Acids Research* **18**, 2643-2648 (1990).

Conaway, R.C. and Conaway, J.W. *Annual Review of Biochemistry* **62**, 161-190 (1993).

Courey, A.J. and Tjian, R. *Cell* **55**, 887-898 (1988).

Courey, A.J., Holtzman, D.A., Jackson, S.P. and Tijan, R. *Cell* **59**, 827-836 (1989).

Cowell, I.G., Skinner, A. and Hurst, H.C. *Molecular and Cellular Biology* **12**, 3070-3077 (1992).

Cowell, I.G. *Trends in Biochemical Sciences* **19**, 38-42 (1994).

Curtin, P. and Kan, Y.W. in *Haemoglobin Switching* (eds. Stamatoyannopoulos, G.) (Alan R. Liss, New York, 1987).

Craddock, C., Vyas, P., Sharpe, J.A., Ayyub, H., Wood, W.G. and Higgs, D.R. EMBO Journal **14**, 1718-1726 (1995)

Curtin, P.T., Liu, D., Liu, W., Chang, J.C. and Kan, Y.W. *Proceedings of the National Academy of Sciences USA* **86**, 7082-7086 (1989).

Dale, T.C., Krnacik, M.J., Schmidhauser, C., Yang, C.L-Q., Bissell, M.J. and Rosen, J. *Molecular and Cellular Biology* **12**, 905-914 (1992).

Davie, J. and Hendzel, M. *Journal of Cellular Biochemistry* **55**, 98-105 (1994).

Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V. and Kopka, M.L. *Science* **216**, 475-485 (1982).

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. *Nucleic Acids Research* **19**, 4008 (1991).

Driscoll, M., Dobkin, C.S. and Alter, B.P. *Proceedings of the National Academey of Sciences USA* **86**, 7470-7474 (1989).

Eick, D., Wedel, A. and Heumann, H. *Trends in Genetics* **10**, 292-296 (1994).

Eissenberg, J.-C. and Elgin, S.C.R. *Trends in Genetics* **7**, 335-340 (1991).

Elgin, E. *Strategies* **4**, 8-9 (1991).

Ellis, J., Talbot, D., Dillon, N. and Grosveld, F. *EMBO Journal* **12**, 127-134 (1993).

Erlich, H. *PCR Technology; Principles and Applications* (Stockton Press, 1989).

Felsenfeld, G. *Nature* **355**, 219-224 (1992).

Forrester, W.C., Vangenderen C., Jenuwein T. and Grosschedl, R. *Science* **265**, 1221-1225 (1994).

Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. *Nucleic Acids Research* **15**, 10159-10177 (1987).

Fraser, P., Hurst, J., Collis, P. and Grosveld, F. *Nucleic Acids Research* **18**, 3503-3508 (1990).

Freshney, R.I. *Culture of Animal Cells. A Manual of Basic Technique* (1983) A. R. Liss, New York.

Garrard, W. *Bioessays* **13**, 87-88 (1991).

Garzon, R. and Zehner, Z. *Molecular and Cellular Biology* **14**, 934-943 (1994).

Gasser, S.M. and Laemmli, U.K. *Cell* **46**, 521-530 (1986).

Getzenberg, R. *Journal of Cellular Biochemistry* **55**, 22-31 (1994).

Gibson, T.J. *Studies on the Epstein-Barr virus genome* (Cambridge University, 1984).

Glassford, S. *High Level Gene Expression in Myeloma Cells* (University of Leicester, 1993)

Gluzman, Y. *Cell* **23** 175-182 (1981)

Grant, S. G., Jessee, J., Bloom, F. R. and Hanahan, D.*Proceedings of the National Academy of Science USA* **87**, 4645-4649 (1990).

Greaves, D.R., Wilson, F.P., Lang, G. and Kioussis, D. *Cell* **56**, 979-986 (1989).

Grosveld, F., Van Assendelft, G. B., Greaves, D. R. and Kollias, G. *Cell* **51**, 975-985 (1987).

Hager, G., Archer, T., Fragoso, G., Bresnick, E., Tsukagoshi, Y., John, S. and Smith, C. *Cold Spring Harbor Symposia on Quantitative Biology* **58**, 63-71 (1993).

Hanahan, D. *Journal of Molecular Biology* **166**, 557-580 (1983).

Hansen, J., van Holde, K. and Lohr, D. *Journal of Biological Chemistry* **266**, 4276-4282 (1991).

Harrison, T.M., Hudson, K., Munson, S.E., Uff, S. and Glassford, S. *Biochimica et Biophysica Acta* **1260**, 147-156 (1995).

Hawkins, J. *Gene Structure and Expression* (Cambridge Unversity Press, 1991).

Higgs, D.R. Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.-M. andAyyub, H. *Genes and Development* **4**, 1588-1601 (1990).

Holstege, F.C.P., van der Vilt, P. and Timmers, H.Th.M. *Biochimica et Biophysica Acta* **1219**, 157-159 (1994).

Hoyle, G., Mercer, E., Palmiter, R. and Brinster, R. *Journal of Neuroscience* **14**, 2455-2463 (1994).

Huber, M., Bosch, F., Sippel, A. and Bonifer, C. *Nucleic Acids Research* **22**, 4195-4201 (1994).

Hudson, K. *Development of vectors allowing efficient heterologous-gene expression in stable myeloma cell transfectants.* (Leicester University) 1989).

Hudson, K. and Harrison, T.M. *Biochemical Society Transactions* **15**, 1033-1034 (1987).

Hull, M., Erickson, J., Johnston, M. and Engelke, D. *Molecular and Cellular Biology* **14**, 1266-1277 (1994).

Huynh, T.V., Young, R.A. and Davis, R.W. *DNA Cloning Techniques : A Practical Approach* (eds. Glover, D.) (IRL Press, Oxford, 1985).

Imbalzano, A.N, Kwon, H., Green, M.R. and Kingston, R.E. *Nature* **370**, 481-485 (1994).

Innis, M., Gelfand, D., Sninsky, J. and White, T. *PCR Protocols: A Guide to Methods and Applications* (Academic Press Inc, San Diego, 1990).

Jendrisak, J., Young, R.A. and Engel, J. *Guide to Molecular Cloning Techniques* (Academic Press, 1987).

Jones, K.A., Yamamoto, K. and Tjian, R. *Cell* **42**, 559-572 (1985).

Jung, A., Sippel, A.E., Grez, M. and Scheutz, G. *Proceedings of the National Academy of Sciences USA* **77**, 5759-5763

Kadonaga, J.T., Jones, K.A. and Tjian, R. *Trends in Biochemical Sciences* **11**, 20-23 (1986).

Kageyama, R. and Pastan I. *Cell* **59**, 815-25 (1989).

Kalos, M. and Fournier, R. *Molecular and Cellular Biology* **15**, 198-207 (1995).

Karin, M., Haslinger, A., Holtgreve, H., Cathala, G., Slater, E. and Baxter, J.D. *Cell* **36**, 371-379 (1984).

Kellum, R. and Schedl, P. *Cell* **64**, 941-950 (1991).

Kielman, M., Smits, R. and Bernini, L. *Genomics* **21**, 431-433 (1994).

Kingsman, S. and Kingsman, A. *Genetic Engineering* (Blackwell Scientific Publications, Oxford, 1988).

Kioussis, D., Vanin, E., deLange, T., Flavell, R.A. and Grosveld, F. *Nature* **306**, 662-666 (1983).

Kollias, G., Hurst J., deBoer, E. and Grosveld, F. *Nucleic Acids Research* **15**, 5739-5747 (1987).

Kornberg, R. and Lorch, Y. *Current Opinion in Cell Biology* **7**, 371-375 (1995).

Laskov, R. and Scharff M. D. *Journal of Experimental Medicine* **131**, 515-541 (1970)

Lawson, G.M., Knoll, B.J., March, C.J., Woo, S.L.C., Tsai, M.-J. and O'Malley, B. *Journal of Biological Chemistry* **257**, 1501-1507 (1982).

Levy-Wilson, B. and Fortier, C. *Journal of Biological Chemistry* **264**, 21196-21204 (1989).

Li, Q. and Stamatoyannopoulas, G. *Blood* **84**, 1399-1401 (1994).

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

Loenen, W. and Brammar, W.J. *Gene* **10**, 249-259 (1980).

Loh, E. *Methods: A Companion to Methods in Enzymology* **2**, 11-19 (1991).

Lopez, C., Dechesnay, A., Tournamille, C., Benghanem, A., Prigent, S., Drouet, X., Lambin, P. and Cartron, J. *Gene* **148**, 285-291 (1994).

Lorch, Y., LaPointe, J.W. and Kornberg, R. *Cell* **49**, 203-210 (1987).

Lu, Q., Wallrath, L.L., Elgin, S.C.R. *Journal of Cellular Biochemistry* **55**, 83-92 (1994).

Lundblad, A.,Stellar, R., Kabat, E., A. Hirst, J. W., Weigert, M. G. and Cohn, M. *Immunocytochemistry* **9**, 535-544 (1972).

Madden, S.L., Cook, D.M. and Rauscher, F.J.*Oncogene* **8**, 1713-1720 (1993).

Marston, F. *Biochemical Journal* **240**, 1-12 (1986).

Mather, E. L., Nelson, K. J., Haimovich, J., and Perry, R. P. *Cell* **36**, 329-338 (1984).

May, M., Grassmann, K., Pfister, H. and Fuchs, P. *Journal of Virology* **68**, 3612-3619 (1994).

McNeil, J. and Friesen, *J. Molec. Gen. Genet.* **184**, 386-393 (1981).

Meissner, P.S., Sisk, W. P. and Berman, M.L. *Proceedings of the National Academy of Sciences USA* **84**, 4171 (1987).

Mitchell, P. and Tjian, R. *Science* **245**, 371-378 (1989).

Mirkovitch, J., Mirault, M.E. and Laemmli, U.K *Cell* **39** 223-232 (1984)

Morley, B.J., Abbott, C.A., Sharpe, J.A., Lida, J., Chan-Thomas, P.S. and Wood, W.G. *Molecular and Cellular Biology* **12**, 2057-2066 (1992).

Needham, M., Gooding, C., Hudson, K., Antoniou, M., Grosveld, F. and Hollis, M. *Nucleic Acids Research* **20**, 997-1003 (1992).

Negri, R., Costanzo, G., Buttinelli, M., Venditti, S. and Dimauro, E. *Biophysical Chemistry* **50**, 169-181 (1994).

Ochman, H., Gerber, A.S. and Hartl, D.L. *Genetics* **120**, 621-623 (1988).

Oi, V.T., Morrison, S.L., Herzenberg, L.A. and Berg, P. *Proceedings National Academy of Sciences USA* **80**, 825-829 (1983).

Old, R. and Primrose, S. *Principles of Gene Manipulation* (Blackwell Scientific Publications, Oxford, 1985).

Padgett, R., Grabowski, P., Konarska, M., Seiler, S. and Sharp, P. *Annual Review of Biochemistry* **55**, 1119-1150 (1986)

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

Paranjape, S., Kamakaka, R. and Kadonaga, J. *Annual Review of Biochemistry* **63**, 265-297 (1994).

Perucho, M., Hanahan, D., Lipsich, L. and Wigler, M. *Nature* **285**, 207-210 (1980).

Peterson, C. and Tamkun, J. *Trends in Biochemical Sciences* **20**, 143-146 (1995).

Phi-Van, L. and Strätling, W.H. *EMBO Journal* **7**, 655-664 (1988).

Philipsen, S., Pruzina, S. and Grosveld, F. *EMBO Journal* **12**, 1077-1085 (1993).

Pinkert, C.A., Ornitz, D.M., Brinster, R.L., Palmiter, R.D. *Genes & Development* **1**, 268-276 (1987).

Platt, T. *Annual Review of Biochemistry* **55**, 339-372 (1986).

Poljak, L., Seum, C., Mattioni, T. and Laemmli, U.K. *Nucleic Acids Research* **22**, 4386-4394 (1994).

Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F. and Philipsen, S. *Nucleic Acids Research* **19**, 1413-1419 (1991).

Queen, C. and Baltimore, D. *Cell* **33**, 741-748 (1983).

Reitman, M., Lee, E., Westphal, H. and Felsenfield, G. *Nature* **348**, 749-752 (1990).

Rich, A., Nordheim,A. and Wang, A. *Annual Review of Biochemistry* **53**, 791-846 (1984).

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes,D. and Klug, A. *Nature* **311**, 532-537 (1984).

Richardson, K., Fostel, J. and Skopek, T. *Nucleic Acids Research* **11**, 8809-8816 (1983)

Ridsdale, A., Hendzel, J., Decluve. P. and Davie, J. *Journal of Biological Chemistry* **265**, 5150-5156 (1990).

Rolfs, A., Schuller, I., Finckh, U. and Weber-Rolfs, I. *PCR: Clinical Diagnostics and Research* (Springer-Verlag, Berlin, 1992).

Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R. and Brinster, R.L. *Genes & Development* **3**, 314-323 (1989).

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

San, K.-Y., Bennett, G.N., Aristidou, A.A. and Chou, C.*Recombinant DNA Technology II* 257-267 (New York Academy of Sciences, 1994).

Sanger, F., Nicklen, S. and Coulson, A.R. *Proceedings National Academy of Sciences USA* **74**, 5463-5467 (1977).

Schule, R., Umesono, K., Mangelsdorf, D., Bolado, J., Pike, J. and Evans, R. *Cell* **61**, 497-504 (1990).

Schultz, L.D., Markus, H.Z., Hofmann, K.J., Montgomery, D.L., Dunwiddie, C.T., Kniskern, P.J., Freedman, R.B., Ellis, R.W. and Tuite, M.F. in *Recombinant DNA Technology II* 148-157 (NewYork Academy of Sciences, 1994).

Schwarz, P. and Hansen, J. *Journal of Biological Chemistry* **269**, 16284-16289 (1994).

Siebert, P.D., Chen, S. and Kellog, D.E. *CLONTECHniques* X, 1-3 (1995).

Siebert, P.D., Chenchik, A., Kellog, D.E., Lukyanov, K.A. and Lukyanov, S.A.*Nucleic Acids Research* **23**, 1087-1088 (1995).

Sippel, A.E., Schafer, G., Faust, N., Saueressig, H., Hecht, A. and Bonifer, C. *Cold Spring Harbor Symposia on Quantitative Biology* **58**, 37-44 (1993).

Slack, J.L., Liska, D.J. and Bornstein, P. *Molecular and Cellular Biology* **11**, 2066-2074 (1991).

Southern, E.M. *Journal of Molecular Biology* **98**, 503-517 (1975).

Southern, P.J. and Berg, P. *Journal of Molecular and Applied Genetics* **1**, 327-341 (1982).

Staneloni, R.J. and Leloir, L.F. *Critical Reviews of Biochemistry* **12**, 298-326 (1982).

Stief, A., Winter, D.M., Stratling, W.H. and Sippel, A.E.*Nature* **341**, 343-345 (1989).

Strauss, E.C. and Orkin, S.H. *Proceedings of the National Academy of Sciences USA* **89**, 5809-5813 (1992).

Struhl, K., Stinchcomb, D., Scherer, S. and Davis, R. *Proceedings National Academy of Sciences USA* **76**, 1035-1039 (1979).

Su, W., Jackson, S., Tjian, R. and Echols, H. *Genes & Development* **5**, 820-826 (1991).

Talbot, D., Descombes, P. and Schibler, U. *Nucleic Acids Research* **22**, 756-766 (1994).

Talbot, D. and Grosveld, F. *EMBO Journal* **10**, 1391-1398 (1991).

Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. *EMBO Journal* **9**, 2169-2178 (1990).

Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. and Greaves, D. *Nature* **338**, 352-355 (1989).

Thoma, J., Koller, T. and Klug A. *Journal of Cell Biology* **83**, 403-427 (1979).

Tjian, R. and Maniatis, T. *Cell* **77**, 5-8 (1994).

Tokugawa, K., Ishil, T., Nakamura, K., Masaki, H. and Uozumi, T. *Journal of Biotechnology* **37**, 33-37 (1994).

Traunecker, A., Oliveri, F. and Karjalainen, K. *Trends in Biotechnology* **9**, 109-113 (1991)

Travers, A. *Bioessays* **16**, 657-662 (1994).

Tuan, D., Solomon, W., Li, Q. and London, I.M. *Proceedings of the National Academy of Sciences USA* **82**, 6384-6388 (1985).

van Heuverswyn, H., Herreweghe, J., Volckaert, G. and Ysebaert, M. *Nature* **273**, 113-120

van Holde, K. *Chromatin* (Springer, New York, 1989).

van Holde, K. and Zlatanova, J. *Journal of Biological Chemistry* **270**, 8373-8376 (1995).

Vettese-Dadey, M., Walter, P., Chen, H., Juan, L. and Workman, J. *Molecular and Cellular Biology* **14**, 970-981 (1994).

Walters, M., Kim, C. and Gelinas, R. *Nucleic Acids Research* **19**, 5385-5393 (1991).

Weidle U. H. and Buckel P. *Gene* **57**, 131-141 (1987)

Werner, R. *Pharmaceutical Technology Europe* May 20-28 (1994).

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

Wolffe, A.P. *Science* **241**, 1626-1632 (1988).

Wolffe, A.P. *Trends in Biochemical Sciences* **19**, 240-244 (1994).

Yang Yen, H., Chambard, J.-C., Sun, Y.-L., Smeal, T., Schmidt, T., Drouin, J. and Karin, M. *Cell* **62**, 1205-1215 (1990).

Yao, C., Lowary, J. and Widom, J. *Biochemistry* **30**, 8408-8414 (1991).

Yu, J.H., Bock, J.H., Slightom, J.L. and Vileponteau, B. *Gene* **139**, 139-145 (1994).