

BROCK UNIVERSITY LIBRARY



3 9157 00928206 5

**Isolation and Characterization of Genomic DNA Sequences That
Enhance the Stability of Plasmid DNA in Mammalian Cells.**

By

Marcus M.S. Manocha, B.Sc. (Hon)

A Thesis

Submitted to the Center for Biotechnology

in partial fulfillment of the requirements

for the degree of

Master of Science

October, 2005

Brock University

St. Catharines, Ontario, Canada

© Marcus Manocha, 2005.

**JAMES A GIBSON LIBRARY
BROCK UNIVERSITY
ST. CATHARINES ON**

Isolation and Characterization of Genomic DNA Sequences That Regulate the Stability of Plasmid DNA in *Escherichia coli*

By

Marlene M. S. Marinho, D. Sc. (1991)

A Thesis

Submitted to the Faculty of the

Graduate School of Arts and Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

at the

University of Maryland

College Park, Maryland

December 1991

U. M. I. 92-00000-10000

92-00000-10000

92-00000-10000

ABSTRACT

The ease of production and manipulation has made plasmid DNA a prime target for its use in gene transfer technologies such as gene therapy and DNA vaccines. The major drawback of plasmid however is its stability within mammalian cells. Plasmid DNA is usually lost by cellular mechanisms or as a result of mitosis by simple dilution. This study set out to search for mammalian genomic DNA sequences that would enhance the stability of plasmid DNA in mammalian cells.

Creating a plasmid based genomic DNA library, we were able to screen the human genome by transfecting the library into Human Embryonic Kidney (HEK 293) Cells. Cells that contained plasmid DNA were selected, using G418 for 14 days. The resulting population was then screened for the presence of biologically active plasmid DNA using the process of transformation as a detector.

A commercially available plasmid DNA isolation kit was modified to extract plasmid DNA from mammalian cells. The standardized protocol had a detection limit of ~0.6 plasmids per cell in one million cells. This allowed for the detection of 45 plasmids that were maintained for 32 days in the HEK 293 cells. Sequencing of selected inserts revealed a significantly higher thymine content in comparison to the human genome. Sequences with high A/T content have been associated with Scaffold/Matrix Attachment Region (S/MAR) sequences in mammalian cells. Therefore, association with the nuclear matrix might be required for the stability of plasmids in mammalian cells.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Y. Haj-Ahmad for support and encouragement throughout my master's degree. He was always able to brighten my spirits even when I thought things were totally and utterly wrong. I would also like to thank Dr. D. Bautista who guided me through the problems encountered in the lab and aided in developing my appreciation for the data I collected. Thank-you to my committee members Dr. A. Castle and Dr. A Skandalis.

My heart felt appreciation also goes out to the staff of Norgen Biotek Corporation. They are the people who have helped to nurture my scientific ability and have been my personal instructors for the last 5 years.

I would like to say a special thanks to Dawn Hall for editing this thesis and to Dr. J. Richardson and Sean Pager for consultation on statistical analysis. My appreciation also goes out to my lab mates Ken Lo, Moemen, Mohamed, Songsong and all my fabulous friends.

I would also like to thank my brother Michael and my mother who have encouraged and supported me throughout my career. It was their love that kept me going.

Finally I would like to thank my father because he is why I am, who I am.

TABLE OF CONTENTS

ABSTRACT	2
ACKNOWLEDGEMENTS	3
TABLE OF CONTENTS	4
LIST OF TABLES	8
LIST OF FIGURES	10
LIST OF ABBREVIATIONS	13
1.0 INTRODUCTION	15
1.1 Viral Vectors	17
1.2 Limitations of Viral Vectors	18
1.3 Artificial Chromosomes	19
1.4 Viral-based Plasmid Vectors	21
1.5 Limitations of Viral-based Plasmid Vectors	24
1.6 Non-viral Plasmid Vectors	25
1.7 Scaffold Matrix Attachment Regions (S/MARs)	28
1.8 Initiation of DNA Replication in Eukaryotes	31
1.9 Autonomous Replicating Sequences (ARS)	34
1.10 Plasmid Delivery	36
1.11 Isolation of Low Molecular Weight DNA	39
1.12 Project Objectives	42
2.0 MATERIALS AND METHODS	43
2.1 Bacterial Strain	43

2.2	Propagation of Bacterial Strain	43
2.3	Preparation of competent cells	43
2.3.1	Sambrook <i>et al.</i> Method (1989)	43
2.3.2	Inoue <i>et al.</i> Method (1990)	44
2.4	Bacterial Transformation	45
2.5	Small Scale Plasmid DNA Isolation from Bacterial Cells	46
2.5.1	Brimboi and Doly Method (1979)	46
2.5.2	Norgen Method	47
2.6	Restriction Enzyme Digestion of Plasmid DNA	47
2.7	Agarose Gel Electrophoresis	48
2.8	Ligation of Restriction Enzyme Digested pDNA Fragments	48
2.9	Shotgun Ligation Library	49
2.10	Large Scale Plasmid DNA Isolation	50
2.11	Shotgun Library Analysis	52
2.12	Determination of Nucleic Acid Concentration	53
2.12.1	Diphenylamine Assay	53
2.12.2	Direct Spectrophotometry	54
2.13	Heat Inactivation of Enzymes	54
2.14	Klenow Treatment of Digested DNA	54
2.15	Excision of Digested pDNA Fragments from Agarose Gels	55
2.16	DNA Sequencing and Analysis	55
2.17	Statistical Analysis	55
2.18	Isolation of Plasmid DNA from Mammalian Cells	56
2.18.1	Mammalian Cell Collection	56
2.18.2	Plasmid DNA Isolation using the Modified Norgen Protocol	56
2.18.3	Modified Hirt Protocol (Hirt, 1967)	58

2.19	Genomic DNA Isolation from HEK 293 Cells	58
2.20	Phenol:Chloroform:Isoamyl Alcohol Extraction	59
2.21	Mammalian Cell Culture Techniques	60
2.21.1	Human Embryonic Kidney Cell (293 Cell) Maintenance	60
2.21.2	HEK 293 Cell Passaging	61
2.21.3	Cell Freezing and Thawing	61
2.21.4	Mammalian Cell Counting	62
2.21.5	Calcium Phosphate Transfection	62
2.21.6	Selection with G418	63
2.21.7	Histochemical Assay for <i>lacZ</i> Activity	64
3.0	RESULTS	65
3.1	pUC19neo	65
3.1.1	Construction of pUC19neo	65
3.1.2	Creation & Confirmation of the plasmid pUC19neo	67
3.2	Genomic DNA Library	70
3.2.1	Construction of Genomic DNA Library	70
3.2.2	Analysis of the Genomic DNA libraries	73
3.3	Transfection and Extraction Procedure	78
3.3.1	Transfection efficiency and variability of HEK 293 cells	78
3.3.2	Determination of optimal G418 concentration during selection	81
3.3.3	Isolation of Low molecular weight DNA from mammalian cells	82
3.3.4	Isolation Sensitivity and Percent Recovery	84
3.3.5	Analysis of the Extraction Assay	86
3.3.6	Second Concentration step	88
3.4	Isolation Analysis	90
3.4.1	Indicators of Plasmid DNA Maintenance	90
3.4.2	Longitudinal Study of Plasmid DNA Stability	93

3.5	Genomic Insert Analysis	97
3.5.1	Insert determination and Size analysis	97
3.5.2	Sequencing and Nucleotide composition	105
3.5.3	Local and Global Alignment of the 12 sequenced inserts	109
3.5.4	Blast Search of the insert sequences and their chromosomal position	112
3.5.5	Analysis of Scaffold Matrix attachment Regions (S/MARs) by MAR-Finder	115
4.0	DISCUSSION	121
4.1	The Genomic DNA library	121
4.2	Isolation & Percent Recovery	122
4.3	Isolation Variability	124
4.4	Plasmid Maintenance	125
4.5	Sequence Analysis	130
4.6	Future Research & Questions	133
5.0	CONCLUSION	134
6.0	LITERATURE CITED	135

LIST OF TABLES

	Page
Table 3.1 Expected fragment sizes based on <i>in silico</i> analysis of the cloning reaction	68
Table 3.2 A Average percentage of colonies containing a recombinant plasmid due to the insertion of <i>EcoRI</i> digested genomic DNA	72
Table 3.3 B Average percentage of colonies containing a recombinant plasmid due to the insertion of <i>Tsp509I</i> digested genomic DNA	72
Table 3.3 A Percentage of the total number of independent recombinants required for 99% probability of having a sequence represented in the human genomic DNA library using the restriction enzyme <i>EcoRI</i>	75
Table 3.3 B Percentage of the total number of independent recombinants required for 99% probability of having a sequence represented in the human genomic DNA library using the restriction enzyme <i>Tsp509I</i>	75
Table 3.4 Number of GEs in the constructed genomic DNA libraries	75
Table 3.5 Results of varying concentration of G418 on 293 cells during a two-week period	81
Table 3.6 Figure legend for 3.17 A, B, C & D with respect to the lane number and plasmid name	100

Table 3.7	Figure legend for 3.18 A, B & C with respect to the lane number and plasmid name	103
Table 3.8	Selected plasmids that were sequenced and separated by their library origins	106
Table 3.9	Blast search results of inserts originating from the 3Mix genomic DNA library	113
Table 3.10	Blast search results of inserts originating from the 3lig509 genomic DNA library	114

LIST OF FIGURES

		Page
Figure 1.1	Survey of clinical trials using gene transfer technology	16
Figure 1.2	Schematic drawing of a mammalian artificial chromosome	19
Figure 1.3	Maintenance mechanisms of viral genomes	22
Figure 1.4	Non-Viral episomally replicating vector based on the SV40 origin of replication	26
Figure 1.5	Initiation of DNA replication	32
Figure 1.6	Delivery of plasmid DNA and the associated barriers	38
Figure 3.1	Methodology for the construction of pUC19neo	66
Figure 3.2	Confirmation of pUC19neo (Reverse)	69
Figure 3.3	Strategy for the production of genomic DNA libraries using pUC19neo as the cloning vector	71
Figure 3.4	Library variability using insert size analysis	76
Figure 3.5	Restriction enzyme analysis of genomic DNA libraries after the large scale propagation and isolation	77
Figure 3.6	Variability seen between transfections on different days	79
Figure 3.7	Transfection variability between samples	80
Figure 3.8	Isolation of plasmid DNA (pUC19neo) from an equal number of transfected mammalian cells	83
Figure 3.9	Isolation of known amounts of plasmid DNA from mammalian cells	85

Figure 3.10	Average number of colonies (n=3) obtained when plasmid DNA (1.0×10^{-4} μ g) is injected into the procedure at three different time points	87
Figure 3.11	Average output of post column preparation of plasmid DNA for transformation (n=3)	89
Figure 3.12	Non-selective maintenance of plasmid DNA in mammalian cells	91
Figure 3.13	Average (n=4) output of blue colony plasmid DNA isolated from the longitudinal study	92
Figure 3.14	Long term output of biologically active plasmid DNA from Mammalian cells	95
Figure 3.15	Average (n=4) amount of plasmid DNA isolated from each sample on Day 32.	96
Figure 3.16	Analysis of 21 plasmids isolated from day 32 mammalian cells which were transfected with the 3Mix (<i>EcoRI</i>) genomic DNA library	100
Figure 3.17	Analysis of 24 plasmids isolated from day 32 mammalian cells originating from the 3Lig509 (<i>Tsp509I</i>) genomic DNA library	102
Figure 3.18	Distribution of insert sizes found in isolated plasmids	104
Figure 3.19	Average nucleotide composition of the 12 sequenced genomic DNA inserts	107
Figure 3.20	Average Trinucleotide content of the 12 sequenced genomic DNA inserts	108

Figure 3.21	Graphical representation of the global alignment of the 11 insert sequences	110
Figure 3.22	Local alignment of the 11 sequences visualized with ABC software	111
Figure 3.23	S/MAR potential (y-axis) with respect to relative base pair position of control plasmids	118
Figure 3.24	S/MAR potential (y-axis) with respect to base pair position of plasmids originating from the 3Mix genomic DNA library	119
Figure 3.25	S/MAR potential (y-axis) with respect to base pair position of plasmids originating from the 3Lig genomic DNA library	120
Figure 4.1	Theoretical effect of maintenance and replication elements on the total plasmid present in a cell population	127

LIST OF ABBREVIATIONS

μg	microgram
μL	microliter
AAV	adeno-associated virus
ACR	ARS consensus sequence
ARS	autonomous replication sequences
BACs	bacteria artificial chromosomes
bp	base-pair
BURs	base-unpairing regions
CHO	chinese hamster ovary
CMV	cytomegalovirus
DmORC	<i>D. melanogaster</i> origin recognition complex
DNA	deoxyribonucleic acid
EBNA-1	Epstein-Barr viral nuclear antigen 1
EBV	Epstien-Barr Virus
EGF	epidermal growth factor
GEs	genome equivalents
HEK	human embryonic kidney
IUPAC	International Union of Pure and Applied Chemistry
MCM	minichromosome maintenance
mL	milliliter
<i>neo</i>	neomycin
NPC	nasopharyngeal carcinoma
NPC	nuclear pore complex

ORCs	origin recognition complexes
pre-RC	pre-replication complexes
RNA	ribonucleic acid
S/MAR	Scaffold Matrix Attachment Region
SIDD	stress-induced duplex destabilization
SV40	Simian Virus 40
T-ag	large T-antigen
YACs	yeast artificial chromosomes

1.0 INTRODUCTION

The maintenance of plasmid DNA in mammalian cells is hindered by the cell's predisposition to lose plasmid DNA by various mechanisms such as: simple dilution by mitosis (Pauletti *et al.*, 1990), cellular degradation by DNases (Wilson *et al.*, 1992) or homologous recombination (Bollag *et al.*, 1989). The study of viruses such as SV40 (reviewed in Cooper *et al.*, 1997) and EBV (reviewed in Lupton and Levine, 1985) has given insight into a mechanism for circular episomal low molecular weight DNA maintenance in mammalian cells. However these mechanisms involve viral encoded proteins which induce immortality in certain cells (reviewed in Cooper *et al.*, 1997) and subsequently have been affiliated with cancer (Chang *et al.*, 2002). However in 1999, Piechaczek and others were able to use sequence from the 5' region of the human interferon- β gene to induce episomal plasmid stability in Chinese Hamster Ovary (CHO) cells.

The intended use of this plasmid maintenance technology is in the area of gene therapy research. Gene therapy is the transfer of genetic material into the cells of a target individual resulting in a therapeutic benefit, for the purpose of preventing, or curing a disease. Such modifications result in the correction of a genetic defect brought about by the absence or alteration of a protein, or the addition of genetic information which will modify the cell's characteristics. Gene and cell therapy can be used for the treatment of various common human diseases where conventional medical procedures fail to be successful.

There are three essential properties to an ideal gene therapy vector. First the vector must efficiently & accurately deliver the transgene of interest to

the specific target tissue. Second the vector must not elicit an adverse immune response or induce undesirable side effects such as cellular immortality, or even death (reviewed in Thomas *et al.*, 2003). Finally, the vector must provide sustained therapeutic expression levels *in vivo*. Gene therapy vectors can be placed in two broad categories, viral and non-viral (plasmid-based). Currently, clinical trials using viral vectors are in the majority (shown in figure 1.1); however each type has its advantages.

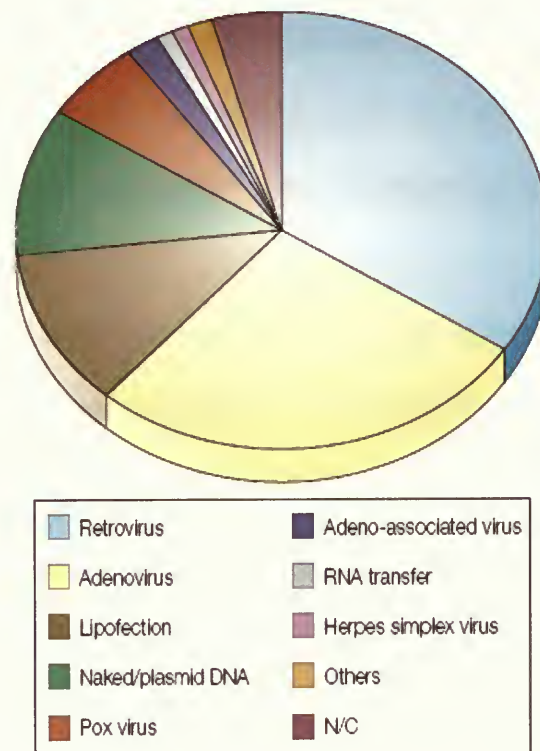


Figure 1.1: Survey of clinical trials using gene transfer technology.

Retrovirus and adenovirus vectors have been the most commonly used vectors in gene transfer trials while non-viral gene transfer has been used in roughly one quarter of all trials (reproduced from Thomas *et al.*, 2003).

1.1 Viral Vectors

Viruses are highly evolved biological vehicles with the innate ability to deliver genetic material to a specific target. Viral vectors harness this ability while suppressing other deleterious attributes associated with viral infection. The basic principal of using viral pathogens as delivery systems relies on an understanding of the viral life cycle, and the ability to separate the components needed for replication from those capable of causing disease (Verma & Weitzman, 2005). This is accomplished by deleting all, or some, of the coding regions from the viral genome but leaving intact the sequences which are necessary for the functionality of the vector (reviewed in Verma & Weitzman, 2005).

The most commonly used vectors are based on retroviruses, a large family of enveloped RNA viruses found in all vertebrates (reviewed by Thomas *et al.*, 2003; Fields *et al.*, 1996). Retroviruses are characterized by the integration of the transgene into the genome of the target cell; allowing for the transgene to be maintained in dividing cells. Integration is not a guarantee of stable transcription, as transgene expression from integrated vector genomes can be gradually silenced by methylation induced by nearby cellular DNA (reviewed by Thomas *et al.*, 2003; reviewed by van Craenenbroeck *et al.*, 2000; Remus *et al.*, 1999).

DNA viruses, adenovirus and the adeno-associated virus (AAV), have also been explored as potential vectors. The adenovirus is a non-enveloped virus with an approximately 36 kb double stranded DNA genome where as the

AAV is much smaller (Stoll & Calos, 2002). Early-generation adenovirus vectors showed potent immunogenicity and short-lived transgene expression, which are undesirable characteristics for most gene-therapy applications. However, this vector will likely find a niche in the treatment of vascular and coronary artery diseases (where transient transgene expression is advantageous) and for cancer treatment (where the cellular toxicity and immunogenicity might enhance anti-tumor effects) (reviewed by Thomas *et al.*, 2003).

Recombinant Adeno-Associated viral vectors (rAAV) are commonly based on the serotype 2 capsid and have been used to infect a large number of cell types and experimental model systems (reviewed in Verma & Weitzman, 2005). The AAV is unique among viruses in that the wild-type virus has never been shown to cause human disease. The small particle size of the non-enveloped capsid allows for the administration of high doses of vector systemically without eliciting an acute inflammatory response or toxic side effect (reviewed in Thomas *et al.*, 2003). The AAV genome is a linear single-stranded molecule of 4608 nucleotides and contains a 145-base terminal repeat sequence at each end (Collaco *et al.*, 1999).

1.2 Limitations of Viral Vectors

Viral vectors are efficient at the delivery of genetic material into a target cell. The limitations that occur in viral vectors are mainly due to the constraints of the physical space available in the vector genome for the incorporation of exogenous DNA (reviewed by Thomas *et al.*, 2003; Stoll &

Calos, 2002). Elimination of the non-essential viral genes creates more room for the incorporation of the desired genetic information (reviewed in Verma & Weitzman, 2005). Other limitations of viral vectors include the deleterious response of the immune system in the presence of viral capsid protein, and the (small yet significant possibility) of viral homologous recombination producing replicative competent viral vectors (reviewed in Verma & Weitzman, 2005; reviewed by Thomas *et al.*, 2003; Colaco *et al.*, 1999).

1.3 Artificial Chromosomes

Artificial chromosomes which are usually, but not always devoid of viral sequences are desirable vectors for gene therapy. They are usually composed of native mammalian chromosomal components (Figure 1.2) and

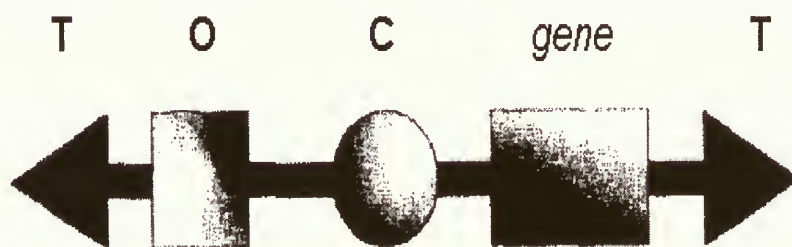


Figure 1.2: Schematic drawing of a mammalian artificial chromosome. The functional elements of this chromosome are telomeres (**T**), a replication origin (**O**) and a centromere (**C**). Any gene (**gene**) of interest could be inserted into such a construct (reproduced from Ascenzioni *et al.*, 1997)

require origins of replication, telomeres and a centromere (reviewed in Ascenzioni *et al.*, 1997; Calos, 1996). Upon successful construction, these vectors should be stably retained in the cell without integration and are present

in only one or two copies per cell (Ascenzioni *et al.*, 1997). The first linear artificial chromosomes were constructed in *Saccharomyces cerevisiae* from cloned telomere repeats, a replication origin sequence and a 125 bp centromere sequence which had been functionally identified using plasmids (Grimes & Cooke, 1998). Yeast artificial chromosomes (YACs) contain all of the DNA sequences necessary for yeast chromosome function and autonomous replication. They are the largest cloning vector available and have allowed the cloning of megabase fragments of mammalian DNA including gene clusters with associated locus control regions (reviewed in Tolmachova *et al.*, 1999).

The bottom up construction of mammalian artificial chromosomes has proven to be problematic since only the telomeres have been structurally defined (reviewed in Grimes and Cook, 1998). Mammalian centromeres have a minimum size of at least several hundred kilobases and the bulk of centromeric DNA is composed of tandem repeats of a 171 bp sequence known as alphoid DNA (reviewed in Calos, 1996). The chromatin associated with this region is tightly packaged due to a high concentration of scaffold matrix attachment region sequences (Ascenzioni *et al.*, 1997). Due to the large size of these mammalian elements stepwise assembly of a chromosome is impractical. A more efficient approach is the progressive reduction of an existing chromosome. Using telomere-mediated fragmentation, telomeric DNA is integrated at random chromosome breaks removing the natural end and forming a new one (reviewed in Lipps *et al.*, 2003). This procedure has led to the reduction of a human X chromosome from 164 Mb to 8Mb in size. These

minichromosomes replicate once per cell cycle and their mitotic stabilities were not noticeably different from those of normal chromosomes.

The large size of artificial chromosomes makes delivery of these molecules challenging, and the use of smaller, modified YACs or bacteria artificial chromosomes (BACs) presents an excellent alternative (Magin-Lachmann *et al.*, 2003; Nielsen *et al.*, 2000; Tolmachavo *et al.*, 1999). YACs containing isolated human origin sequences from human artificial chromosomes were shown to replicate, for at least one round of semi-conservative replication, and maintain an episomal un-rearranged form in HeLa cells under G418 selection (Nielsen *et al.*, 2000). A BAC system, carrying the *oriP* and EBNA-1 protein, allowed the stable maintenance of a 150 kb sequence, without rearrangement for 5 weeks in B16F10 murine melanoma cells (Magin-Lachmann *et al.*, 2003). The delivery of genomic DNA would enable genes to be transferred as complete loci, including regulatory elements, introns and native promoter regions, which might be important for the prolonged and tissue specific expression of the transgene (Wade-Martins *et al.*, 2000).

1.4 Viral-based Plasmid Vectors

Stability of a delivered transgene and its subsequent long term expression is one of the most important obstacles to gene therapy. The insertion of DNA into the host genome may cause insertional mutagenesis or silencing by *de novo* methylation of the introduced gene by nearby cellular DNA (reviewed by Craenenbroeck *et al.*, 2000; Remus *et al.*, 1999). The use

of replicating episomal vectors presents an appealing alternative to integrating viral vectors, as they associate with the host chromosome in a non-covalent manner but still segregate during mitosis (figure 1.3) (reviewed in Calos, 1998). Unlike viral vectors, plasmid vectors have almost no limitations on the size of the therapeutic transcription unit, with inserts in excess of 300 kb being possible (reviewed in Chow *et al.*, 2002; Wade-Martins *et al.*, 2000). Plasmid vectors are usually introduced into the target cells using non-viral delivery

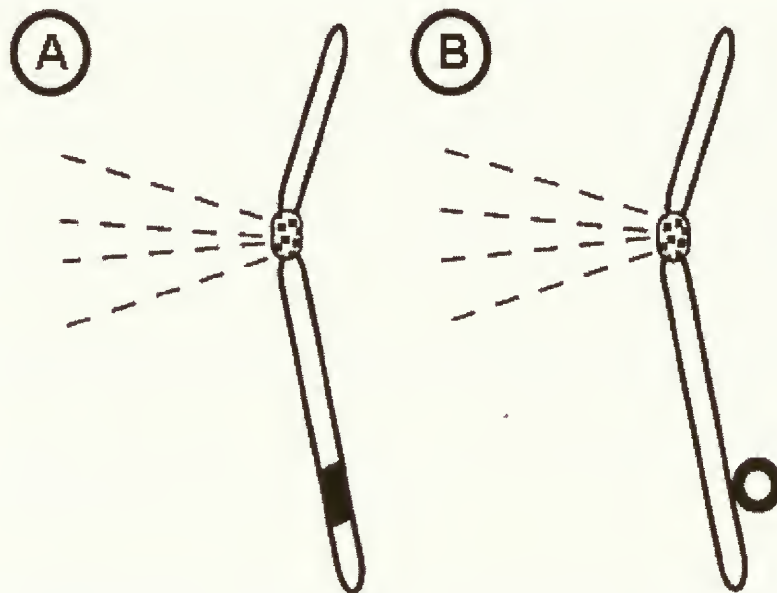


Figure 1.3: Maintenance mechanisms of viral genomes in mammalian cells. The virus exploits the chromosome centromere for long-term maintenance by (A) covalent integration or (B) non-covalent close association (reproduced from Calos, 1998).

systems which have the advantage of being non-pathogenic and non-immunogenic. These DNA-delivery methods are less efficient at introducing foreign DNA into the target cell than viral vectors, however unlike viral vectors

they lend themselves to the industrial “scaling-up” that is required for the economical production of therapeutic gene vectors (reviewed in Thomas & Klibanov, 2003, reviewed in Chow *et al.*, 2002).

Viral-based plasmid vectors developed from the understanding and characterization of viruses with genomes that could replicate episomally, and therefore were uncoupled from the once per cell cycle constraint of the host cellular genome (Craenenbroeck *et al.*, 2000; Fields *et al.*, 1996). One of the first vectors was derived from the polyomavirus simian virus 40 (SV40), which requires the Large T-antigen (T-ag) and the SV40 origin for the maintenance of an episomal vector (reviewed in Stenlund, 2003; Craenenbroeck *et al.*, 2000; Cooper *et al.*, 1997). These two viral elements allow for multiple copies of the vector to be produced, during the cell cycle, resulting in high levels of gene expression. The high copy number also aids in the maintenance of high expression by facilitating the vertical transfer of the episomes during cell division (reviewed in Cooper *et al.*, 1997).

EBV based vectors have become increasingly useful in gene therapy applications due to their large wild-type genome (165 kb) which is maintained as a large episome (Fields *et al.*, 1996). EBV is a member of the family *herpesvirinae* which establishes latent infection in lymphocytes and is associated with cell proliferation (Craenenbroeck *et al.*, 2000). Unlike SV40 which can replicate throughout the cell cycle, EBV is faithfully synthesized only once during the S-phase of the cell cycle (Aiyar *et al.*, 1998). Synthesis and maintenance of this episome is mediated by a viral sequence called *oriP*, and

a single viral protein, the Epstein-Barr viral nuclear antigen 1 (EBNA-1) (Lupton & Levine, 1985).

The viral origin, *oriP* consists of a 24-member family of tandem 30 bp repeats which may be involved in plasmid maintenance; a region of dyad symmetry, and four partial copies of the 30 bp repeats located 1Kb away which are important for EBNA-1 binding (reviewed in Bode *et al.*, 2001; reviewed by Krysan *et al.*, 1989; Lupton & Levine, 1985). Mutations in the EBNA-1 protein inhibit the stable maintenance of the EBV vector without affecting the ability of the episome to replicate in a transient assay. Deletion of the dyad sequence does not permit replication, however when replaced by human genomic DNA sequences, replication can be rescued in long-term assays (Krysan *et al.*, 1989; Lupton & Levine, 1985). Finally, it has been found that plasmids which do not contain the family of repeats region or EBNA-1 are lost rapidly from cells (Aiyar *et al.*, 1998). This suggests that in EBV-based vector maintenance is independent of replication but is mediated by the same protein. If this protein or the sequence which it binds is lost, the plasmid is rapidly degraded by cellular mechanisms. Even with all of the components present, the retention mechanism is imperfect resulting in a slow loss (2-8% per cell generation) over time, in the absence of selection (Sclimenti & Calos, 1998).

1.5 Limitations of Viral-based Plasmid Vectors

EBV is closely associated with several malignancies originating from lymphoid or epithelial cells, such as endemic Burkitt's lymphoma and

nasopharyngeal carcinoma (NPC). EBV has also been shown to persist in an integrated state in the genomes of nasopharyngeal carcinoma cells (Chang *et al.*, 2002). The major drawback for these viral-based episomal expression vectors in gene therapy is that they require the presence of a viral *trans*-acting factor, which could lead to the transformation of transfected cells (reviewed in Craenenbroeck *et al.*, 2000). Specifically, the large T-antigen associates with host tumor suppressor gene products, including p53, retinoblastoma, and retinoblastoma-related proteins (reviewed in Craenenbroeck *et al.*, 2000; reviewed in Cooper *et al.*, 1997).

1.6 Non-Viral Plasmid Vectors

Taking qualities from both viral-based vectors and artificial chromosomes, non-viral plasmid vectors attempt to be perceived by the cell as a chromosome, while still allowing the ease of replication in bacterial cells. Both replicative and non-replicative plasmid have been shown to assemble into minichromosomes when introduced to mammalian cells, however only 60-70% of the plasmids reaching the nucleus after calcium phosphate transfection are actually assembled into the minichromosome (Reeves *et al.*, 1985). Without the addition of a maintenance sequence, both replicative and non-replicative plasmid DNA is lost from the cell over time by, simple dilution of the plasmid (Pauletti *et al.*, 1990), cellular degradation by DNases (Wilson *et al.*, 1992) or homologous recombination (Bollag *et al.*, 1989).

There is only one non-viral plasmid vector available which replicates episomally and does not contain any viral encoded proteins. This plasmid is

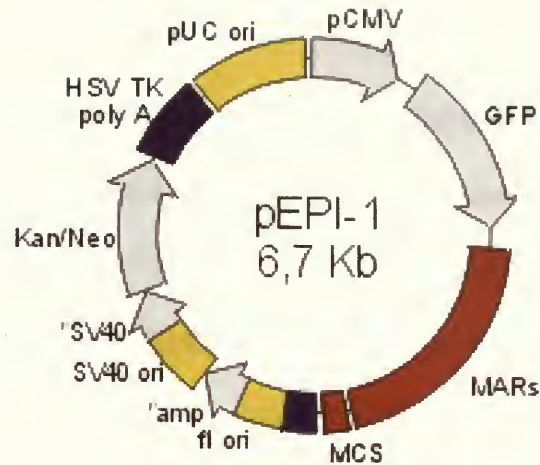


Figure 1.4: Non-Viral episomally replicating vector based on the SV40 origin of replication. The plasmid also contains a Neomycin cassette for antibiotic selection in mammalian cells. A second marker (Green Fluorescent Protein; GFP) driven by the Cytomegalovirus promoter (pCMV) contains a Scaffold/Matrix Attachment Region (S/MAR) downstream of the transcribed region (reproduced from Piechaczek *et al.*, 1999).

constructed with an SV40 origin of replication and promoter system. The large T-antigen is however replaced by a genomic scaffold/matrix attachment region (S/MAR) DNA sequence originating from the 5'-region of the human interferon β -gene. Piechaczek *et al.* (1999) showed that the plasmid pEPI-1 (figure 1.4) replicated efficiently and episomally in Chinese hamster ovary (CHO) cells for more than 100 generations without selection. They also observed an absence of random integration of the plasmid; however the plasmid did integrate into the host genome when only carrying the SV40 *ori* or the S/MAR region. The absence of integration inhibits epigenetic silencing of the transgene by cytosine methylation (Jenke *et al.*, 2004a).

Analysis of 10 different clones by fluorescent *in situ* hybridization on metaphase spreads over time revealed that the number of vector molecules per cell were between 4 and 13 copies and that these figures did not significantly change between 20 and 50 generations after transfection (Baiker *et al.*, 2000). Determined by nuclear fractionation and southwestern analysis, the S/MAR containing vector bound to proteins found in the filament nuclear matrix network (Baiker *et al.*, 2000) such as scaffold attachment factor A (SAF-A) (Jenke *et al.*, 2002). Therefore these S/MARs may be able to recruit cellular factors from the nuclear matrix that have a similar function to the SV40 large T-antigen by enhancing the extrachromosomal stability of the plasmid.

Scaffold matrix attachment region sequences have a high A/T nucleotide content and range in size from hundreds to thousands of base pairs (van Driel & Otte, 1997). However, the AT-rich sequences NTS-1 and NTS-2 which are associated with endogenous origin of replication in mouse cells, are unable to support an episomal state (Baiker *et al.*, 2000). Therefore A/T nucleotide content is not sufficient to confer S/MAR characteristics which may be related instead to topological or structural features which are not strictly linked to the primary sequence (Benham *et al.*, 1997). Minimal functional analysis of the human interferon β -gene S/MAR region has revealed a 155-bp element that as a tetramer can functionally replace the original S/MAR sequence when placed downstream from a transcribed region (Jenke *et al.*, 2004b).

The mechanism by which this occurs is similar to the Epstein-Barr virus mode of maintenance (Figure 1.3) during which EBNA-1 facilitates the attachment of the EBV genome to the host chromosome during mitosis. This close association of the vector and the nuclear matrix allows the vector access to cellular replication machinery, and explains the mitotic stability and the once per cell cycle replication seen in early S phase (Jenke *et al.*, 2004b). This mode of replication in eukaryotic cells is further supported by observations that the nuclear scaffold matrix is necessary for replication of host genomic DNA (reviewed in Frouin *et al.*, 2003).

1.7 Scaffold Matrix Attachment Regions (S/MARs)

Eukaryotic genomes are thought to be organized into independently regulated chromatin loop domains which contain coding and non-coding sequences as well as *cis*-regulatory DNA elements (Miekle *et al.*, 2002). A relatively novel addition to the class of *cis*-regulatory DNA sequences are Scaffold Matrix Attachment Regions (S/MARs) which are thought to interact with the protein backbone of the nucleus (Bode *et al.*, 1998) and act as the physical anchor of chromatin loop domains (van Driel & Otte, 1997). S/MARs have also been implicated in a variety of functions such as genome organization, and may also contribute to chromosome condensation during nuclear division (Liebich *et al.*, 2002). S/MARs are usually found at putative borders of several chromatin domains and in close association with certain enhancers (Bode *et al.*, 1998) leading to increased transcription activity (Liebich *et al.*, 2002). S/MARs also contribute to the enhancer-mediated

control of the immunoglobulin-gene demonstrating that together, enhancers and S/MARs are sufficient for a functional locus control region (Lipps *et al.*, 2003).

S/MARs are typically A/T rich (~70% A/T content) and contain arrays of oligo dA/dT called A-tracts. Some A-tracts contain sequences similar to the topoisomerase II consensus sequence (A)ATATTT(TT), however there is no S/MAR consensus sequence (van Driel & Otte, 1997). Some structural features of S/MARs include bending, a narrow minor groove due to the increased content of oligo(dA/dT) and single strandedness (Bode *et al.*, 1992). This propensity for sequence induced base-unpairing has resulted in a precise correlation between S/MARs and sites of predicted stress-induced DNA duplex destabilization (SIDD) (Mielke *et al.*, 2002; Bode *et al.*, 2001).

S/MARs were originally identified as DNA with a high binding affinity for the nuclear matrix and associated proteins (Bode *et al.*, 1992; van Driel & Otte, 1997). Several S/MAR binding proteins have been characterized, including ubiquitous and abundant proteins such as topoisomerase II, histone H1, lamin B1, HMG I/Y, and nucleolin (reviewed in Kipp *et al.*, 2000). Elucidation of the molecular mechanism behind binding specificity is not complete even though these proteins have been thoroughly characterized (van Driel & Otte, 1997). However the mechanism behind binding of both the Scaffold Attachment Factor-A (SAF-A) and High Mobility Group protein (HMG I/Y) has been identified. SAF-A binding of DNA is mediated by a multiple SAF-box motif which is structurally related to homeodomains and is known to fold into a hook-like structure composed of two α -helices separated by a turn

(Kipps *et al.*, 2000). HMGI/Y binds DNA through the small AT-hook motif containing a sequence pattern centered on the Glycine-Arginine-Proline (GRP) tripeptide (Aravind & Landsman, 1998). Both motifs are evolutionarily highly conserved and interact with DNA through a mass binding mechanism which is characterized by the binding of a large number of low-affinity and specificity binding motifs that together produce a high-affinity and specificity interaction (Kipp *et al.*, 2000).

The current experimental methods for the identification of S/MAR regions requires considerable effort and is unsuitable for large scale screening of genomic sequences (Frisch *et al.*, 2001). *In silico* methods however provide a functional means to screen large amounts of sequence data. There are currently three software tools available for the prediction of S/MARs all of which demonstrates the feasibility of *in silico* methods to solve this problem of screening huge amounts of information.

Mar-Finder (Singh *et al.*, 1997) is based on the statistical occurrence of S/MAR motifs described as consensus sequences based on the International Union of Pure and Applied Chemistry (IUPAC) code for nucleotide sequences. These motifs are based upon the conserved characteristics for the origins of replication, TG-rich sequences, curved DNA, kinked DNA, topoisomerase II sites, and AT-rich sequences (Singh *et al.*, 1997). The stress-induced duplex destabilization (SIDD) program (Benham *et al.*, 1997) allows for the visualization of base-unpairing regions (BURs) by calculating and plotting the incremental free energy required to induce separation of each base pair in a supercoiled molecule (Benham *et al.*, 1997; Benham *et al.*, 1993;).

Finally SMARTest (Frisch *et al.*, 2001) is based on a density analysis of S/MAR-associated patterns represented by a weighted matrix library. The weighted matrices were derived from known S/MAR sequences and can be easily updated when new experimental data are available. Unfortunately the use of *in silico* bioinformatics tools is plagued with the propensity of false-positives and false negatives and should be used as a starting point for future experimental analysis (Frisch *et al.*, 2001; Benham *et al.*, 1997; Singh *et al.*, 1997).

1.8 Initiation of DNA Replication in Eukaryotes

Advances in the study of the initiation of eukaryotic DNA replication are due to the improved understanding and study of minimal systems such as viral-based vectors. Initiation of DNA replication generally involves four steps: recognition (figure 1.5A), melting (figure 1.5B), unwinding (figure 1.5C) and recruitment (figure 1.5D) of replication factors. The cellular initiator proteins from both eukaryotes and prokaryotes carry out a subset of these functions (reviewed in Stenlund *et al.*, 2003), and can be classified into three categories based on function.

The first and simplest category is represented by the origin recognition complexes (ORCs) which are the conserved initiator in all eukaryotes. ORCs are responsible for two functions in DNA replication; 1) recognition and binding to the replicator and 2) recruitment of the factors that are required for initiation (Reviewed in Stenlund *et al.*, 2003). The second category of initiator proteins include the DnaA protein, which is the ubiquitous initiator in bacteria which, in

addition to recognition and recruitment, is directly involved in the melting of the DNA template (Reviewed in Stenlund *et al.*, 2003).

The final category includes initiators from DNA viruses such as the papillomavirus E1 protein, and the polyomavirus large T-antigen, which, in addition to recognition, melting and recruitment, facilitates the unwinding of the DNA in front of the replication fork in a DNA helicase capacity (Reviewed in Stenlund *et al.*, 2003).

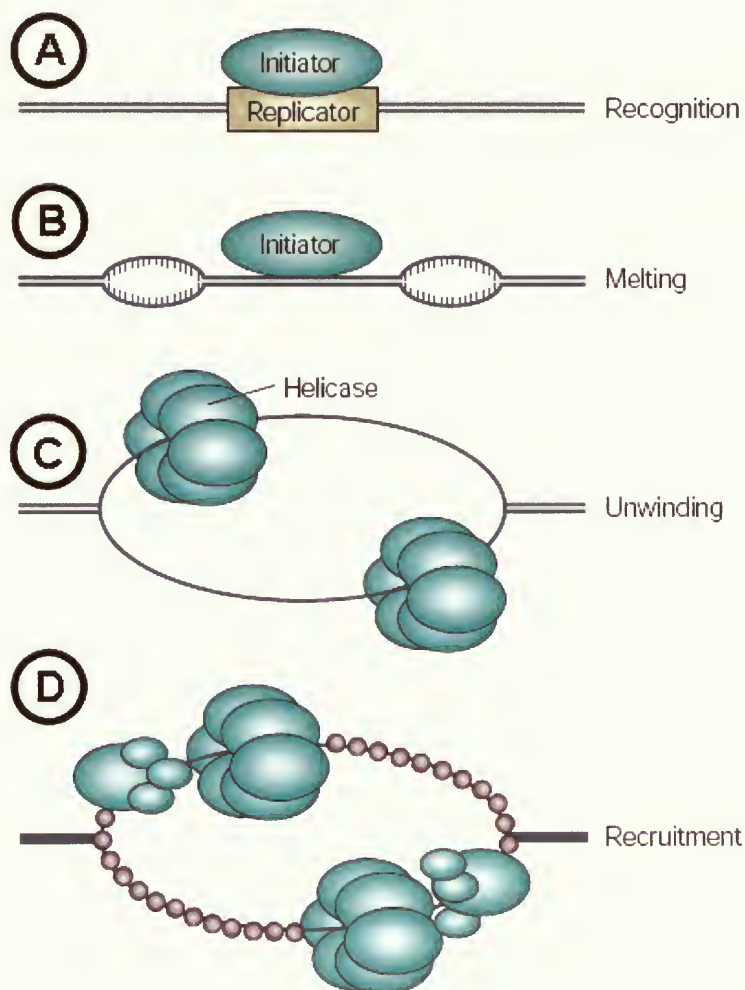


Figure 1.5: Initiation of DNA replication which can involve four steps: recognition (A), melting (B), unwinding (C) and recruitment (D) of replication factors (reproduced from Stenlund *et al.*, 2003).

Studies of the replication of plasmid DNA containing the viral SV40 origin originally identified DNA polymerase- α /primase, polymerase- δ , replication protein A (RPA), RNase H1 and DNA Ligase as the essential components of the eukaryotic replisome (reviewed in Frouin *et al.*, 2003). The metabolism of nucleic acids at the replisome is thought to be carried out in spatially organized domains of the nucleus. These domains are organized by non-chromatin nuclear structures such as the nuclear matrix, and the nucleoskeleton (Reviewed in Fujita *et al.*, 2002). S/MAR sequences are also found frequently at sites close to known DNA replication origins in the mammalian genome (Reviewed in Fujita *et al.*, 2002; reviewed by Shimizu *et al.*, 2001). Therefore it can be hypothesized that S/MAR sequences and some of the associated proteins may also play a role in the initiation of DNA replication, in addition to episomal plasmid maintenance.

In mammalian cells, preparation for DNA replication begins in the last phases of mitosis and continues into G1-phase, when the ORC, Cdc6, Cdt1 and other proteins cooperate to load the minichromosome maintenance (MCM) proteins onto chromatin to forming the pre-replication complexes (pre-RC) at potential replication origins (reviewed in Huberman, 2003). The processes of ORC binding and MCM protein loading are highly regulated by post-translational modifications and accessory proteins (Cdc6 & Cdt1). However, the sequence to initiate binding of the ORC to the chromosomal DNA appears to be random. There is a preference of the ORC to bind to A/T rich regions in the mammalian genome (Vashee *et al.*, 2003), which may explain why many different DNA fragment can serve as replicators in human

cells (Krysan *et al.*, 1989), suggesting that the initiation of DNA replication requires more than nucleotide sequence alone.

Preliminary elucidation of consensus sequence required for DNA replication produced a 36-bp human consensus sequence which determines autonomous DNA replication in eukaryotic cells (Price *et al.*, 2003). The degeneracy of this consensus sequence is quite extensive, however sequence homologies have been observed in CpG islands, within which eukaryotic DNA replication has been shown to initiate (reviewed in Price *et al.*, 2003). Experiments using episomal plasmids as a minimal system to study DNA replication demonstrate that DNA replication is contingent upon the interactions of many different cellular functions such as transcription, nuclear matrix interaction and chromatin structure (Stehle *et al.*, 2003). This epigenetic mechanism in concert with specific sequences likely determines the initiation sites for mammalian DNA replication (Schaarchmidt *et al.*, 2004).

1.9 Autonomous Replicating Sequences (ARS)

The duality of replication initiation and extrachromosomal maintenance is best defined in the model organism *Saccharomyces cerevisiae* which has a well defined AT-rich replication origin sequence that when integrated into a plasmid can promote autonomous replication (reviewed in Stehle *et al.*, 2003, Clyne & Kelly, 1997). Autonomous Replication Sequences (ARS) were discovered during research into the *LEU2* gene when researchers observed integration of some plasmids into the yeast chromosome, while other plasmids remained episomal and capable of replication without integration (reviewed in

Clyne & Kelly, 1997). Now known as the ARS assay, ARS elements could be determined by isolation of the episomal elements, while under metabolic or antibiotic selection (Clyne & Kelly, 1997). The identification and characterization of these ARS ultimately lead to the identification of the first ORC which are initiator proteins that recognizes an 11-bp ARS consensus sequence (ACR) in an ATP-dependent manner (reviewed in Vashee *et al.*, 2003). Recently it has been found that these ARS regions have an increased susceptibility to SIDD, on average less free energy is needed to separate the DNA duplex then their immediate surrounding sequences, this makes these sequences more than 2000 times easier to open (Ak & Benham, 2005).

Homologues of ORC subunits have been identified in other eukaryotic species (*Schizosaccharomyces pombe*, *Xenopus laevis* and *Drosophila melanogaster*) and in several cases genetic or biochemical data have indicated that the same ORC subunits are required for DNA replication (reviewed in Chuang & Kelly, 1999). Since the ORC appears to be highly conserved in eukaryotes, it is likely that there is a common mechanism for the initiation of eukaryotic DNA replication. However unlike *S.cerevisiae*, other eukaryotes do not have the same concise sequence for the initiation of DNA replication and therefore their origin sequences have been harder to locate (Clyne & Kelly, 1997). Using the ARS assay method, ARS sequences have been identified in *S. pombe*. At 500-1000 bp in size *S. pombe* ARS are longer than *S.cerevisiae* ARS and do not contain a consensus sequence, but possess elements rich in AT base pairs which serve as binding sites for ORC (Chuang & Kelly, 1999). Other experimentally characterized eukaryotic ARS

elements have revealed that sequence similarity between the elements is limited to short stretches of AT base pairs (reviewed in Schaarschmidt *et al.*, 2004).

The apparent lack of sequence similarity between ARS elements in *S.pombe* is similar to the lack of sequence homology within mammalian origins, with the exception of a statistically higher AT content (Vashee *et al.*, 2003). The high AT content in ARS elements of *S.pombe* is important for the functionality of the ORC, a protein complex made up of 6 different protein subunits (ORC1-6), within which ORC4p (origin recognition complex subunit 4 protein) preferentially binds AT-rich sequences (Chuang & Kelly, 1999). The interaction between the ORC of *S.pombe* and the AT-rich regions is mediated by the N-terminal domain of the ORC4p which contains nine copies of the AT-hook motif (Lee *et al.*, 2001; Chuang & Kelly, 1999). AT-hook motifs are known (from biochemical and structural studies) to mediate binding to the minor groove of AT-tracts which are approximately 4-6 nucleotides in length (reviewed in Chuang & Kelly, 1999). The *D. melanogaster* origin recognition complex (DmORC) has also been shown *in vivo* and *in vitro* to bind the AT-rich chorion gene amplification control elements *ACE3* and *ori-β*, however no distinct ARS like element has been determined at these sites (reviewed in Vashee *et al.*, 2003).

1.10 Plasmid Delivery

The plasmid enters the mammalian cell by means of a delivery mechanism such as a molecular conjugate or viral vector (reviewed in Verma

& Weitzman, 2005; reviewed in Schaffer & Lauffenburger, 1998). Molecular conjugates, such as calcium phosphate are not as efficient nor specific as viral vectors, however for use in basic research they are quite effective with transfection efficiencies being reported as high as 30-60% (Jordan *et al.*, 1996). The majority of naked plasmid DNA never reaches the cell and is degraded by serum nucleases, in particular supercoiled plasmid DNA has a half-life of 1.2 minutes in rat plasma (figure1.6A); which can be extended by the use of molecular conjugates (Houk *et al.*, 1999). Those plasmids which make it to the cellular membrane must find a way into the cell by endocytosis (figure1.6B) which could be facilitated by piggybacking during the internalization of receptors such as the epidermal growth factor (EGF) receptor or fibroblast growth factor receptor (Schaffer & Lauffenburger, 1998). Upon endocytosis the plasmid DNA must escape from the endocytic network or risk being degraded by intra-cellular DNAses (figure1.6C) (reviewed in Schaffer & Lauffenburger, 1998).

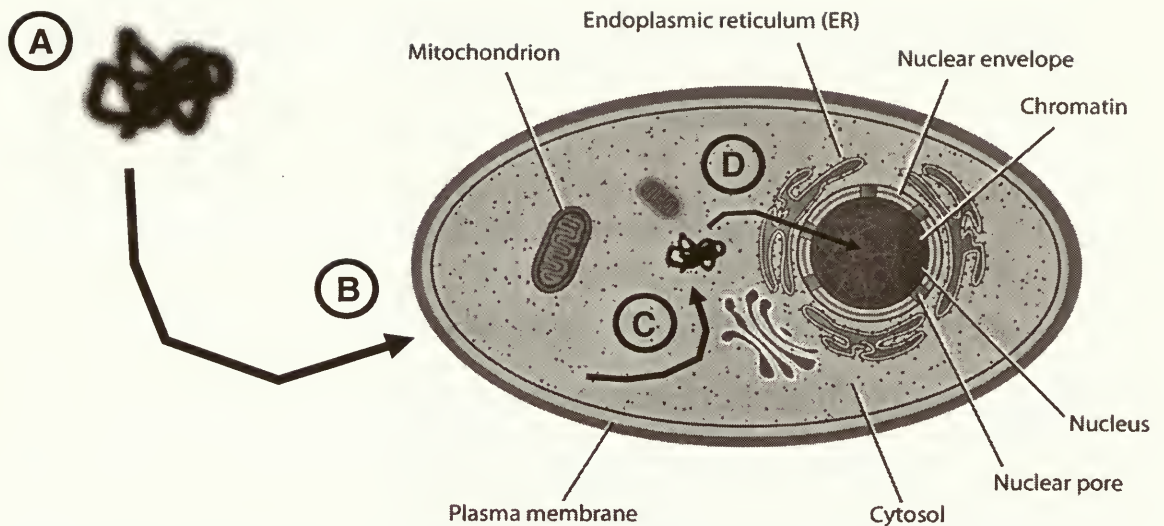


Figure 1.6: Delivery of plasmid DNA and the associated barriers which it must overcome in order to enter the nucleus. The plasmid DNA (A) must first pass through the plasmid membrane (B) by means of endocytosis. Once inside the cell the plasmid DNA must escape the endocytic network (C) and enter the nucleus during mitosis or through the nuclear pore complex (D) (<http://www.press.uillinois.edu/epub/books/brown/images/fig8.1b.gif>, Accessed August 10, 2005).

The final intracellular barrier is entry into the nucleus of the cell (figure 1.6D). In rapidly dividing cells entry can be facilitated by the disassembly of the nuclear envelope during mitosis, thus eliminating the nuclear barrier for a short period of time (reviewed in Dean *et al.*, 2005). Nuclear import of plasmid DNA into non-dividing cells is more complicated, however it is essential for many viral life cycles as well as for the success of many gene therapy protocols (reviewed in Dean, 1997). The nuclear import and export of proteins is facilitated by the nuclear pore complex (NPC) by

means of a nuclear localization signal (Dean *et al.*, 2005). However, nuclear import of endogenous DNA is sequence specific and in particular the SV40 genome contains this sequence within the promoter and origin region (Dean, 1997).

1.11 Isolation of Low Molecular Weight DNA

The isolation of low molecular weight DNA from mammalian cells, though required for this project, also has uses in the study of other eukaryotic systems such as cDNA expression cloning (Bowers *et al.*, 1999), recombination (Bollag *et al.*, 1989) and construction of viral episomal vectors (Lupton & Levine, 1985). This process involves the separation of a very small quantity of plasmid DNA from a much larger quantity of genomic DNA. Low molecular weight DNA has traditionally been isolated using the Hirt extraction method which is contingent upon the preferential precipitation of un-degraded cellular DNA in the presence of SDS and NaCl (Hirt, 1967). However this method involves an extensive precipitation, phenol/chloroform extraction and 2-3 days of time (Siebenkolten *et al.*, 1995).

Plasmid DNA exists in 3 possible forms (not including aggregates (concatamers)), supercoiled, open circular and linear. Supercoiled plasmid is a circular double-stranded DNA that has increased torsional pressure about the right handed-turn of the double helix (Dale, 1998). This causes the macromolecule to twist upon itself and travel the furthest in an agarose gel due to its compact size (Clemson & Kelly, 2003). If a break in the phosphate back bone occurs in a supercoiled plasmid the torsional pressure is released

and the plasmid uncoils to become the open circular or relaxed form. If a second break occurs on the opposite phosphate sugar backbone the molecule can no longer maintain an opened circular form and is therefore said to be linear. The supercoiled form of the plasmid exhibits the maximal biological activity, and is therefore the form which is sought after when purifying plasmid DNA for use in vaccines and gene-therapy applications (Clemson & Kelly, 2003)

In bacteria, the most common method for the isolation of plasmid DNA employs the use of rapid changes in pH to separate the genomic DNA from the plasmid DNA. This method uses 3 solutions each of which are important and can not be changed and switched with the other. The resuspension solution is an isotonic solution which is added first to allow the cells to disassociate from one another and facilitate maximal efficiency during the lysis processes. The lysis solution is an alkaline solution containing an anionic detergent (SDS) that binds to proteins and causes linearity in them, and also causes instability in most biological membranes. The NaOH in the solution aids this process by reacting with the cell wall and inducing pores. However the essential property of an alkaline pH is its selected denaturation of chromosomal DNA at pH 12-12.5. When supercoiled DNA is exposed to >pH 13 it gives rise to a denatured form which is not readily renaturable (Birnboim & Doly, 1979). The optimum concentration of SDS/NaOH depends on the number of cells (or biomass) being processed (Clemson & Kelly, 2003). The addition of the third solution neutralizes the alkalinity of the solution and stops the lysis process. This rapid neutralization causes the large genomic DNA to

aggregate during the hybridization process; however the much smaller plasmid remains in solution. The high concentration of sodium acetate in the neutralization solution causes the precipitation of protein-SDS complexes and of high molecular weight RNA (Birnboim & Doly, 1979). Plasmid DNA can then be absorbed from the supernatant onto glass or silica in the presence of chaotropic salts. This makes phenol/chloroform extraction and ethanol precipitation unnecessary (Carter & Milton, 1993).

Commercially available kits which employ an alkaline lysis front end and a silica based column back end are numerous (Qiagen, Eppendorf, Sigma, Norgen & Promega) and produce similar recoveries. Use of these commercially available kits for the isolation of episomes from mammalian cells is limited, with only one publication using Promega's Wizard Miniprep DNA Purification System. Bowers *et al.* (1999) used the plasmid pCEP4 for the expression cloning of potential oncogenes in nonmalignant breast epithelial cell line MCF-10A. The plasmid pCEP4 contains an Epstein Barr virus origin of replication and the Epstein-Barr nuclear antigen-1 (EBNA) gene which facilitates the episomal replication of the plasmid in eukaryotic cells. After selection for 21 days and then additional days for clonal propagation pCEP4 episomes were consistently harvested from transfected cells (516 ± 112 pg/harvest, mean \pm standard deviation; $n=11$). Qiagen also includes an application note which compares the QIAprep Spin plasmid kit with the Hirt extraction in the isolation of plasmid from transiently transfected murine pre-B cell lines (Siebenkolten G., *et al.*, 1995). Using the transformation of *E.coli* as

a detection method, it was concluded that the QIAprep procedure yielded equivalent amounts and quality of plasmid DNA.

1.12 Project objectives

The objective of this project is to isolate and characterize human genomic DNA sequences which, when inserted into plasmids, enhance plasmid stability in mammalian cells.

2.0 MATERIALS AND METHODS

2.1 Bacterial Strain

The bacterial strain used for cloning and propagation of plasmid DNA was *Escherichia coli* DH5 α , with the genotype (F^- , $\Phi 80$, $\delta lacZ$, $\Delta M15$, $\Delta(lacZYA-argF)$, U169, *deoR*, *recA1*, *endA1*, *hsdR17*(rk^- , mk^+), *phoA*, *supE44*, *thi-1*, λ^- , *gryA96*, *relA1*) (Woodcock *et al.*, 1989).

2.2 Propagation and Maintenance of Bacterial Cultures

Bacterial cultures were grown in sterile Luria-Bertani broth (LB; 1% bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl, 4N NaOH to pH 7.1) overnight at 37°C. For long-term storage, cells were frozen in 15% glycerol, and stored at -70°C. To propagate culture from frozen stocks, cells were thawed then equally distributed into larger volumes of LB. Bacterial colonies were grown at 37°C on solid LB medium (LB plates) containing bacto-agar (20% w/v) and stored at 4°C for short periods of time.

2.3 Preparation of Competent Cells

2.3.1 Sambrook *et al.* Method (1989)

Competent bacterial cells were prepared according to Sambrook *et al.* (1989). *E.coli* was streaked on agar plates to allow for the isolation of a single colony. Half of the single colony was inoculated into 2 mL of LB and the other half of the colony was inoculated into 2 mL of LB supplemented with 100 μ g/mL ampicillin and grown overnight in a 37°C shaker bath. If the colony inoculated into LB with ampicillin showed no growth, then a 50 mL flask of LB was inoculated with 2 mL of overnight culture from the LB tube and grown to a mid-

log phase ($O.D._{600} = 0.45-0.55$). The cells were then placed on ice for 15 minutes and centrifuged for 10 minutes at $2700 \times g$ at $4^{\circ}C$ in a Beckman GPR centrifuge with 50 mL conical tube bucket inserts. The supernatant was decanted; the cells were resuspended in 20 mL of cold transformation buffer (75 mM $CaCl_2$, 5 mM Tris, pH 7.6) and incubated for one hour on ice. The cells were centrifuged again at $2700 \times g$ for 10 minutes at $4^{\circ}C$, the supernatant was decanted, and the cells were resuspended in 4 mL of cold transformation buffer and incubated on ice for 45 minutes. For long-term storage, 50% glycerol was added to the cells to a final concentration of 15%. The cell suspension was placed in 1.5 mL sterile Eppendorf tubes in 200 μ L aliquots and stored at $-70^{\circ}C$ for up to 3 months.

2.3.2 Inoue et al. Method (1990)

Competent bacterial cells were prepared according to Inoue *et al.* (1990). Single colonies were obtained from a stock cell population and inoculated on plates containing LB with no antibiotics. A single colony was isolated and grown in 2mL of LB overnight at $37^{\circ}C$ in a shaker bath. The inoculum was transferred the next day to 200mL of SOB Medium (2% (w/v) bacto tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, pH 7.0) and grown to mid-log phase ($O.D._{600} = 0.6$). The culture was aliquoted into 50 mL conical tubes and placed on ice for 10 minutes. The bacterial culture was then centrifuged for 10 minutes at $2700 \times g$ at $4^{\circ}C$ in a Beckman GPR centrifuge with 50 mL conical tube bucket inserts. The supernatant was removed and the cells were resuspended in 64mL transformation buffer (TB) (10mM Pipes, 55 mM $MnCl_2$, 15 mM $CaCl_2$, 250mM KCl, pH 6.7), and placed on ice for 10 minutes.

The cells were centrifuged again for 10 minutes at 2700 x g at 4°C. The supernatant was removed and the cells were then gently resuspended in TB and 7% DMSO (Dimethylsulfoxide) and placed on ice for 10 minutes. The competent cells were then aliquoted and frozen immediately at -70°C.

2.4 Bacterial Transformation

Plasmid DNA was suspended in water or TE buffer (1mM ethylene diamine tetra-acetic acid (EDTA), 10mM TrisHCl) to a maximum volume of 20 µL. Frozen competent *E.coli* stocks were thawed at room temperature, and 200 µL was then used to inoculate the plasmid DNA suspension. The mixture was subsequently incubated at 0-4°C for 30 minutes in an ice water bath. At the end of 30 minutes, the samples were subjected to a 43°C heat-shock for 45 seconds, and then returned immediately to the ice water bath for an additional 2-5 minutes. To encourage bacterial out-growth, 1 mL of SOC Medium (2% (w/v) bacto tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20mM Glucose, pH 7.0) was added to the mixture, which was then placed at 37°C for 1 hour.

For samples which had a large or more than 1 ng of DNA initial amount of plasmid DNA (>1ng), 50 to 200 µL aliquots were plated on LB plates containing the appropriate antibacterial selection agent. For plasmids such as pUC19 and pCMV-beta ampicillin (100 µg/mL) was used as the selecting agent in the LB plates. For plasmids which contained the Neomycin resistance gene (pUC19neo) both kanamycin (25 µg/mL) and ampicillin (50 µg/mL) was used in the LB plates. Finally all plates contained 5-bromo-4-chloro-3-indolyl-bD-galactoside (X-Gal; 60 µg/mL).

2.5 Small Scale Plasmid DNA isolation from Bacteria Cells

2.5.1 Birnboim and Doly Method (1979)

Single colonies grown on LB plates containing antibiotics were picked and grown overnight in 2 mL of liquid LB media containing the same antibiotics. The overnight culture was transferred to a sterile 1.5mL microfuge tube and centrifuged at 16000 x g for 2 minutes in a Eppendorf 5402 desktop centrifuge. The medium was removed by vacuum aspiration and the cell pellets resuspended in 100 μ L of sterile resuspension solution (50 mM Tris-HCl (pH7.5), 10 mM EDTA, 100 μ g/mL RNase A) per tube by vortexing. After 3-5 mins 100 μ L of cell lysis solution (1% w/v sodium dodecyl sulphate (SDS), 0.4 N NaOH) was added to each tube and mixed by inversion for 3-5 minutes. The solution was neutralized by adding 100 μ L of neutralization solution (3M sodium acetate (NaOAc), pH 4.75) to each tube and inverting the tube 6-8 times, the samples were then placed in the fridge at 4°C for 30 minutes.

Samples were transferred from the fridge to a 4°C Eppendorf 5402 desktop centrifuge and centrifuged for 10 minutes at 16000 x g. The lysate was transferred from the microfuge tube into a fresh 1.5 mL microfuge tube. Three volumes of cold 95% Ethanol (900 μ L) was added to each sample, the tubes were then placed at -20°C for 20 minutes to facilitate plasmid DNA precipitation. The samples were then centrifuged at 16000 x g for 10 minutes in a 4°C Eppendorf 5402 desktop centrifuge, the supernatant was decanted and the tubes dried for 10 minutes. The DNA pellet was gently resuspended in 200 μ L of autoclaved water and three volumes of cold 95% ethanol was added. The samples were incubated at -20°C for 20 minutes, and centrifuged for 10 minutes

at 16000 x g in a 4°C Eppendorf 5402 desktop centrifuge. The supernatant was decanted and the tubes were evaporated to dryness. The plasmid DNA was then resuspended in 80 µL of TE Buffer.

2.5.2 Norgen Method

The Norgen Biotek Corp. (St.Catharines, Ontario, Canada) MiniPrep plasmid DNA isolation kit was used for the isolation of plasmid DNA from bacterial colonies after transformation with plasmid isolated from mammalian cells. Single colonies grown on LB plates containing antibiotics were picked individually and grown overnight in 2 mL of liquid LB medium containing the same antibiotics. The overnight culture was transferred to a sterile 1.5mL microfuge tube and centrifuged at 16000 x g for 2 minutes in an Eppendorf 5402 desktop centrifuge. The solutions and method were carried out in accordance with the Norgen Mini prep Plasmid DNA Isolation kit.

2.6 Restriction Enzyme Digestion of Plasmid DNA

Procedures involving specific restriction enzymes were carried out in accordance with the manufacturer's guidelines (New England Biolabs or Fermentas). Incubation times varied with respect to the amount of DNA found in the sample in proportion to the concentration of enzyme.

Typically for screening purposes, 0.5 µg of plasmid DNA in TE buffer, 1.5 µL of the recommended 10 X restriction enzyme buffer, 1 unit/µg DNA of the restriction enzyme and autoclaved ddH₂O water to a final volume of 15 µL were mixed together. The sample was incubated at the appropriate temperature for 1-2 hours to ensure complete digestion. The reaction was stopped by heat

inactivation or by the addition of DNA Loading Dye containing SDS (0.05% bromophenol blue, 20% glycerol, 2% SDS).

Genomic DNA digestions were carried out in 50 μL samples using 4-5 μg of genomic DNA, 5 μL of the recommended 10 X restriction enzyme buffer, 1 unit/ μg of DNA and autoclaved ddH₂O water to a final volume of 50 μL . The sample was incubated at the appropriate temperature overnight to ensure complete digestion. The reaction was stopped by heat inactivation.

2.7 Agarose Gel Electrophoresis

Genomic and plasmid DNA were separated and visualized by agarose gel electrophoresis. Agarose gels were prepared by dissolving 1.0% (w/v) of agarose in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) and then heating in a microwave oven. The gel was allowed to cool to approximately 45°C before the addition of 0.5 $\mu\text{g/mL}$ ethidium bromide, the gel then was allowed to solidify in a gel tray with an appropriately sized gel comb inserted in the gel.

Upon solidification of the agarose, the gel comb was removed and the gel placed in the gel box with TAE Buffer covering the gel. DNA 6X loading dye was mixed with the DNA sample prior to loading into the wells. Ten μL of the DNA ladder High Ranger Plus (Norgen Biotek Corp.) was used to facilitate the estimation of the DNA fragment sizes. The DNA fragments were separated at 1-10 V/cm and were viewed and photographed using the Alpha Innotech Alphamager 2200 and AlphaEaseFC software.

2.8 Ligation of Restriction Enzyme Digested pDNA Fragments

Bacteriophage T4 DNA ligase was used to catalyze the ligation of DNA fragments. Reaction mixtures were set up according to the manufacturer's

guidelines (New England Biolabs). A typical reaction mixture contained 15 μL heat-inactivated restriction enzyme digestion mixture containing the smaller plasmid DNA fragment, 5 μL heat-inactivated restriction enzyme digestion mixture containing the larger plasmid DNA fragment, 4 μL 10X T4 DNA ligase reaction buffer, 1 μL T4 DNA ligase, 15 μL double distilled water.

The ligations of restriction enzyme digested plasmid DNA fragments were carried out at 16 °C overnight (15-16 hours). The reaction mixture was then used to transform *E.coli* DH5 α .

2.9 Shotgun Ligation Library

The plasmid pUC19Neo was digested with *EcoRI*, genomic DNA was digested with either *EcoRI* or *Tsp509I* (according to section 2.8). The digestion was stopped by phenol chloroform extraction (Sambrook *et al.*, 1989) followed by ethanol precipitation. DNA concentration was determined by absorbance at 260nm.

From the DNA concentration data, three different ligation mixtures were made with varying ratios of vector to insert (1:1, 1:5 & 1:10). Each individual reaction had a final volume of 20 μL which included 2 μL of 10X ligation buffer and 0.5 μL of T4 DNA Ligase (200 Units, New England Biolabs). The ligation reaction was performed at 16°C overnight, and was heat inactivated at 65°C for 20 minutes the following day. The entire ligation product was then used for transformation (refer to section 2.4) of bacterial cells. Following a 1 hour incubation at 37°C the cell culture was transferred to 30 mL of LB medium containing both ampicillan (50 $\mu\text{g/mL}$) and kanamycin (25 $\mu\text{g/mL}$).

2.10 Large Scale plasmid DNA Isolation

For single plasmid cultures, 1mL of frozen cell stock was inoculated into 30mL of LB medium containing appropriate antibiotics. For multi-plasmid cultures (shotgun library) the plasmids were already pre-inoculated into 30 mL of culture containing both ampicillin and kanamycin. The 30 mL was incubated for 4-6 hours or until an OD_{600} of 0.4 to 0.6 was reached. The 30 mL of culture was inoculated into 6L of LB medium containing appropriate antibiotics. The culture was then incubated at 37°C overnight (16-18 hours).

Cultures were harvested by centrifugation in 1 liter bottles at 4560 x g for 20 minutes in a Beckman model J-6B centrifuge equipped with a JS-4.2 rotor. The supernatant was decanted and the bacterial pellets were dried for 5 minutes. The cells were resuspended in 50 mL of resuspension solution (50mM Tris, 10 mM EDTA & 100 ug/mL RNase A) by gentle vortexing and incubated at room temperature for 10 minutes. An equal volume of lysis solution (1% SDS & 0.4 M NaOH) was added to the cells and the combined solutions were mixed by inverting the bottles 6-8 times for no more than 10 minutes, at which time 50 mL of neutralization solution (3M Sodium Acetate pH 4.75) was added to the 1 L bottles. The bottles were inverted repeatedly and vortexed for 5-10 seconds to ensure adequate mixing the solutions. The bottles were placed in the fridge at 4°C for 30 minutes or overnight to facilitate the formation of genomic DNA and protein aggregates. The contents of the 1 L bottles were transferred to 250 mL centrifugation bottles, and centrifuged at 30100 x g for 20 minutes in a Beckman J2-M1 centrifuge equipped with a JA-14 rotor. The lysate was filtered into fresh 1 L bottles using sterilized coffee filters placed in a funnel. For every 150 mL of

lysate (equivalent to 1 L of bacterial culture), 500 mL of ethanol was added and placed in a freezer (-20°C) for 1 hour or overnight to precipitate the plasmid DNA.

The plasmid DNA solution was transferred into 250 mL centrifuge bottles and precipitated at 30100 x g for 20 minutes using the Beckman J2-M1 centrifuge equipped with the JA-14 rotor. The plasmid DNA pellets were evaporated to dryness at room temperature for 10-15 minutes. The plasmid DNA was then resuspended in 6-7 mL of water or TE Buffer for every liter of bacterial culture harvested and placed in the freezer (-20°C) until plasmid verification was completed. An aliquot of plasmid DNA was obtained by a second ethanol precipitation and verified by agarose gel electrophoresis for the presence of the desired plasmid prior to cesium chloride (CsCl) banding.

The density of the resuspended plasmid DNA was adjusted to 1.57-1.59 g/mL using CsCl. Density was determined by solution weight. Two hundred μ L of EtBr (10 mg/mL) was added for every liter of bacterial culture harvested. (Note: Once the EtBr solution had been added, the plasmid DNA solution container was covered with aluminum foil at all times to prevent exposure to light.) Equal volumes of the plasmid DNA solution was transferred to optiseal ultracentrifuge tubes (1 tube for every liter of bacterial culture harvested). If necessary the volume of the tube was adjusted with TE buffer and CsCl (1.59 g/mL) to reach a final volume of approximately 8.9 mL. The tubes were placed in a type 70.1 Ti rotor which was subsequently placed in the Beckman L8-80M ultracentrifuge and centrifuged at 246 960 x g for 22 hours at 23°C.

After centrifugation, the lower supercoiled plasmid DNA band was removed from the tube using an 18.5 gauge needle and a 10 mL syringe and

transferred to a 15 mL tube. Extraction of the EtBr from the isolated supercoil plasmid DNA was performed by repeated mixing with CsCl saturated isoamyl alcohol. To facilitate the separation of the two phases, the emulsion was centrifuged for 2-3 minutes at 2000 x g in a Beckman GPR centrifuge equipped with a JS-42 rotor. The top isoamyl alcohol layer was subsequently removed using a sterile Pasteur pipette. This process was repeated until no EtBr could be seen in both the aqueous plasmid DNA solution and the isoamyl alcohol.

The plasmid DNA solution was transferred into boiled, clip-sealed Spectra/Por 7 dialysis tubing (Spectrum Laboratories, Inc.), and dialyzed against 30 liters (3 by 10 liters) of TE buffer. The buffer was changed twice at 16-24 hour intervals. After dialysis, the plasmid DNA was filter sterilized using a 0.2 µm syringe filter and aliquoted into sterile microfuge tubes for storage at -20°C.

2.11 Shot gun Library Analysis

Ligation efficiency was determined after large scale plasmid DNA isolation by CsCl banding to incorporate the changes which might have occurred during the large scale growth processes. The purified shotgun plasmid DNA (0.5 µg) was transformed back into *E.coli* along with a known concentration of pUC19neo. The transformation efficiency of the experiment was determined by the number of colonies obtained from the known concentration of pUC19neo. The percentage of insert containing plasmid in the library was determined using blue/white selection.

2.12 Determination of Nucleic Acid Concentration

2.12.1 Diphenylamine Assay

The assay consists of deoxyribose sugar hydrolysis using a strong acid followed by the binding of the diphenylamine compound to the open sugar ring resulting in the formation of a blue colour (reviewed in Thompson & Dvorak, 1989). A predetermined concentration of DNA was used as a standard, varying from 0 μg (water only) to 15 μg . Varying amounts of the sample DNA was also used for accurate determination of DNA concentration. The samples were placed in 1.5 mL microfuge tubes and the volume was equalized to 100 μL with autoclaved distilled water. One hundred μL of perchloric acid (20% v/v in ddH₂O) was added to each tube followed by the addition of 200 μL of diphenylamine solution (4% w/v in glacial acetic acid), and finally 20 μL of acetaldehyde (0.16% v/v in ddH₂O). The tubes were mix by inversion 6-8 times then centrifuged briefly in an Eppendorf 5402 desktop centrifuge to ensure no liquid remained on the side of the tube. The samples were then incubated at room temperature overnight (approximately 16 hours).

Sample absorbance was determined at 595 nm and 700 nm wavelengths of light in a Jenway 6405 Spectrophotometer. The water-only control (no DNA) was used as a blank. The absorbance (at 595 nm, 595-700 nm & 595+700 nm) from the standard DNA dilutions was plotted against the known concentration of DNA. A linear trend line was determined for each of the three graphs. The trend line which best fit the curve was used to determine the concentration of the unknown samples.

2.12.2 Direct Spectrophotometry

The concentration of genomic DNA and plasmid DNA was also determined by $A_{260} - A_{280}$ (Sambrook et al., 1989). Plasmid DNA was diluted in TE buffer and the absorption measured at 260 nm and 280 nm wavelength light using the Janway 6405 spectrometer. The formula,

$$[\text{pDNA}] = \frac{(A_{260})(\text{dilution factor})(50)}{1000}$$

was applied to determine the concentration of DNA in the sample. Only samples with an A_{260}/A_{280} of approximately 1.8 were considered to be accurate, and if this was not the case a diphenylamine assay was performed to accurately determine the DNA concentration.

2.13 Heat Inactivation of Enzymes

For some applications it was necessary to inactivate certain restriction enzymes prior to use in downstream applications such as ligations and transformations. Enzyme inactivation was carried out according to the manufacturer's instructions.

2.14 Klenow treatment of Digested pDNA

During the creation of the vector pUC19neo the formation of blunt ends was necessary to facilitate the cloning of the neo cassette into pUC19. The Klenow fragment of *E.coli* DNA polymerase I (New England Biolabs) was used to fill in the overhangs generated by cohesive end generating restriction enzymes. Typically 1 µg of digested plasmid DNA was incubated with 2.5 units of Klenow fragment in the presence of 200 µM dNTP and the appropriate volume of 10X

EcoPol reaction buffer for 15 minutes at room temperature. The reaction was stopped using phenol:chloroform: isoamyl alcohol extraction.

2.15 Excision of Digested pDNA Fragments from Agarose Gels

After the separation of plasmid DNA fragments by agarose gel electrophoresis, individual bands were characterized visually against a known marker. Fragments which had the correct, pre-determined size were excised from the gel using a sharp razor under UV light. The fragment was placed into a pre-weighed microfuge tube; the tube was weighed again to determine the weight of the gel slice. The plasmid DNA fragment was purified from the agarose gel using the Norgen Biotek Corp. Gel Extraction Kit.

2.16 DNA Sequencing & Analysis

DNA sequencing was performed by the Robarts Research Institute (London, ON), using an ABI GeneAmp 9600 or 9700 Thermocycler and an Applied Biosystems 3730 Analyzer. *In silico* manipulation of genomic DNA inserts and plasmids for cloning strategy was performed using the Vector NTI suit of programs (Invitrogen).

Nucleotide composition was determined using Annhyb (Friard, 2002) and the output data was analyzed using Excel (Microsoft Corp., USA).

2.17 Statistical Analysis

All Statistical analysis was performed using Excel (Microsoft Corp., USA) or SAS (Statistical Analysis Software).

2.18 Isolation of DNA from mammalian cells

2.18.1 Mammalian Cell Collection

Mammalian cells were grown to approximately 85-95% confluency in 100 mm tissue culture plates before isolation of plasmid DNA was performed. Citric Saline (135 mM potassium chloride & 15mM sodium citrate) was used to lift cells from the plate. The cells were washed with 6 mL of Citric Saline which was removed promptly by aspiration, followed by a second volume (2 mL) of Citric Saline was added and incubated for 5-10 minutes at 37°C. Upon visual conformation of cellular release from the plate, 6 mL of phosphate buffered saline (PBS, 140 mM NaCl, 2.6 mM KCl, 4 mM Na₂HPO₄ and 1.5 mM KH₂PO₄, pH 7.3) was added to the plate. The contents of the plate were agitated using gentle tituration to separate the cell aggregates, and then half the contents (4 mL) of the plate was placed in a 15 mL plastic tube (Sarstedt, Germany).

2.18.2 Plasmid DNA Isolation using the Modified Norgen Protocol

Mammalian cells were collected from 100 mm tissue culture plates according to the protocol above. The tubes were centrifuged at 670 x g in a Beckman Sprinchron centrifuge equipped with a GH 3.8 rotor for 5 minutes. The supernatant was decanted and the remaining contents of the tube were gently resuspended in 1 mL of PBS using a P1000 pipette with a cut tip so as not to shear the cells. One mL of the sample was transferred into a 2 mL microfuge tube (the remaining cells in the tube were used for cell counting). The microfuge tube was centrifuged for 5 minutes at 670 x g in the Beckman Spinchron centrifuge, and the supernatant was decanted.

The remained liquid was used to resuspend the cells by gentle flicking. Three hundred μl of Norgen Resuspension Solution was added to the cells and the cells were further resuspended. Three hundred μl of Norgen Lysis Solution was added and the samples were capped and inverted 3-5 times. The solution was incubated at room temperature for 5 minutes. Finally 900 μl of Neutralization Solution was added and the sample was capped and inverted for mixing. The samples were incubated at 4°C for 30 minutes to facilitate the precipitation of genomic DNA and proteins from the sample. The pellicle was separated from the lysate by centrifugation at $16000 \times g$ in an Eppendorf 5402 centrifuge at 4°C . The lysate was applied to the Norgen column in two 900 μl aliquots following centrifugation at $16000 \times g$ for 1 minute. The column was washed twice with 500 μl of wash buffer and a final spin was used to remove any residual wash buffer. Two 25 μl elutions were performed using the Norgen Elution Buffer. The elution of the plasmid was done at $2285 \times g$ to facilitate the total elution of all the plasmid from the column. A final $16000 \times g$ spin was used to make sure all liquid was removed from the column.

Six volumes of isopropanol were added to the sample to facilitate the precipitation of the plasmid DNA. The samples were placed at -20°C for no less than 30 minutes, at which point the plasmid DNA was precipitated by centrifugation at $16000 \times g$ in a 4°C Eppendorf 5402 centrifuge for 20 minutes. The isopropanol was removed by aspiration and the pellet was dried for 10 minutes at 37°C . The pellet was then resuspended in 20 μl of TE Buffer which was used for the transformation of bacterial cells.

2.18.3 Modified Hirt Protocol (Hirt, 1967)

Mammalian cells were collected from 100 mm tissue culture plates according to the protocol above. The tubes were then centrifuged at 670 x g in a Beckman Sprinchron centrifuge equipped with a GH 3.8 rotor for 5 minutes. The supernatant was decanted and the remaining contents of the tube were gently resuspended in 1 mL of PBS and transferred into a 2 mL microfuge tube (the remaining cells in the tube were used for cell counting). The microfuge tube was centrifuged for 5 minutes at 670 x g in a Beckman Spinchron centrifuge, and the supernatant was decanted. One hundred μ l of STE buffer (100 mM NaCl, 10 mM Tris-HCl & 1 mM EDTA, pH 7.3) was added to resuspend the cells by gently flicking. An equal volume of modified HIRT lysis buffer (0.6% SDS, 50 mM Tris-HCl, 10mM EDTA, 1 μ g/mL Proteinase K) was added and the solution was mixed by inversion 8-10 times.

The tubes were placed at 56°C for 1 hour to reduce DNase activity and allow the proteinase K to degrade a portion of the cellular proteins. One quarter volume of 5 M NaCl was added and the sample was incubated at 4°C overnight. The next day, the tubes were spun at 16000 x g in an Eppendorf 5402 centrifuge at 4°C for 45 minutes. The supernatant was removed and placed in a new microfuge tube. A Phenol:Choloroform:Isoamyl alcohol DNA extraction was then performed on the sample. The final DNA pellet was resuspended in 20 μ l of TE buffer which was used to transform bacterial cells.

2.19 Genomic DNA Isolation from HEK 293 cells

Genomic DNA was isolated from HEK 293 cells on a 150 mm tissue culture dish according to a modified genomic DNA isolation protocol (Sambrook

et al., 1989). Medium was aspirated from the plates, and the cells were washed with PBS. One mL of PBS was added to the plate followed by 1 mL of Pronase-SDS lysis buffer (0.01 M Tris-HCl, 0.01 M EDTA, 0.6% SDS, 1 mg/mL Pronase; pH 7.8) which was mixed by swirling. The plate was incubated at 37°C for 30 minutes at which time the cell lysate was collected into sterile 2 mL microfuge tubes and placed at 56°C overnight. The next day a phenol:chloroform:isoamyl alcohol extraction was performed on the samples (please refer to section 2.19). Upon the addition of 2 volumes of 95% ethanol, high molecular weight DNA appears in the tube. The high molecular weight DNA was removed from the tube using a sterile pipette tip and allowed to air dry for 5 minutes. The tip was then placed in 500 µl of TE buffer to dissolve the genomic DNA.

2.20 Phenol:Choloroform:Isoamyl Alcohol Extraction

An equal volume of 25:24:1 (v:v:v) phenol:choloroform:isoamyl alcohol (BioShop) capped with TE buffer (pH 8.5) and the solution containing the DNA of interest were mixed in a microfuge tube by vigorous vortexing until an emulsion had formed. The tube was centrifuged at 16000 x g for 5 minutes in an Eppendorf 5402 desktop centrifuge at room temperature. Being careful not to remove any of the white precipitate at the organic/aqueous phase boundary, only half of the organic phase was removed and placed in a fresh microfuge tube. The volume which was removed from the aqueous phase was replaced with the same volume of double distilled autoclaved water and the formation of an emulsion and centrifugation was repeated. Half of the volume of the aqueous phase was again removed and combined with the other half.

An equal volume of water saturated chloroform was added to the aqueous layer containing the DNA. The tube was vortexed until an emulsion was formed, and was centrifuged for 5 minutes. The aqueous layer was removed, and placed in a sterile microfuge tube. The concentration of sodium acetate was adjusted using 3M sodium acetate (pH 7.0) to a final concentration of 0.3M. Three volumes of cold ethanol were then added to the tube and the solution was stored at -20°C for one hour before centrifugation at 16000 x g for 10 minutes at 4°C. The supernatant was decanted and the pellet was dried at room temperature for 10 minutes. The pellet was resuspended in 70% ethanol and placed at -20°C for an additional hour before centrifugation at 16000 x g for 10 minutes at 4°C. The supernatant was decanted and the remaining liquid was aspirated, taking care not to disturb the pellet. The DNA pellet was dried at 37°C for 10 minutes before resuspension in an appropriate volume of TE buffer.

2.21 Mammalian Cell Culture Techniques

2.21.1 Human Embryonic Kidney Cell (293 cell) maintenance

Human Embryonic Kidney 293 cells (Microbix Inc., ATCC CRL-1573) were maintained as monolayer cultures at 37°C in a humid 5% CO₂ atmosphere in a water jacketed incubator (Fisher Scientific, Pittsburg PA). The medium utilized was autoclavable MEM (Invitrogen Corp., Gibco) supplemented with 0.225% (w/v) sodium bicarbonate (Invitrogen Corp., Gibco), 0.292 µg/mL L-Glutamine, 10% (v/v) heat inactivated fetal bovine serum (FBS) (CanSera International, Inc.), and was fortified with 1% Antibiotic-Antimycotic (10,000 units/mL penicillin G sodium, 25 µg/mL amphotericin B, and 10,000 units/mL streptomycin sulphate,

Invitrogen Corp., Gibco). All tissue culture dishes and tubes were obtained from Sarstedt.

All cell manipulations were carried out in a Nuaire biological safety cabinet, after 20 minutes of UV sterilization. Aseptic technique was used in all cell manipulations. Cells were observed using an inverted microscope (Dexter Instruments Co. Inc. San Antonio TX) with Olympus optics.

2.21.2 HEK 293 Cell Passaging

Nearly confluent 293 cell monolayers were passaged by using sterile saline citrate (15 mM sodium citrate, 135mM potassium chloride) to facilitate cell lifting from culture dishes. The volumes used were dependent upon the size of the dish used (ie. the smaller the dish the less saline citrate was used). The medium from a 150mm cell culture dish was aspirated from the plate, and the cells washed once with 6 mL of saline citrate. The wash was then aspirated and 4 mL of saline citrate was applied to the monolayer. The cells were then incubated at 37°C until cell rounding and lifting from the surface was observed. The cells were dislodged by gentle tapping of the culture dish. Eight mL of new medium was then added to the cell suspension and the cells were separated by gentle tituration. Six mL of the cell suspension was then added to a new culture dish, and the volume was adjusted with new medium to a final volume of 20 mL. The cells were then allowed to reattach overnight at 37°C.

2.21.3 Cell Freezing and Thawing

This procedure was modified slightly from the method outlined by Freshney (2000) for the freezing and storage of mammalian cells. Cells were grown to late log phase, lifted off the plate using saline citrate, and then

centrifuged at 670 x g for 5 minutes at room temperature (Beckman Spinchron Centrifuge). The medium was aspirated and the cells were resuspended in 4 mL of fetal bovine serum containing 10% dimethyl sulfoxide (DMSO). The cells were dispensed into pre-labeled 2 mL cryogenic vials (Nalgene). To freeze the cells slowly, the vials were rapped in paper towel and placed in a cardboard box at -20°C overnight, after which they were transferred to a -70 °C freezer.

To recover frozen cell stocks, cells in cryogenic vials were thawed by heating quickly in a 37°C water bath. Once thawed, the cells were poured into a 150 mm tissue culture plate containing medium supplemented with 10% serum and were swirled well to mix. As soon as the cells were attached to the plate, the medium was changed to remove the toxic DMSO.

2.21.4 Mammalian Cell counting

Cells were lifted from the tissue culture dish using saline citrate and PBS. The cells were transferred to a 15 mL plastic tube where the volume of the sample was determined. A 50 µl sample was removed from the tube and placed on the hemocytometer. The average number of cells was determined in the area defined by the hemocytometer, and the total number of cells in the sample could be calculated from the original volume of the sample.

2.21.5 Calcium Phosphate Transfection

The calcium phosphate procedure used to transfect the 293 cells in this study was based on the method developed by Graham and van der Eb (1973). Twenty-four hours prior to transfection, cells were subcultured and plated on a 100 mm plate to assure 70-80% confluency for transfection. Three hours prior to transfection, medium was removed and replaced with fresh medium. All buffers

and solutions were warmed to room temperature prior to transfection. For each 100 mm tissue culture plate 15 µg of plasmid DNA was mixed in TE buffer (pH 8.0) to a final volume of 470 µl, to which 55 µl of 2.5M CaCl₂ (filter sterilized) was added. For samples containing the genomic DNA libraries, 5 µg of each of the libraries (a total of 15 µg) were added and mixed in TE buffer to a final volume of 470 µl to which the CaCl₂ was then added. The DNA-CaCl₂ mixture was added dropwise to 525 µl of 2X HEPES (280 mM NaCl, 10 mM KCl, 1.5 mM Na₂HPO₄•2H₂O, 12 mM dextrose, 50 mM HEPES, pH 7.05, filter sterilized) with constant mixing. Once combined, the mixture was incubated at room temperature for 5 minutes to allow for the formation of calcium phosphate – DNA complexes. After the incubation period, the solution was briefly vortexed and added dropwise over the entire surface of the cells. The cells were incubated at 37 °C, and the medium was changed after 6-8 hours. Cells were then grown for 3-4 days, at which time half was used for the isolation of plasmid DNA while the other half was replated on the 100 mm tissue culture plate. The following day, the cells were then placed on selective medium.

2.21.6 Selection with G418

Cultured cells differ extensively in their sensitivity to Geneticin (G418, BioShop). The appropriate concentration to use for selecting transfected cells must be determined experimentally. Non-transfected 293 cells were subjected to varying levels of G418 during a 2-week period to determine the lowest concentration that would bring about total cell death in that period. It was determined that 400 µg/mL of G418 would bring about the formation of foci within a 2 week time period. Medium containing G418 was changed every 2-3 days

and selection was maintained for 2 weeks or until all the control cells were killed. The cells were then allowed to recover and grow in non-selective medium.

2.21.7 Histochemical Assay for lacZ Activity

Cells which were transfected with a plasmid carrying the *lacZ* reporter gene were stained for *lacZ* activity. Medium was aspirated from the cells, which were then washed with 1 mL of PBS. The cells were then overlaid with 1 mL of 4% paraformaldehyde (pH 7.2) and incubated at 37°C for 3 minutes. The paraformaldehyde was then aspirated and the cells were gently washed twice with 1 mL of PBS. The cells were overlaid with 2-3 mL of X-gal stain (35 mM $K_3Fe(CN)_6$, 35 mM $K_4Fe(CN)_6$, 1 mM $MgCl_2$, in PBS; pH 7.4, supplemented with 1 mg/mL of X-gal) and incubated at 37°C overnight. The following day, the number of blue cells was counted and transfection efficiency determined.

3.0 RESULTS

3.1 pUC19neo

3.1.1 Construction of pUC19neo

The pUC19neo plasmid was constructed by inserting the neomycin cassette from pSVneo (Figure 3.1B) into the pUC19 backbone (Figure 3.1A). The neomycin cassette contains the neomycin (*neo^R*) resistance gene which encodes the aminoglycoside-3'-phosphotransferase protein allowing for selection with G418 in mammalian cells and kanamycin in bacterial cells (Yaloop & Svendsen, 2001).

Typically the neomycin cassette can be removed using the restriction enzymes *Nde* I (5'...CA↓TATG...3'; New England Biolabs) and *Bam*HI (5'...GGATCC...3'; New England Biolabs). The resulting restriction fragments would result in a size difference of approximately 100 bp between the cassette (2915 bp) and the vector backbone (2814 bp) which is inadequate for the proper separation and subsequent gel extraction (Figure 3.1C) of the desired 2915 bp neomycin cassette. Therefore, pSV2-neo was first digested with the restriction enzyme *Eco*RI (5'...G↓AATTC...3'; Fermentas Inc.), then digested with the enzymes *Nde*I and *Bam*HI. This strategy resulted in fragment sizes of 2915 bp, 2062 bp & 750 bp which facilitated the excision and gel extraction of the upper band, corresponding to the neomycin cassette.

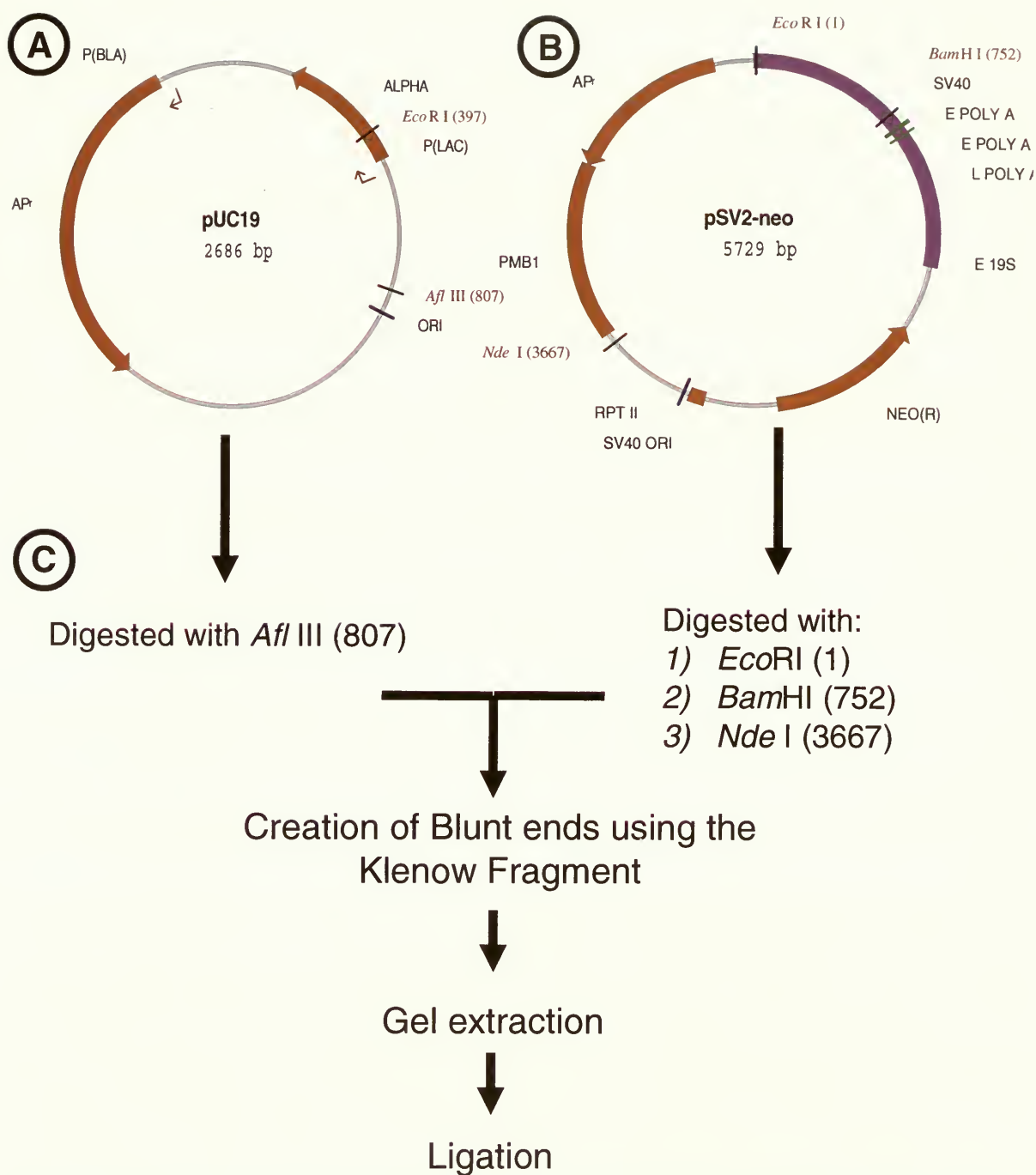


Figure 3.1: Methodology for the construction of pUC19neo. Restriction map of the plasmids pUC19 (A) and pSV2-neo (B) containing the relevant restriction enzyme sites. (C) Strategy for the construction of the final plasmid product.

3.1.2 Creation & Confirmation of the plasmid pUC19neo

The plasmid, pUC19 (Figure 3.1A) was digested with *Afl*III (5'...A*CRYGT...3'; New England Biolabs) which allowed for the insertion of the neo cassette into the pUC19 backbone. All fragments were treated with the large klenow fragment before separation by agarose gel electrophoresis. The Klenow fragment was used to fill in overhangs and create blunt ends (Figure 3.1C). Agarose gel electrophoresis was used for size separation and means for removal of contaminants which may inhibit the ligation process.

There are two possible outcomes from this ligation, a forward orientation of the neomycin cassette or a reverse orientation. Insert orientation was confirmed by restriction enzyme digestion (outlined in table 3.1) followed by analysis of the size of the linear DNA fragments (Figure 3.2 B). Dual selection (ampicillin and kanamycin) demonstrated that ligation efficiency was extremely low, with only one colony being produced from the reaction. This colony contained the neomycin cassette in the reverse orientation (Figure 3.2 A) and was used for subsequent experiments.

Table 3.1: Expected fragment sizes based on *in silico* analysis of the cloning reaction. Restriction enzyme digests and agarose gel electrophoresis were used to determine the orientation of the insert.

Restriction Enzyme	Forward Orientation (fragment sizes, bp)	Reverse Orientation (fragment sizes, bp)
<i>EcoRI</i>	5607	5607
<i>BglII</i>	5607	5607
<i>EcoRI</i> & <i>BglII</i>	3172, 2435	4301, 1306
<i>SmaI</i>	4189, 1418	3316, 2291
<i>HindIII</i>	2903, 2704	4672, 935
<i>BamHI</i>	5214, 393	3277, 2330

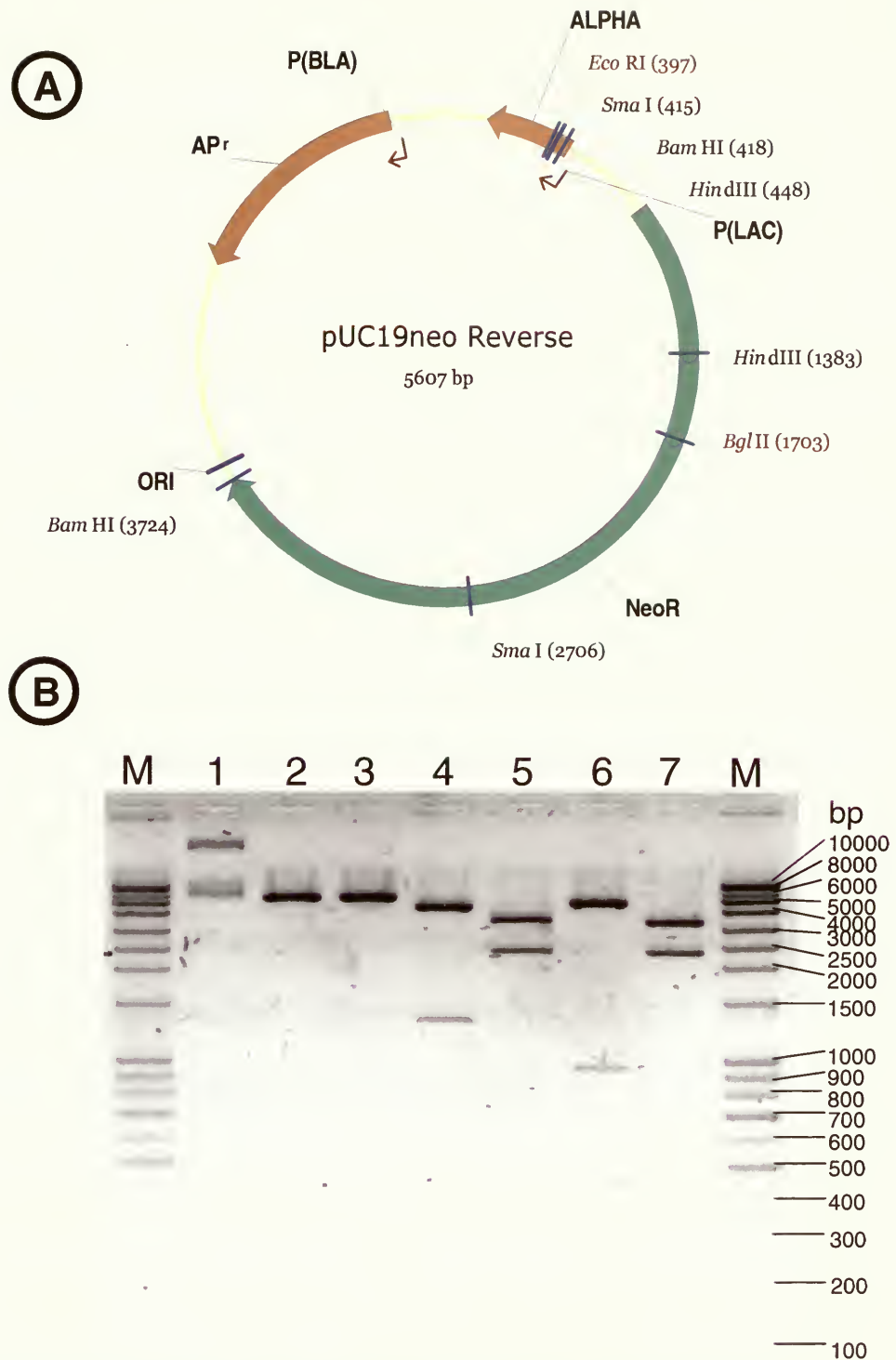


Figure 3.2: Confirmation of pUC19neo (Reverse). **(A)** The restriction map showing the relevant restriction enzymes used in the confirmation. **(B)** Agarose gel (1%) separating the restriction fragments of pUC19neo. The M denotes the Norgen HighRangerPlus marker, lane 1 is undigested plasmid DNA, lane 2-7 are the following restriction enzyme digestions: *EcoRI*; *BglII*; *EcoRI* & *BglII*; *SmaI*; *HindIII* & *BamHI*.

3.2 Genomic DNA Library

3.2.1 Construction of Genomic DNA Library

Genomic DNA libraries have traditionally been used for the sequencing of genomes, and require that individual clones be obtained and sequenced. Due to the insert size limitations of plasmid DNA most libraries have been based on other vectors such as the lambda phage or BAC (reviewed in Lodish *et al.*, 2000). This project however did not require the individual selection of clones (Figure 3.3), and therefore the size limitation of Petri dishes only became applicable when determining the percent of the culture population which contained an insert (Tables 3.2 A&B). This advantage allows for the efficient screening of the entire human genome assuming that insertion of the genomic fragments is completely random. Therefore to increase the variability and the relative transformation efficiency of the shotgun libraries, the ratio of vector to insert was manipulated so that 3 different ratios (1:1, 1:5, 1:10) were used in each of the six individual genomic DNA libraries (3Mix.1, 3Mix.2, 3Mix.3, 3Lig509.1, 3Lig509.2 & 3Lig509.3) created for this project.

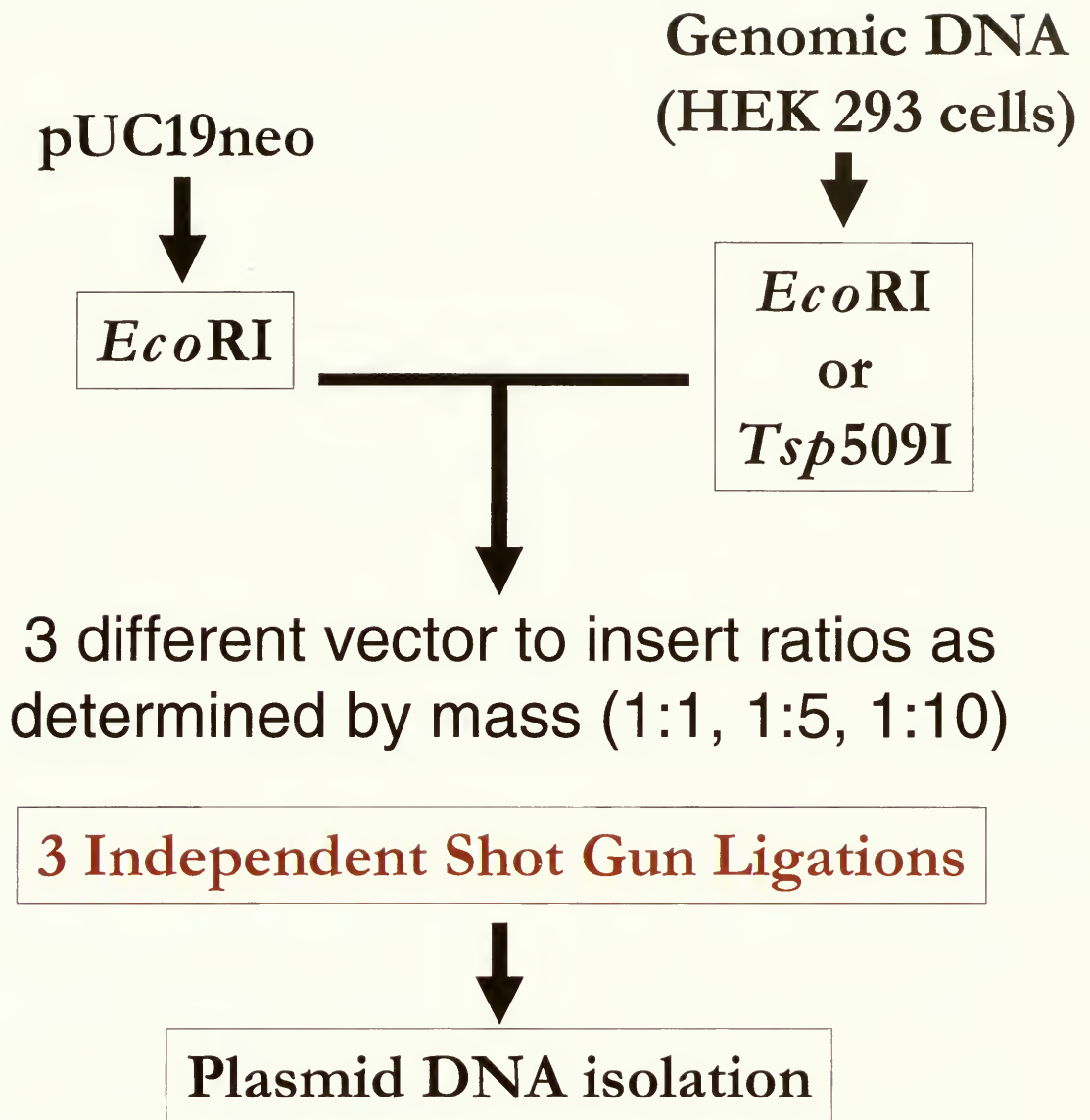


Figure 3.3: Strategy for the production of genomic DNA libraries using pUC19neo as the cloning vector.

Table 3.2 A: Average percentage of colonies containing a recombinant plasmid due to the insertion of *EcoRI* digested genomic DNA. The three libraries (3Mix.1, 3Mix.2 & 3Mix.3) were mixed equally by mass in later experiments to facilitate its transfection into HEK 293 cells. The combined libraries are referred to as 3Mix. The percentage of white colonies is an average of the 3 different ratios (1:1, 1:5, 1:10) used to create the genomic DNA library.

Library Name	Transformation Efficiency (CFU/ug)	White Colonies (%)
3Mix.1	5.90E+06	30.4
3Mix.2	6.70E+06	39.1
3Mix.3	3.80E+07	8.7

Table 3.2 B: Average percentage of colonies containing a recombinant plasmid due to the insertion of *Tsp509I* digested genomic DNA. The three libraries (3Lig509.1, 3Lig509.2 & 3Lig509.3) were mixed equally by mass in later experiments to facilitate its transfection into HEK 293 cells. The combined libraries are referred to as 3Lig509. The percentage of white colonies is an average of the 3 different ratios (1:1, 1:5, 1:10) used to create the genomic DNA library.

Library Name	Transformation Efficiency (CFU/ug)	White Colonies (%)
3Lig509.1	5.90E+06	41.5
3Lig509.2	6.70E+06	24.1
3Lig509.3	3.80E+07	52.7

3.2.2 Analysis of the Genomic DNA libraries

When digesting large sequences, the average size of the fragments is dictated by the restriction enzyme recognition sequence. For example, assuming a GC content of 50%, a six cutter such as *EcoRI* will cut on average every 4096 bp (4^6 bp) where as a 4 cutter such as *Tsp509I* will cut on average every 256 bp (4^4 bp). Given this information, an estimation of the probability of having any given DNA sequence in the library can be calculated from the equation (Blaber, 1998)

$$N = \frac{\ln(1-P)}{\ln(1-f)}$$

Where P is the desired probability, f is the fractional proportion of the genome in a single recombinant and N is the necessary number of recombinants. Using *EcoRI* as the restriction enzyme, 3.37×10^6 individual clones would be needed to effectively screen 99% of the human genome, where as when using *Tsp509I*, 5.40×10^7 clones would be needed.

The total number of recombinants within the genomic DNA libraries can theoretically be determined from the percent of white colonies and the transformation efficiency. Using these numbers in comparison to the calculated number of clones needed to effectively screen 99% of the human genome, a percentage of the total genome represented by the shotgun ligations can be determined. The *EcoRI* (table 3.3A) digested genomic DNA library showed an average coverage of 51.4%, where as the *Tsp509I* (table 3.3B) covered approximately 5.37% of the human genome.

Genome equivalents (GEs) describe the complexity (number of independent DNA clones) of a genomic DNA library, and can be used (as another way) to determine the quality of the library (Wendl *et al.*, 2001). During the ligation process, multiple copies of the human genome are present in the mixture. For example a total of 6.4 ug of genomic DNA was used in the ligations with varying ratios of vector to insert; this mass corresponded to 7.98×10^4 human genomes. Due to random chance, duplications of certain sequences can occur and therefore the GEs of the libraries must be considerably greater than 1 in order to have a high probability of including all the sequences of a genome.

The shotgun libraries used in this project had genomic equivalencies (Table 3.4) of 7.09 for the 3Mix (*EcoRI*) series of libraries, and 0.73 for the 3Lig (*Tsp509I*) series of libraries. Though the genomic equivalency of the 3Lig library is not as high as expected, when combined, the two sets of genomic DNA libraries cover a large portion of the human genome (7.82 GEs). This is exemplified by analysis of the insert sizes shown in figure 3.4 which revealed that there was great variability between individual clones, with insert sizes ranging from 200bp to 10000 bp or more. Insert size variability was also illustrated in figure 3.5 as seen by a smear in the background of the intense vector fragments. This background smear is very similar to the smear seen in the lanes containing digested genomic DNA.

Table 3.3 A: Percentage of the total number of individual recombinants required for 99% probability of having a sequence represented in the human genomic DNA library using the restriction enzyme *EcoRI*.

Library	# of recombinant plasmids	Percentage of the Required Number of Recombinants
3Mix.1	5.38E+06	53.22%
3Mix.2	7.86E+06	77.74%
3Mix.3	9.92E+06	22.72%

Table 3.3 B: Percentage of the total number of individual recombinants required for 99% probability of having a sequence represented in the human genomic DNA library using the restriction enzyme *Tsp509I*.

Library	# of recombinant plasmids	Percentage of the Required Number of Recombinants
3Mix.1	7.35E+06	4.53%
3Mix.2	4.84E+06	2.99%
3Mix.3	6.01E+07	8.59%

Table 3.4: Number of GEs in the constructed genomic DNA libraries.

Library	1 GE	Total number of independent clones	Number of genomic equivalents
3Mix (<i>EcoRI</i>)	7.3×10^5	5.18×10^6	7.09
3Lig (<i>Tsp509I</i>)	1.2×10^7	8.70×10^6	0.73

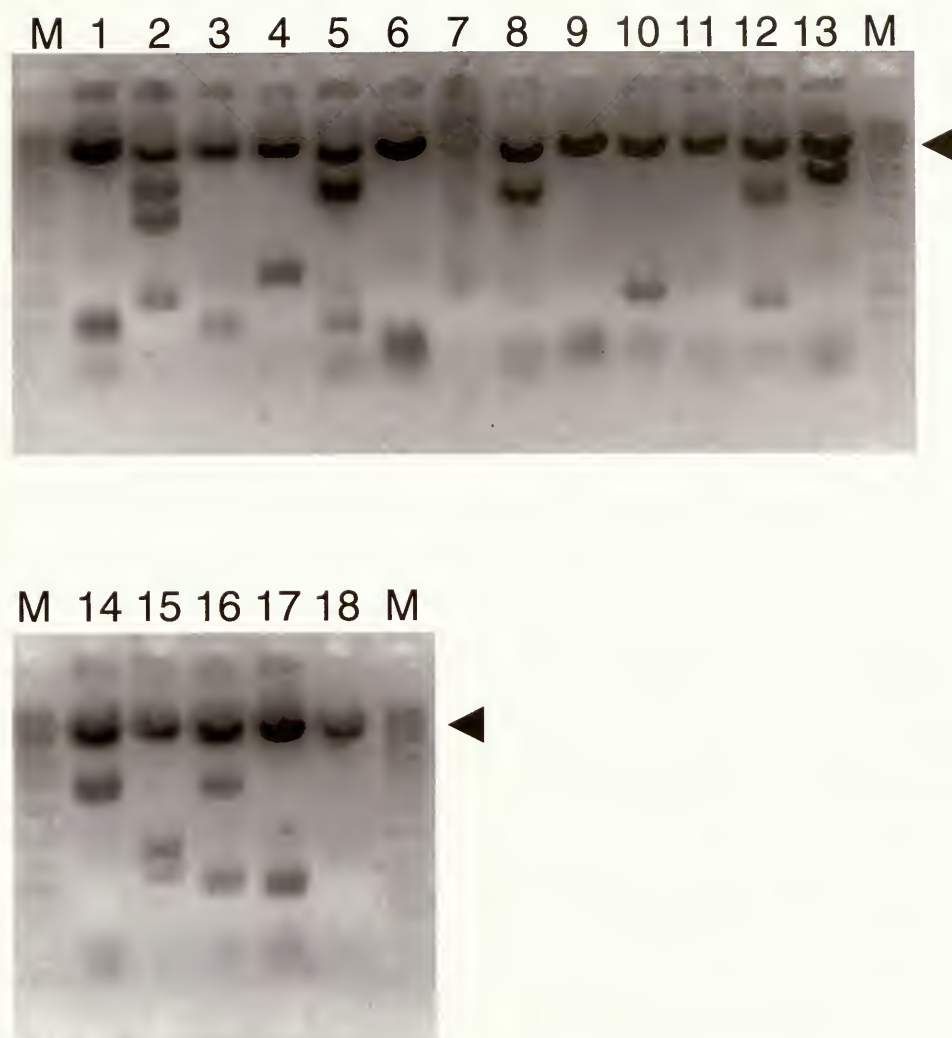


Figure 3.4: Library variability using insert size analysis. White colonies were obtained before the large scale grow-up and screened using the restriction enzyme *EcoRI*. The fragment pattern was then compared to others in an attempt to find similar sized fragments. The letter M indicates the Norgen HighRangerPlus marker, lanes 1-17 are the screened white colonies with the (◄) symbol representing the position of the pUC19neo vector. Lane 18 is the control pUC19neo digestion.

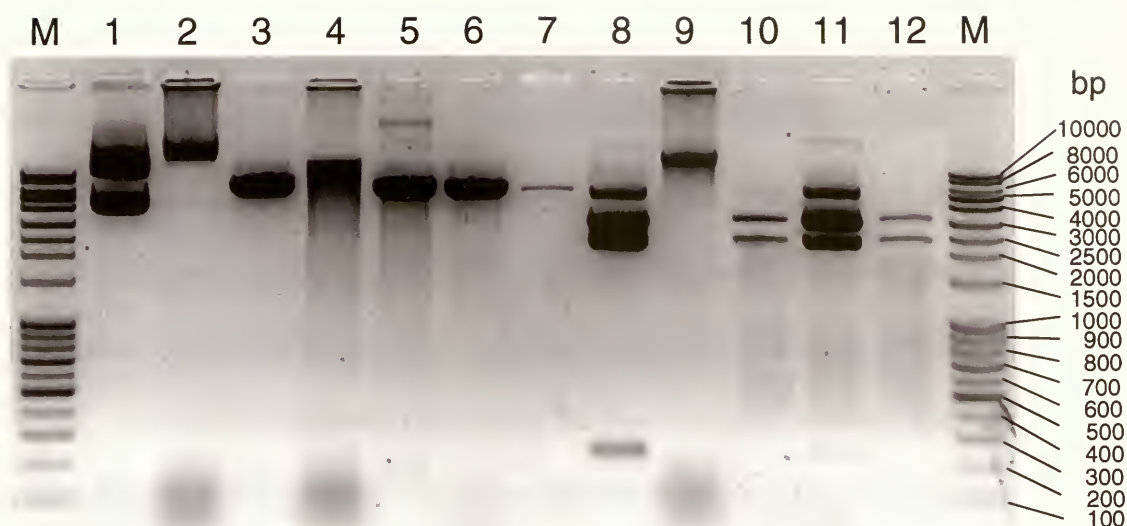


Figure 3.5: Restriction enzyme analysis of genomic DNA libraries after the large scale propagation and isolation. Lanes 5-7 contain the 3Mix library series (3Mix.1-3 respectively) and were digested with *EcoRI* to release the genomic DNA insert. Vector (pUC19neo) and genomic DNA (lanes 3 & 4 respectively) was also digested with *EcoRI*. Lanes 10-12 contain the 3Lig library series (3Lig.1-3 respectively) and were digested with *NdeI* and *SmaI* to release the genomic DNA insert. Vector and genomic DNA (lanes 8 & 9 respectively) was also digested with *NdeI* & *SmaI*. Vector and genomic DNA (lanes 1 & 2 respectively) loaded as undigested controls. The marker (lane M) was HighRangerPlus (Norgen Biotek Corp.).

3.3: Transfection and Extraction Procedure

3.3.1: *Transfection efficiency and variability of HEK 293 cells.*

Transfections using calcium phosphate are inherently variable due to small changes in parameters such as pH or temperature which greatly affect the formation properties of the DNA calcium phosphate complex (Jordan *et al.*, 1996). pH can be controlled by using the same stock solutions throughout the entire length of the project. However factors such as ambient temperature and cellular condition are not as easily controlled and therefore will attribute to the variability seen in calcium phosphate transfections. Shown in figure 3.6, HEK cells were transfected on different days with the same amount of pCMV- β and subsequently stained with X-gal. During the first transfection, 1.04% of the cells were able to take up pCMV- β and express the β -galactosidase gene; the second transfection had a much greater efficiency with 19.30%. The difference between the two days being 18.26% represents a substantial and significant difference between the two transfection events (according to the Student's t-test $p = 5.46 \times 10^{-7}$, 4). Samples of the same day and from the same reaction showed only small variability.

Sample to sample variability was examined using equal amounts of pCMV- β transfected on the same day (Figure 3.7). The first sample was able to transfect an average of 33929 cells, the second sample contained the same amount of DNA but was only able to transfect an average of 19704 cells. This was a significant decrease, according to the students t-test ($p=0.003$, 4), in the number of cells transfected. Such variability can be beneficial in the randomization of plasmids which are taken up by the cells. Unfortunately, this

variability is also detrimental, especially during the analysis of the longitudinal study of plasmid maintenance.

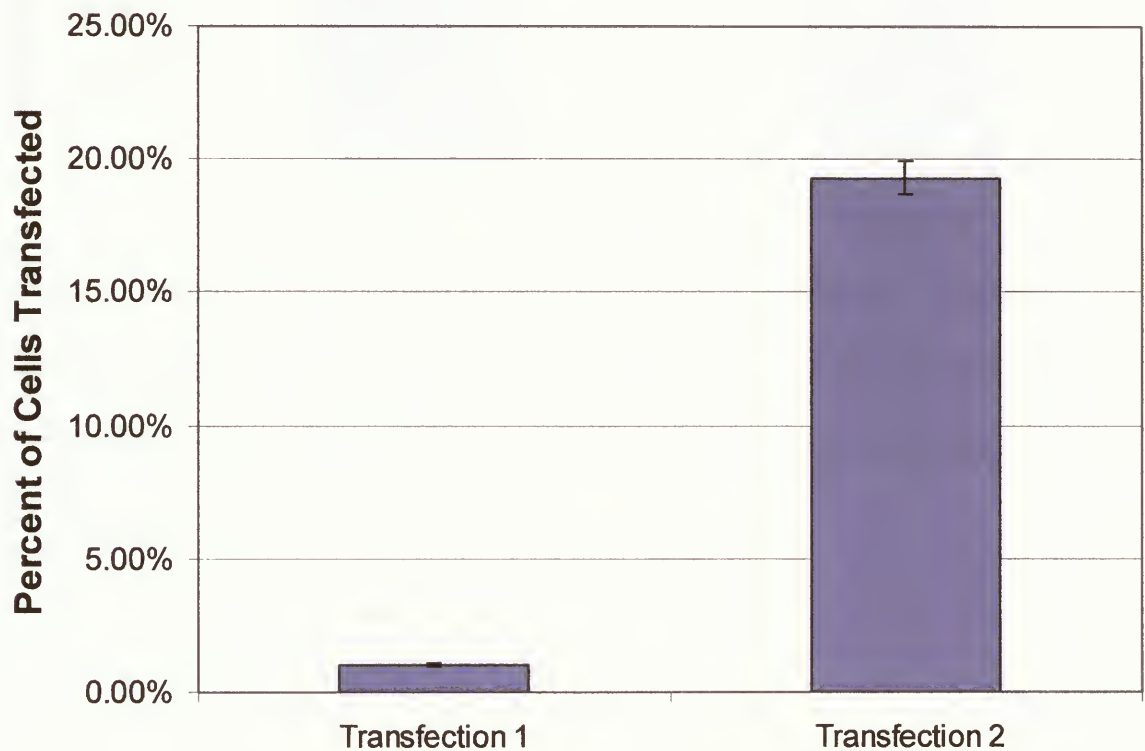


Figure 3.6: Variability seen between transfections on different days. Difference between two samples are significant according to the Student's t-test ($p=5.46 \times 10^{-7}$, 4).

ability is also detrimental, especially during the analysis of the literature study of plant material.



Figure 3.6: Variability seen between transistors on different days. The difference between two samples was significant according to the Student's t -test ($p < 0.05$).

10.4

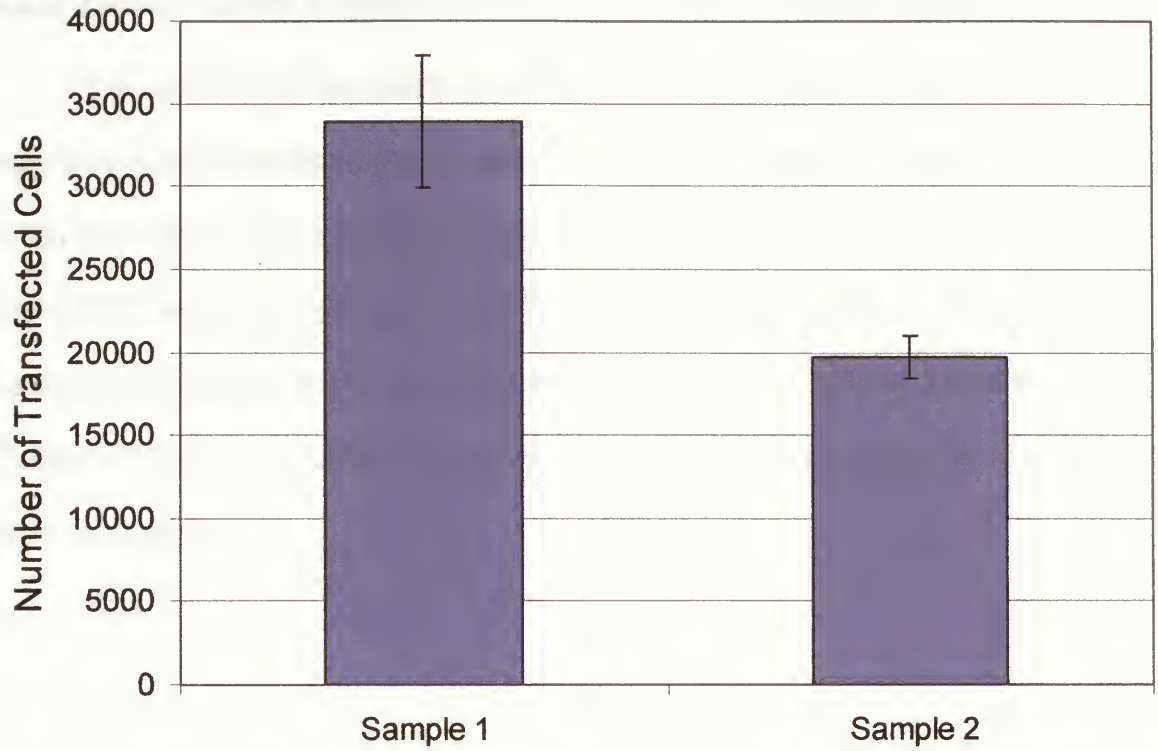


Figure 3.7: Transfection variability between samples. Transfection of sample 1 & 2 were performed on the same day into the same population of cells and with the same amount of plasmid; however the samples were prepared separately.

($p=0.003$, 4)



Figure 3. T-helper cell counts in two groups. The Y-axis represents the number of T-helper cells. The X-axis represents the two groups. Error bars represent standard deviation. $p < 0.001$.

3.3.2: Determination of optimal G418 concentration during selection

The optimal concentration of G418 present in the medium had to be experimentally determined due to variable sensitivity within cell lines. Therefore cells which did not contain any plasmid were subjected to varying amounts of G418 (200, 400, 600, 800 & 1000 $\mu\text{g/mL}$). Visualization of the cells over the 2 week selection process revealed that 400 $\mu\text{g/mL}$ was the optimal concentration (Table 3.5) for the complete removal of cells which did not contain the neomycin resistance gene.

Table 3.5: Results of varying concentration of G418 on 293 cells during a two-week period (percentages are estimates of the number of surviving cells, where 100% is fully confluent).

G418 Concentration ($\mu\text{g/mL}$)	t = 0 weeks	t = 1 week	t = 1.5 weeks	t = 2 weeks
200	100%	100%	75%	45%
400	100%	70%	30%	0%
600	100%	50%	25%	0%
800	100%	25%	5%	0%
1000	100%	15%	0%	0%
0 (Control)	100%	100%	100%	100%

3.3.3: Isolation of Low molecular weight DNA from mammalian cells.

The Hirt method (Hirt, 1967) is the current method for the isolation of low molecular weight DNA from mammalian cells and is time consuming and cumbersome. The Birnboim and Doly method (1979) is much quicker, however its use in extracting low molecular weight DNA from mammalian cells has not been extensively explored. This method is coupled to a DNA concentrating step usually accomplished by ethanol precipitation. This step can be replaced with the use of a resin based chromatography column. The various methods available at the time were explored and the quality of the sample was determined by transformation of bacterial cells. It should be noted that the volumes used with the Norgen Isolation kit, with respect to the three buffers (resuspension, lysis & neutralization) were increased 3-fold to accommodate for the excessive amount of biological material seen in mammalian cells. The results show (Figure 3.8) that the Hirt extraction method was significantly better at extracting plasmid DNA from mammalian cells, followed by the modified Norgen protocol, which was significantly better than both the Qiagen and the original Birnboim and Doly protocol ($p=0.0001$, 8, 27.54).

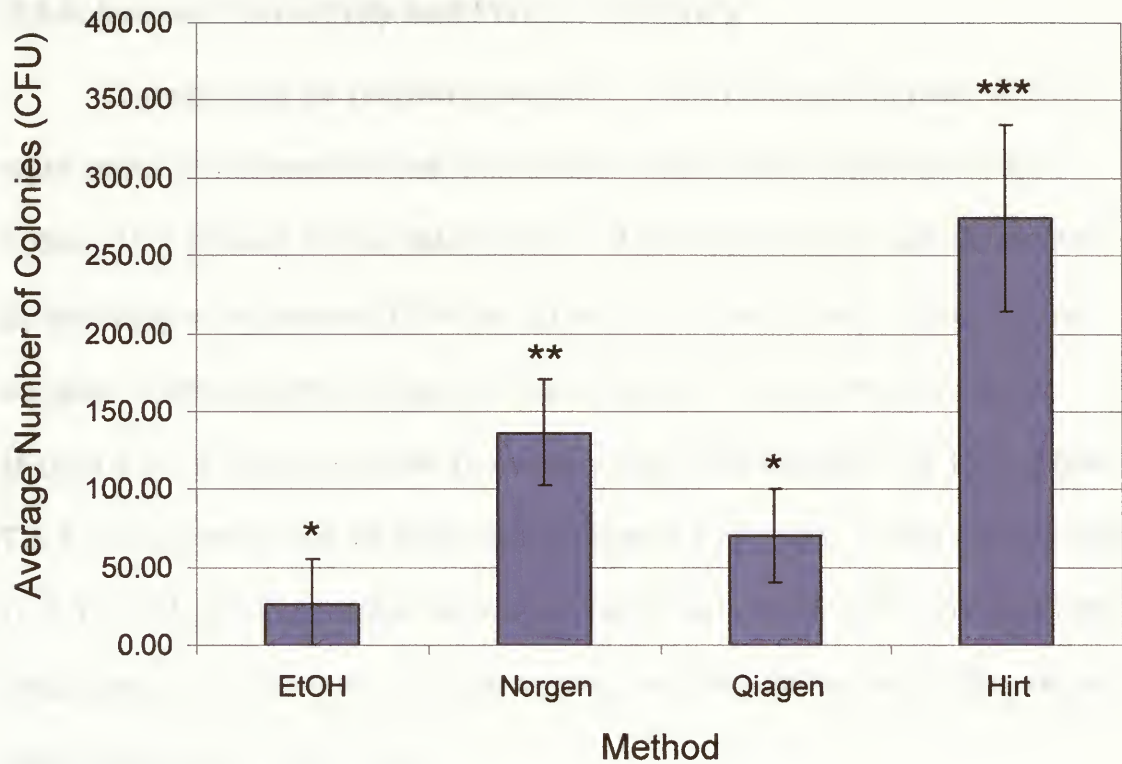


Figure 3.8: Isolation of plasmid DNA (pUC19neo) from an equal number of transfected mammalian cells. The amount of DNA isolated from the Norgen method is significantly greater than the amount isolated using the Qiagen method according to a one-way ANOVA with a subsequent Scheffe post-hoc comparison ($p=0.0001$, 8, 27.54). Stars indicate significant differences.

3.3.4: Isolation Sensitivity and Percent Recovery

To determine the recovery sensitivity, known amounts of plasmid DNA were added to mammalian cells immediately before the addition of the lysis buffer. This allowed for the determination of the approximate plasmid recovery by detecting intact plasmid DNA using bacterial transformation from a known sample. In the picogram range the dose response is not completely linear (Figure 3.9). A sharp increase in recovery was seen between $5 \times 10^{-4} \mu\text{g}$ and $7.5 \times 10^{-4} \mu\text{g}$ which can be attributed to plasmid saturation. In this plateau region ($> 7.5 \times 10^{-4} \mu\text{g}$), the percent recovery is not as accurately calculated as in the linear range. However, the average percent recovery for the entire data set was determined to be $7.65 \pm 2.62\%$.

The lower linear range between $0 \mu\text{g}$ and $5 \times 10^{-4} \mu\text{g}$ is unsaturated and allows a more accurate description of the system. During the time course experiments, the number of colonies per plate did not exceed 150 colonies, reinforcing the importance of this particular region of the graph. The average percent recovery of the linear range was determined to be $5.77 \pm 1.12\%$. Theoretical determination of the lower detection limit was obtained using the slope of the line and letting $y=1$ (1 colony obtained). The resulting theoretical lower detection limit was calculated to be $3.43 \times 10^{-4} \mu\text{g}$. This limit corresponds to a sensitivity of 0.566 plasmids per cell in 1 million cells.

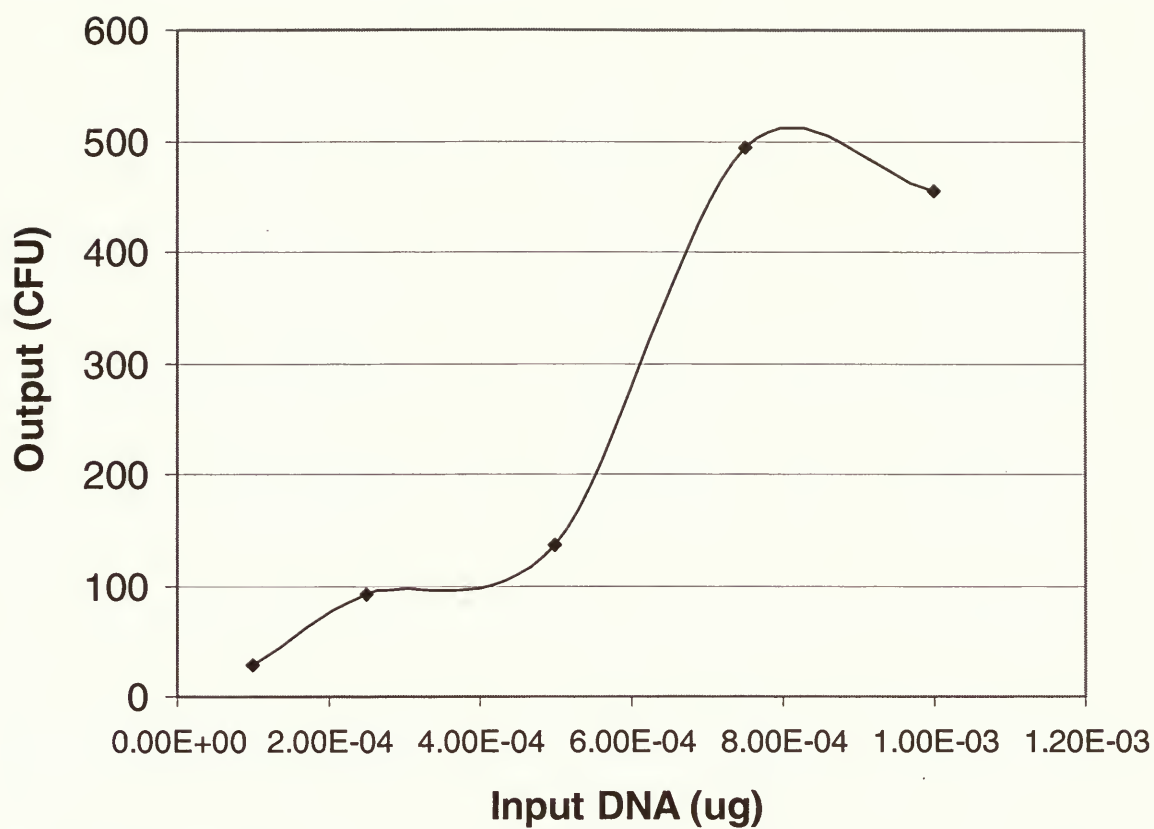


Figure 3.9: Isolation of known amounts of plasmid DNA from mammalian cells.

The dose response to each particular input amount was calculated by the number of colony forming units (CFU) (n=3).

3.3.5: Analysis of the Extraction Assay

The extraction assay is comprised of two parts, lysis & transformation. By using known amounts of input DNA (1.0×10^{-4} μ g) preliminary determination of plasmid loss can assist in improving the sensitivity of the assay (Figure 3.10). Plasmids of various sizes, pUC19 (2686 bp), pUC19neo (NEO, 5607 bp) and pCMV-beta (7164 bp) were also studied to determine the affect of size on the various stages of the extraction assay. The majority of the plasmid DNA is lost during the lysis step, with the greatest loss being observed in larger plasmids.

When adding pDNA with elution from columns containing just cellular lysate, an increase was observed in relationship to the size of the plasmid. However, the process of transformation itself is not affected by plasmid size.

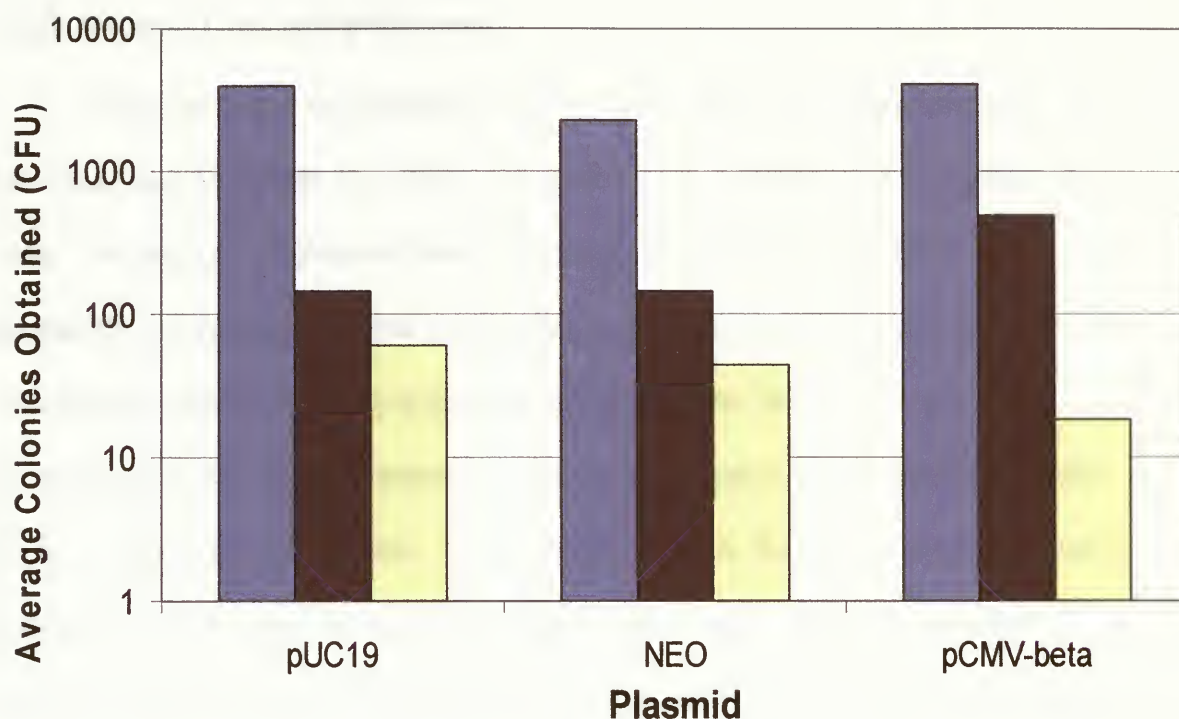


Figure 3.10: Average number of colonies ($n=3$) obtained when plasmid DNA (1.0×10^{-4} μg) is injected into the procedure at three different time points. The plasmids pUC19 (2686 bp), pUC19neo (NEO, 5607 bp) and pCMV-beta (7164 bp) were used to identify bias related to the size of the plasmid. The standard (■) was determined from the transformation of the plasmid DNA in just TE buffer. The “Before Transformation” sample (■) was obtained by lysis of cells without plasmid but was spiked just before the transformation step. Finally to mimic the actual extraction of plasmid from cells, plasmid DNA was added just before the lysis step of the protocol (■).

3.3.6: Second Concentration step

The goal of the mammalian plasmid DNA extraction assay employed in this study was to screen the entire population cells isolated for intact plasmid DNA. The amount of plasmid detected was determined from plasmid DNA standards which accompanied each experiment (the plasmid amounts for all subsequent experiments were determined in a similar fashion). A second concentration step was necessary to reduce the volume of the sample to under 20 μ L for use in transformation. Either three volumes of ethanol or six volumes of isopropanol were used for concentration with both resulting in the isolation of similar amounts of plasmid DNA (Figure 3.11). Using non-concentrated samples, we were able to detect significantly more plasmid DNA than in the concentrated samples. This statement however was only true when larger quantities of the plasmid were present (ie. Days 1 & 4). By day 8 & 12 much of the original plasmid would have been removed and therefore the concentration of the plasmid in the system would be greatly reduced. This low concentration of plasmid DNA was undetectable by the transformation assay when non-concentrated sample was used. Concentration of the sample by either isopropanol or ethanol allowed for the detection of plasmid DNA on day 12.

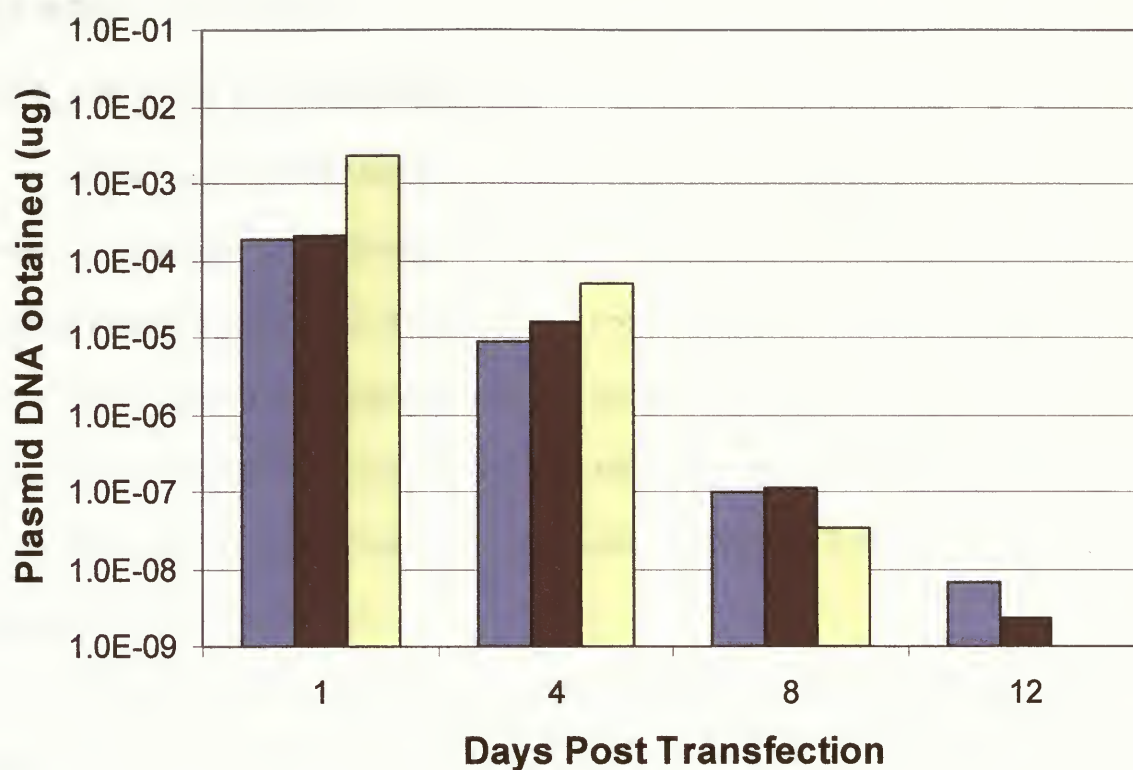


Figure 3.11: Average output of post column preparation of plasmid DNA for transformation (n=3). Plasmid DNA was extracted from the same number of cells transfected with pUC19neo. Samples isolated from the column were either not concentrated (yellow) a second time or concentrated with ethanol (blue) or isopropanol (dark blue).

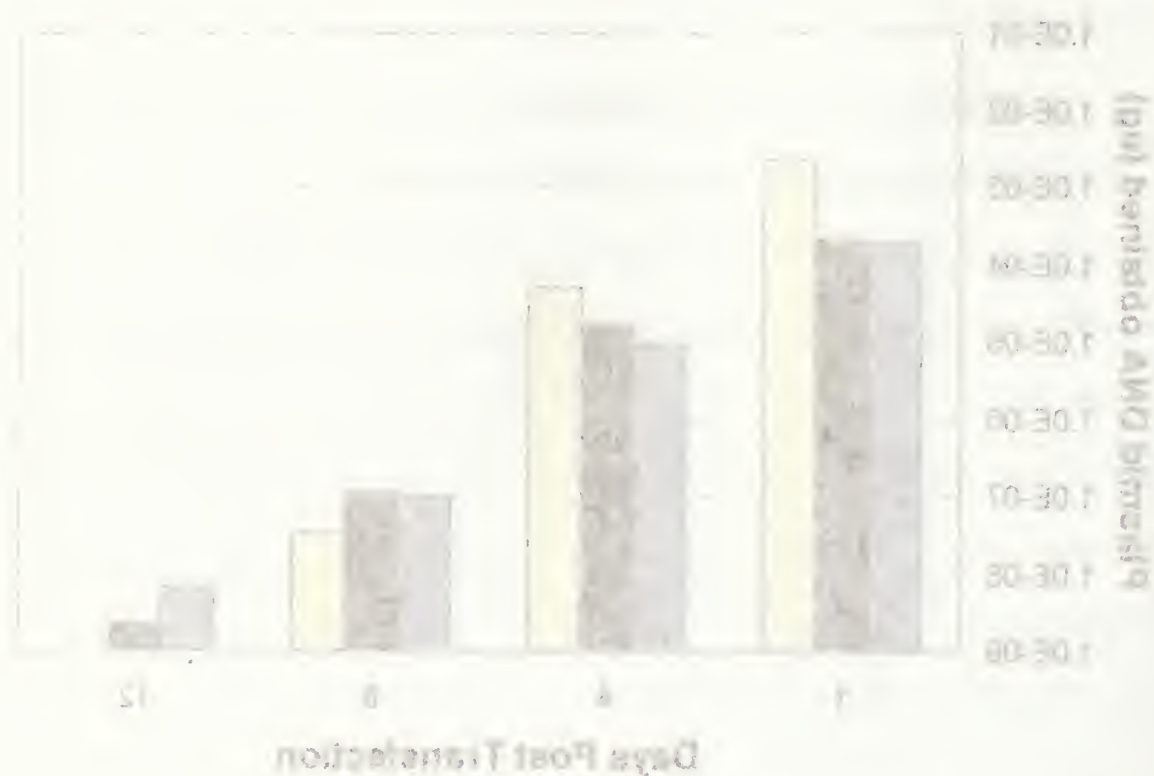


Figure 2.11 Average output of post column preparation of plasmid DNA for transfection (1-2). Plasmid DNA was extracted from the same number of cells transfected with pUC19. Samples removed from the column were either not concentrated () or concentrated with ethanol () at a second time at concentration ()

3.4 Isolation Analysis

3.4.1. *Indicators of Plasmid DNA Maintenance*

Eight days after transfection no plasmid DNA was detected in cells that were not subjected to G418 selection (Figure 3.12). This downward trend is also evident in previous sections (Figure 3.11). From these two results, a 12 day interval was chosen to be adequate for the removal of background plasmids. This interval was employed during the long term selection assay which resulted in choosing day 32 as the final extraction point. Day 32 is 13 days after the cessation of G418 selection.

The genomic DNA libraries contained plasmids which produce both blue (re-ligated pUC19neo) and white (insert containing) colonies. The blue colonies, in the libraries, can be used as an indicator for the presence of background plasmid. From the longitudinal experiment (section 3.4.2) the number of blue colonies present decreases over time between days 24 to 32 (Figure 3.13). This is similar to the decrease observed in both figures 3.11 & 3.12. Also noteworthy is the complete lack of blue colonies observed in the 3Lig509 sample on day 32. It will be shown in the next section that this same sample also had the largest number of white colonies on that day.

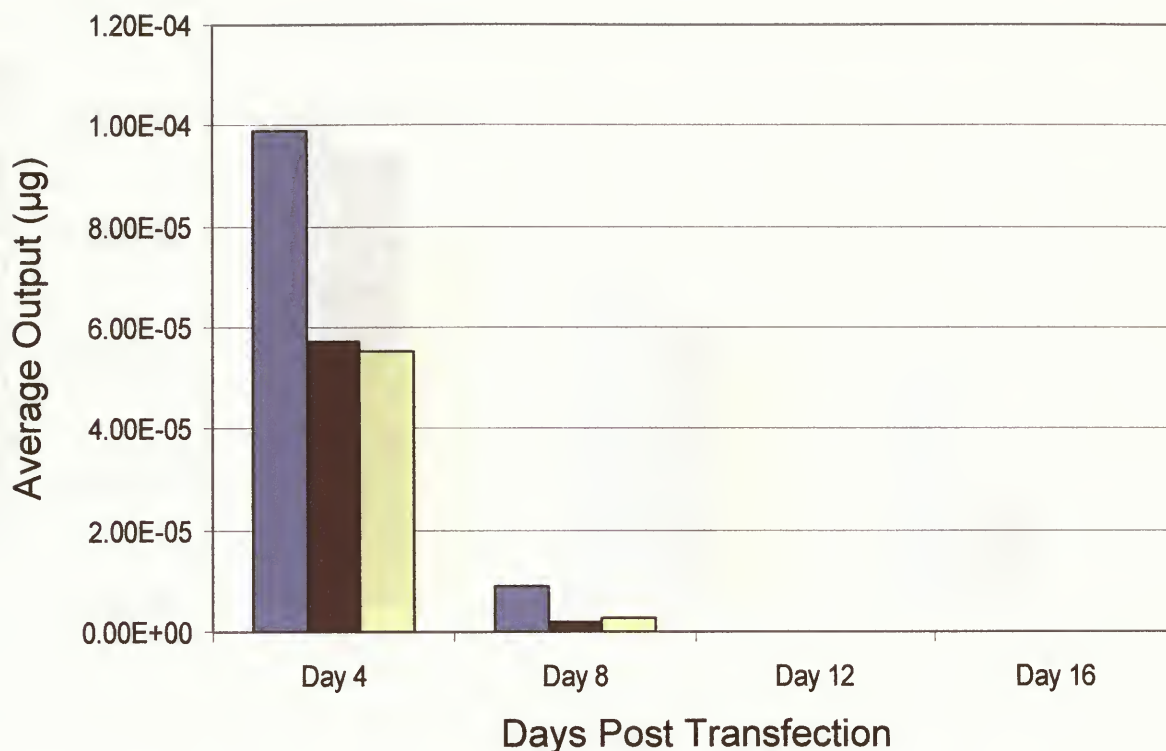


Figure 3.12: Non-selective maintenance of plasmid DNA in mammalian cells.

During this experiment pUC19neo (■), the 3Mix (■) and 3Lig509 (■) genomic DNA libraries were maintained in mammalian cells without selection. The value for pUC19neo represent the average (n=3) output of plasmid DNA as determined by the number of blue colonies obtained in a blue/white screen (no white colonies were obtained from these particular samples). The values from the 3Mix and 3Lig509 samples were determined from the number of white colonies obtained in a blue/white screen.



Figure 3.12. Non-selective measurement of plasma DNA in newborn cells. During the experiment 3LgG2 (grey), the 3LgG1 (black) and 3LgG2 + 3LgG1 (white) DNA samples were measured in newborn cells without selection. The values for 3LgG2 represent the average log output of plasma DNA as determined by the number of blue colonies obtained in a three-fold serial dilution and were obtained from three separate samples. The values for the 3LgG1 and 3LgG2 + 3LgG1 samples were determined from the number of blue colonies obtained in a three-fold serial dilution. The values are the average of three separate samples.

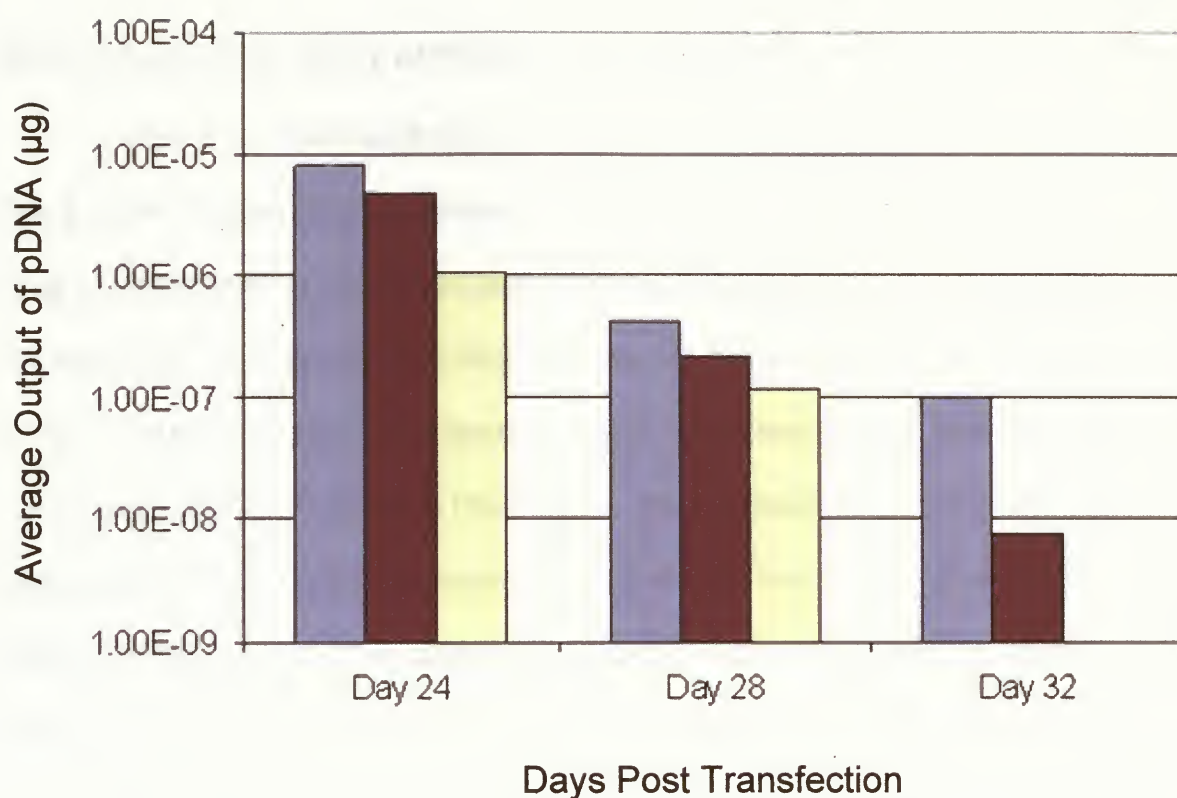


Figure 3.13: Average ($n=4$) output of the control plasmid (pUC19neo) during the longitudinal study (section 3.4.2, with G418 selection). The plasmid DNA was isolated from mammalian cells which were transfected with the 3Mix (■) and 3Lig509 (■) genomic DNA libraries and the control plasmid pUC19neo (■). No blue colonies were observed in the 3lig509 sample isolated on Day 32.

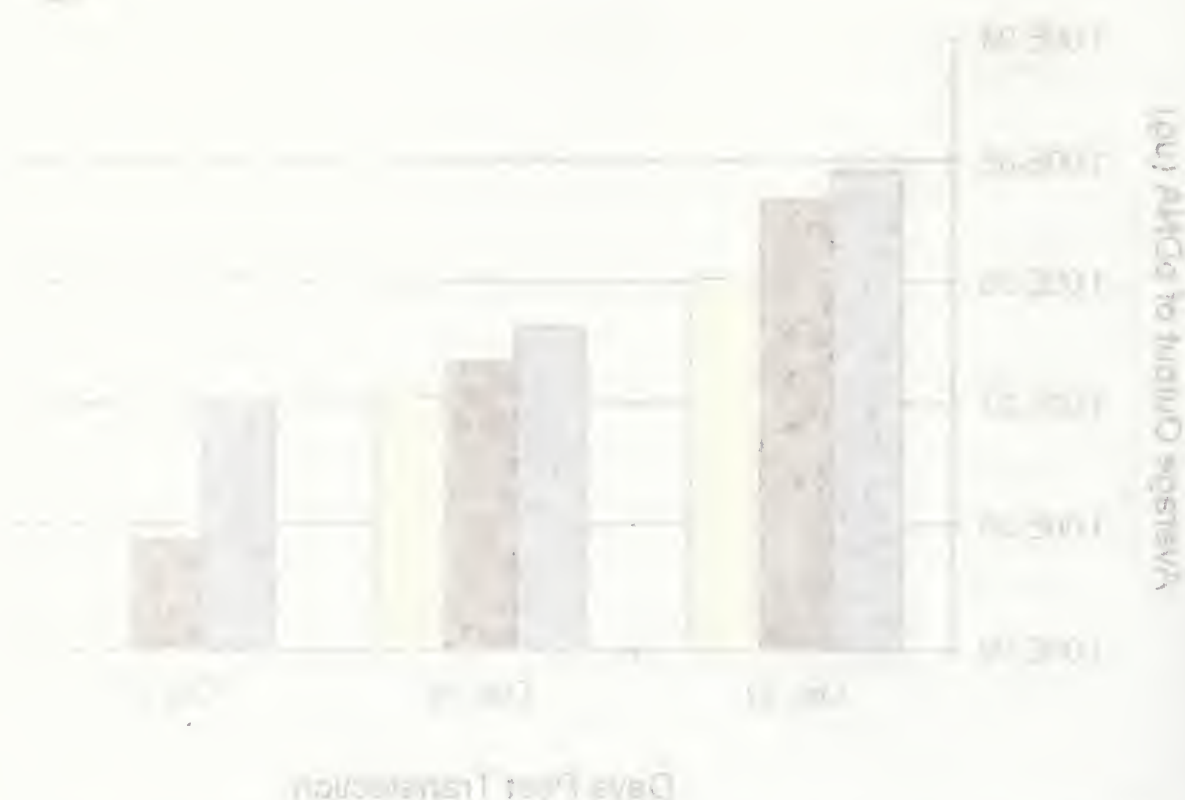


Figure 3.13: Relative expression of the control gene (GAPDH) in the control group (GAPDH) and the experimental group (GAPDH) at Day 1, Day 2, and Day 3. The control group shows a steady increase in expression over time, while the experimental group shows a sharp increase at Day 3. The control group shows a steady increase in expression over time, while the experimental group shows a sharp increase at Day 3. The control group shows a steady increase in expression over time, while the experimental group shows a sharp increase at Day 3.

3.4.2: Longitudinal Study of Plasmid DNA Stability.

Figure 3.12 indicates that no maintenance of any plasmid occurred after day 8 since no plasmid was isolated. However, plasmid DNA was detected up until 32 days post transfection in cells which were subjected to G418 selection (Figure 3.14). During selection very little plasmid is lost, however, upon removal of the selective pressure pUC19neo is readily lost between days 19 & 32. Cells containing pDNA with genomic DNA inserts show a reduction in the amount of plasmid DNA until day 28. Between days 28 and 32 the number of white colonies produced from the samples indicate an increasing amount of insert containing plasmids.

On Day 32 the amount of plasmid DNA isolated from the pUC19neo sample is less than the 3Mix and 3Lig509 sample (Figure 3.15). Using the parametric statistical analysis ANOVA we found that the differences between the samples were not statistically significant NS($p=0.50$, $df=9$, $F=0.743$). In an attempt to determine if the variability between replicates can be attributed to random variation we used a two factor ANOVA to firstly determine if the variability between replicates was significant and secondly to determine the variability between treatments. Using the Statistical Analysis System (SAS, www.sas.com) we found that there was significant variability between replicates ($p=0.0421$, $df=3$, $F=5.18$), however once this was accounted for we determined that there was no significant difference between treatments NS($p=0.2477$, $df=2$, $F=1.78$).

Parametric analysis relies on the assumption that the differences between the samples are normally distributed. Due to the high number of zeros present in

the data, the probability of a skewed distribution is quite high. The Mann-Whitney U-test was employed as the non-parametric equivalent to the Student's t-test. Analysis of differences between pUC19neo and 3Lig509 indicate a significant difference between the two samples ($p = 0.05$, $U=1$).

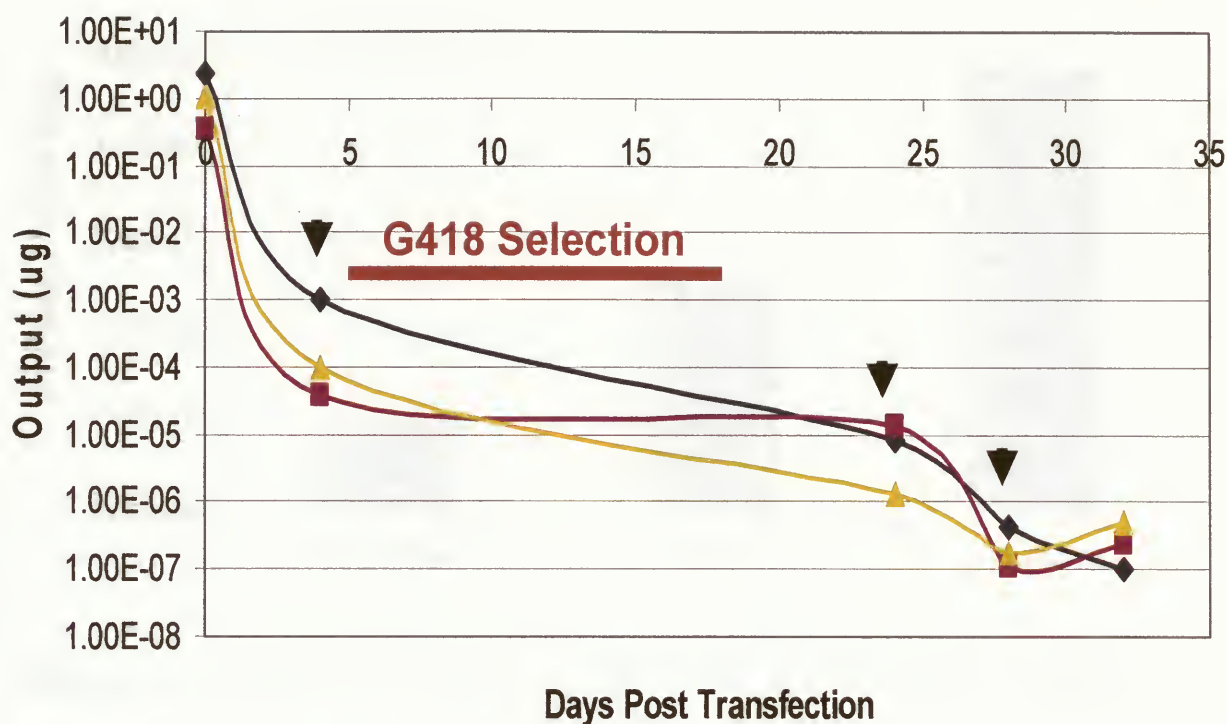


Figure 3.14: Long term output of biologically active plasmid DNA from Mammalian cells. Cells were passaged (▼) during this month long assay on days 4, 24 & 28. Selection of neomycin resistant cells with **G418** (400 $\mu\text{g/mL}$) occurred during a two week period between the days 5 & 19 post transfection. The outputs of both 3Mix (—■—) and 3Lig509 (—▲—) were determined from only the white colonies obtained. The output for pUC19neo (—◆—) was determined from only the blue colonies obtained. The data points represent an average of 4 independent assays.

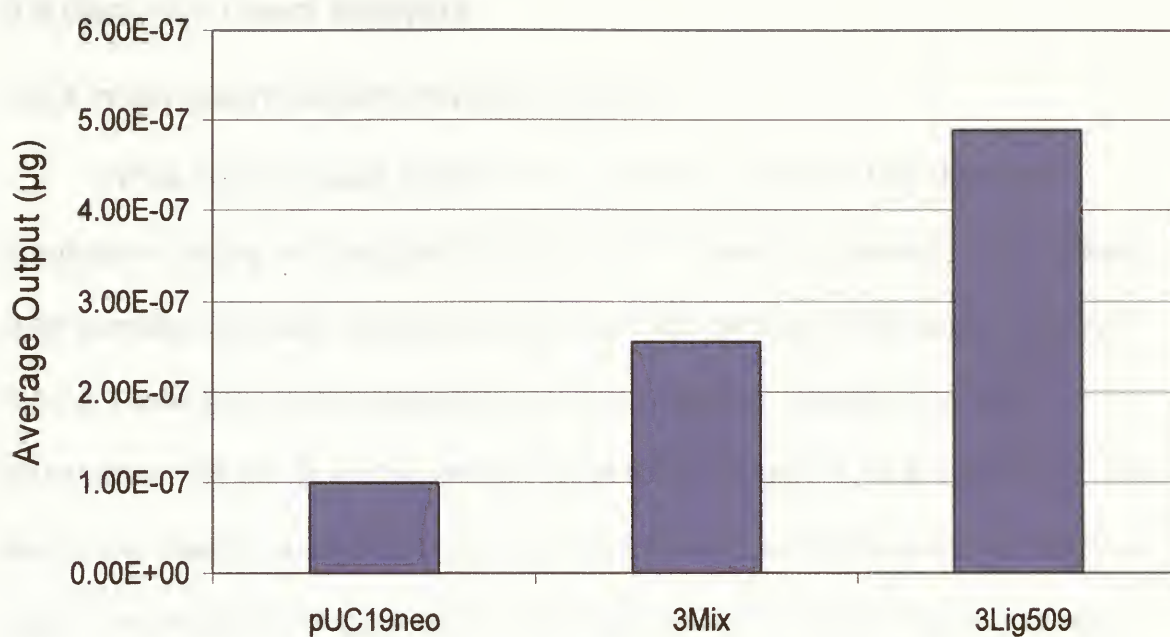


Figure 3.15: Average (n=4) amount of plasmid DNA isolated from each sample on Day 32.



Figure 3.12: Amount of glucose (mg) released from each sample over time (hours).

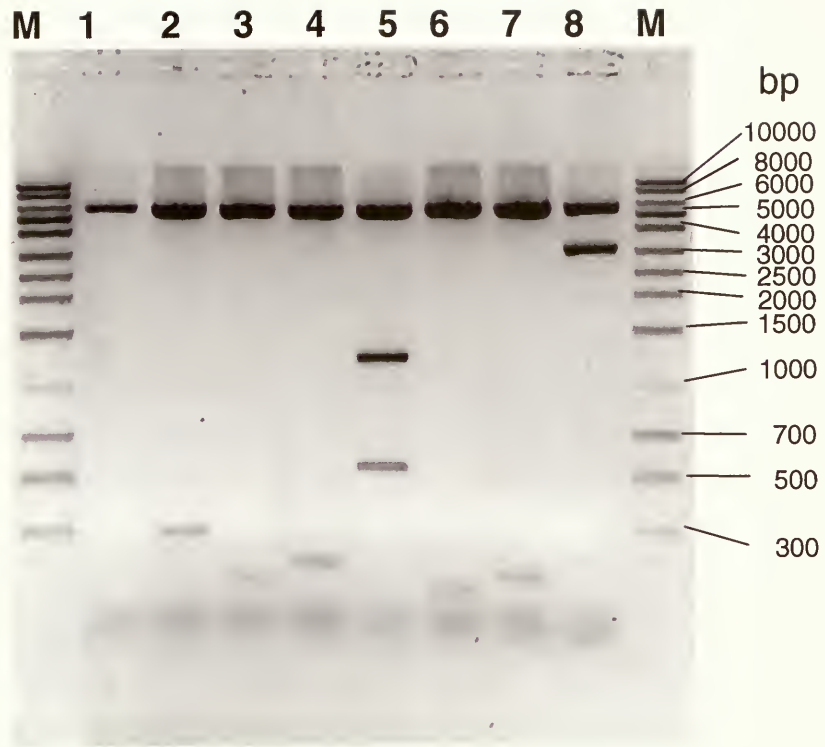
on Day 35

3.5 Genomic Insert Analysis

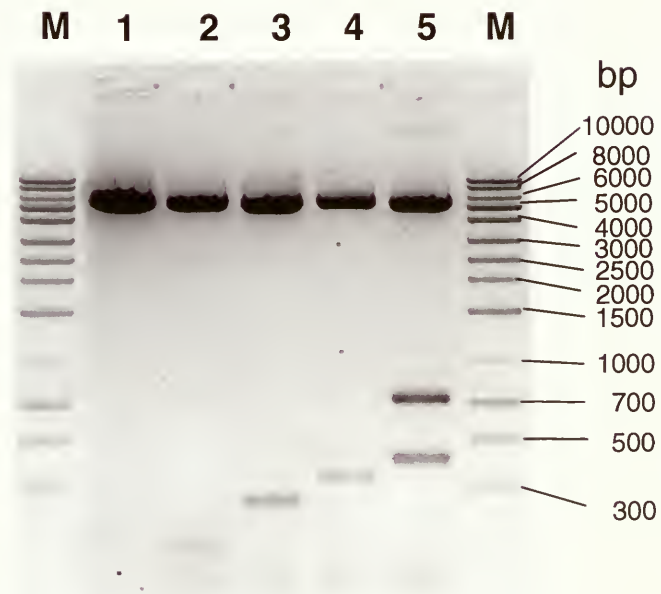
3.5.1 Insert determination and Size analysis.

White colonies were picked from samples isolated on Day 32 of the longitudinal assay and analyzed (Figure 3.15). In total 21 plasmids were isolated from samples originally transfected with the 3Mix genomic DNA library (Figure 3.16 & Table 3.6) and 24 plasmids were isolated from samples originally transfected with the 3Lig509 genomic DNA library (Figure 3.17 & Table 3.7). The size of the inserts ranged from less than 100 bp to over 3000 bp with an average insert size of 900 bp (Figure 3.18). The majority of the inserts had an average size of approximately 200-400 bp.

(A)



(B)



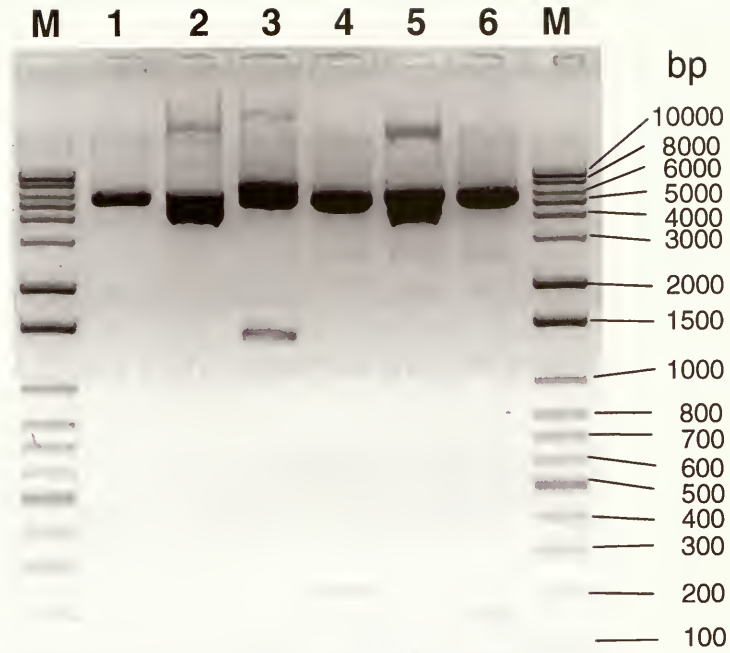
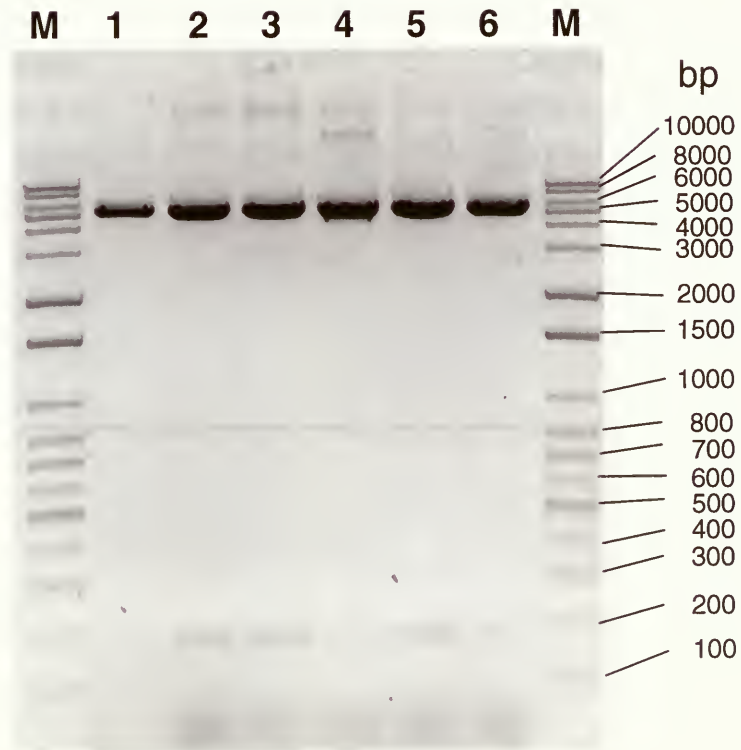
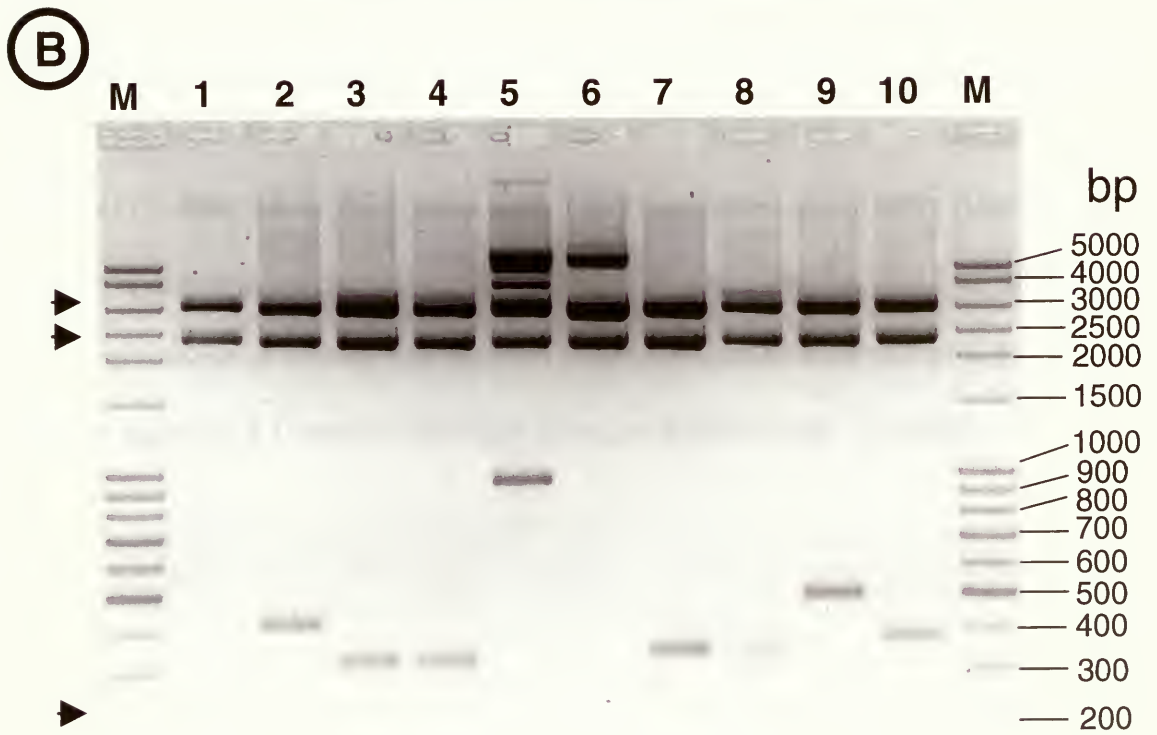
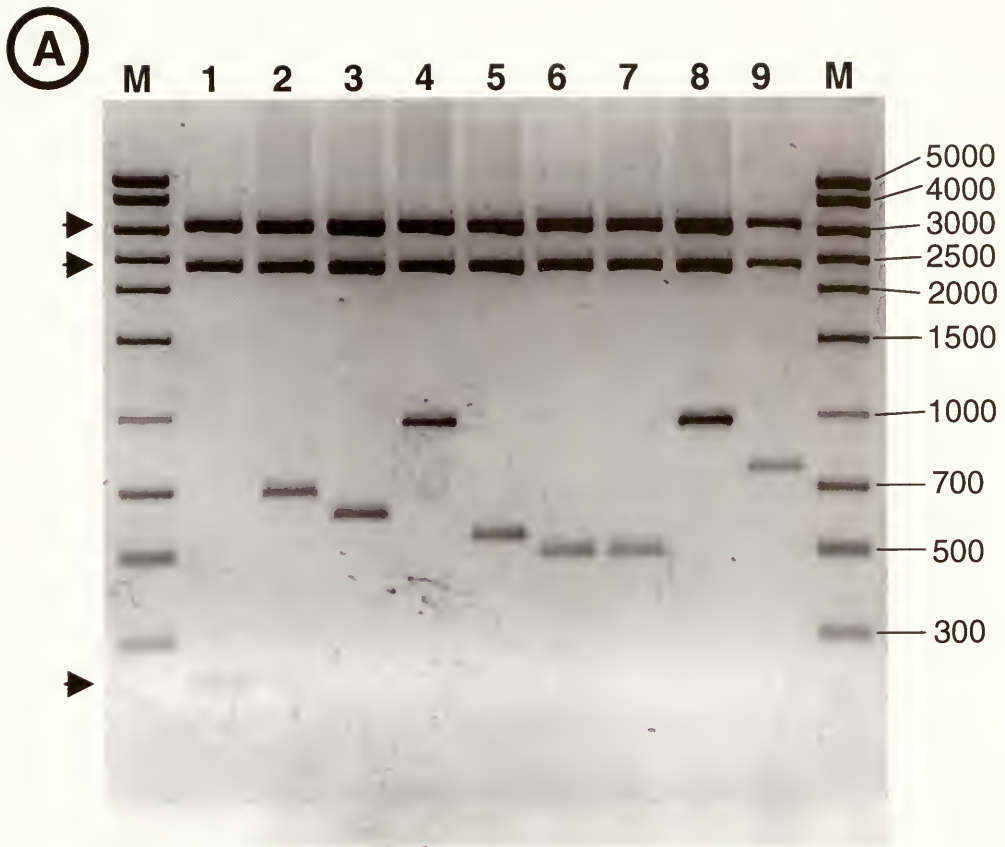
©**©**

Figure 3.16: Analysis of 21 plasmids isolated from day 32 mammalian cells which were transfected with the 3Mix (*EcoRI*) genomic DNA library mix. The plasmids were digested with the restriction enzyme *EcoRI* and given names as outlined in table 3.6. In all figures lane 1 is the control plasmid pUC19neo, the reference marker (**M**) used in figures A & B was HighRanger and the marker used in figure C & D was HighRangerPlus (Norgen Biotek Corp., Canada).

Table 3.6: Figure legend for 3.17 A, B, C & D with respect to the lane number and plasmid name.

Lane Number	Figure Reference			
	Figure 3.17 A	Figure 3.17 B	Figure 3.17 C	Figure 3.17 D
1	Control	Control	Control	Control
2	pE059-5	pE059-24	pE059-31	pE067-9
3	pE059-6	pE059-25	pE059-32	pE067-10
4	pE059-7	pE059-26	pE059-33	pE067-11
5	pE059-8	pE059-27	pE059-34	pE067-12
6	pE059-9		pE059-35	pE067-13
7	pE059-10			
8	pE059-11			



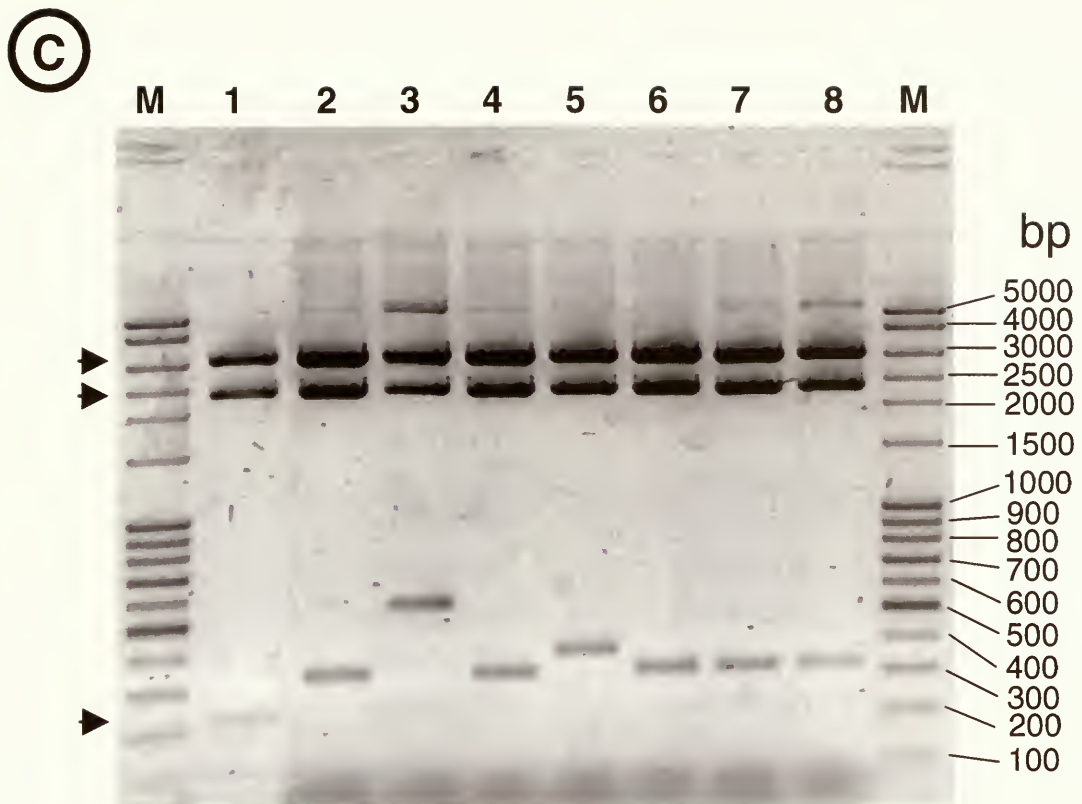


Figure 3.17: Analysis of 24 plasmids isolated from day 32 mammalian cells originating from the 3Lig509 (*Tsp509I*) genomic DNA library mix. The plasmids were digested with the restriction enzymes *NdeI* and *SmaI* and given names as outlined in table 3.7. In all figures lane 1 is the plasmid pUC19neo (Control), digestion of this plasmid results in 3 bands (▶) at 3086 bp, 2291 bp and 230 bp. The reference marker (M) used in figures A was MidRanger and the marker used in figure B & C was FullRanger (Norgen Biotek Corp., Canada).

Table 3.7: Figure legend for 3.18 A, B & C with respect to the lane number and plasmid name.

Lane Number	Figure Reference		
	Figure 3.18 A	Figure 3.18 B	Figure 3.18 C
1	Control	Control	Control
2	pE059-2	pE066-1	pE067-2
3	pE059-13	pE067-1	pE067-3
4	pE059-14	pE059-36	pE067-4
5	pE059-15	pE059-37	pE067-5
6	pE059-16	pE059-38	pE067-6
7	pE059-17	pE059-39	pE067-7
8	pE059-28	pE059-40	pE067-8
9	pE059-30	pE059-41	
10		pE059-42	

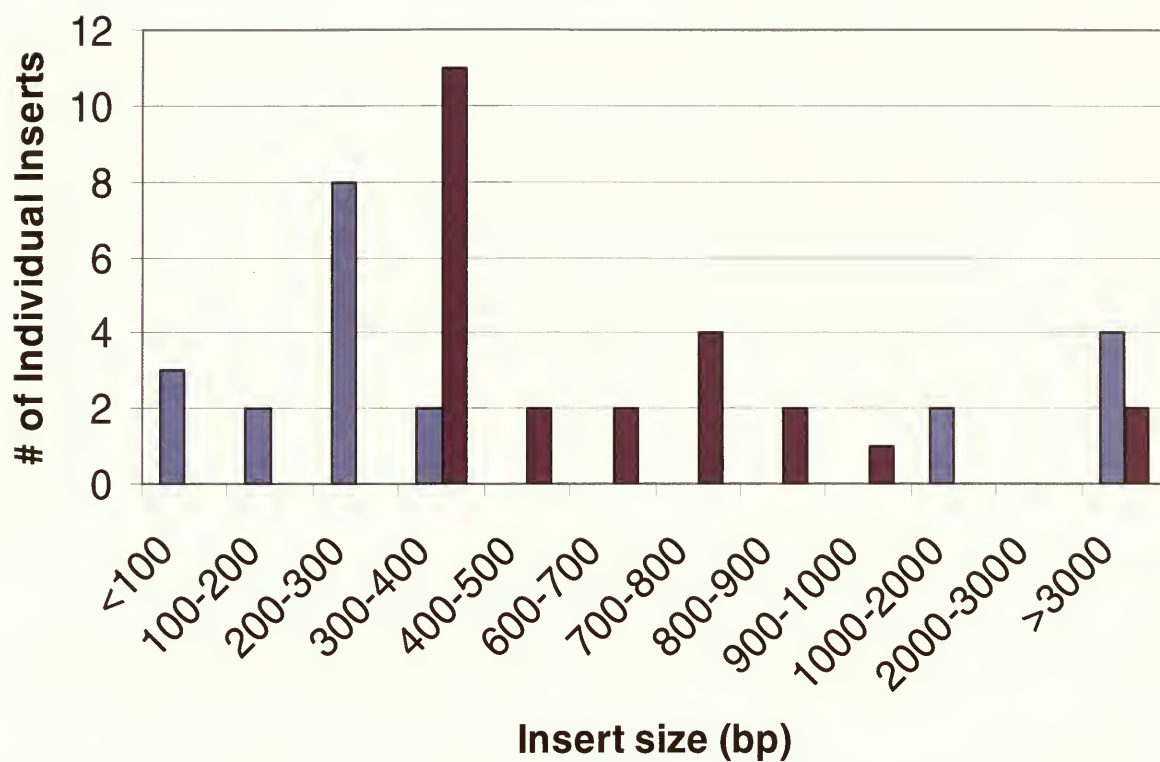


Figure 3.18: Distribution of insert sizes found in isolated plasmids. A total of 45 plasmids were obtained from mammalian cells that contained the 3Mix (■) and 3Lig (■) genomic DNA libraries which were maintained for approximately 32 days.

3.5.2: Sequencing and Nucleotide composition

Twelve plasmids, 6 from the 3mix library and 6 from the 3Lig library, were arbitrarily selected for sequencing (Table 3.8) and insert sizes were determined from the sequence data. The average insert size was determined to be 366 bp. This value was used to generate a random DNA sequence using the sequence manipulation suite (Stothard, 2000) for use as a negative control.

Analysis of the average nucleotide content of the genomic insert sequences (Figure 3.19) revealed a significantly higher thymine (T) content and a lower guanine (G) content in comparison to the average nucleotide content of the human genome (Karolchik *et al.*, 2003; Kent *et al.*, 2002). Trinucleotide analysis determined that permutations of the triplet oligo A or T were present in the genomic insert sequences more often than by random chance as determined by the randomly generated sequence (Figure 3.20).

Table 3.8: Selected plasmids that were sequenced and separated by their library origins.

3Mix Library		3Lig Library	
Plasmid Name	Insert Size (bp)	Plasmid Name	Insert Size (bp)
pE059-5	279	pE059-13	395
pE059-6	143	pE059-14	752
pE059-7	187	pE059-15	325
pE059-8	121	pE059-16	275
pE059-27	1133	pE059-17	275
pE059-33	230	pE059-41	269

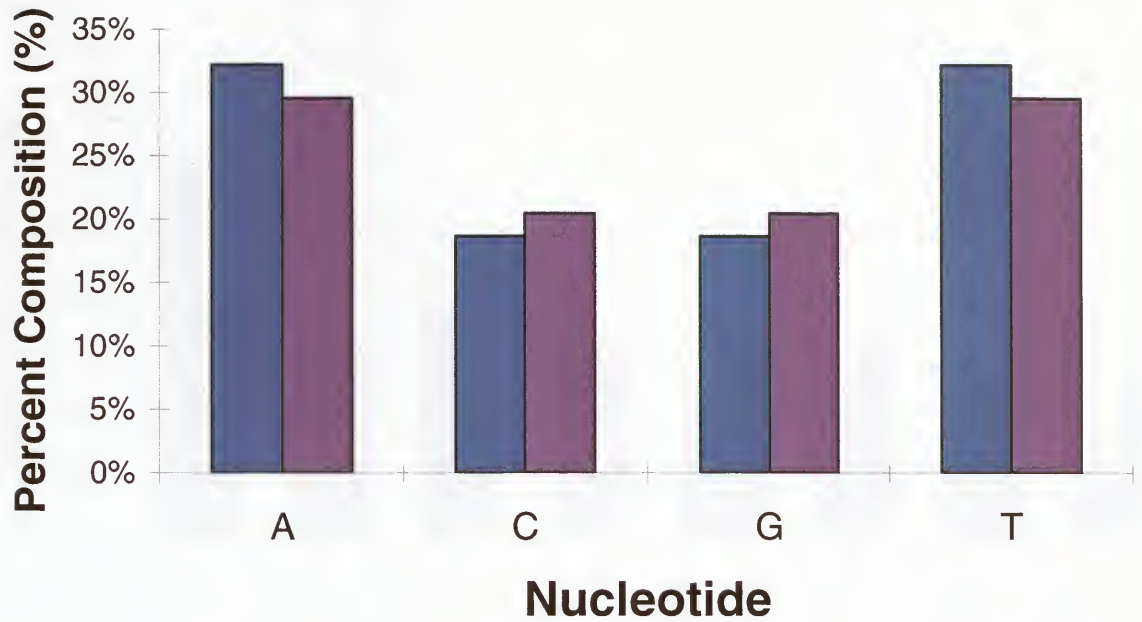


Figure 3.19: Average nucleotide composition of the 12 sequenced genomic DNA inserts. Nucleotide composition of the inserts (■) is in comparison to the nucleotide composition of the human genome (■). Average nucleotide composition of the human genome was determined from the UCSC Genome database (Karolchik *et al.*, 2003; Kent *et al.*, 2002). The confidence interval was determined for the composition of the insert sequences with $\alpha = 0.05$.

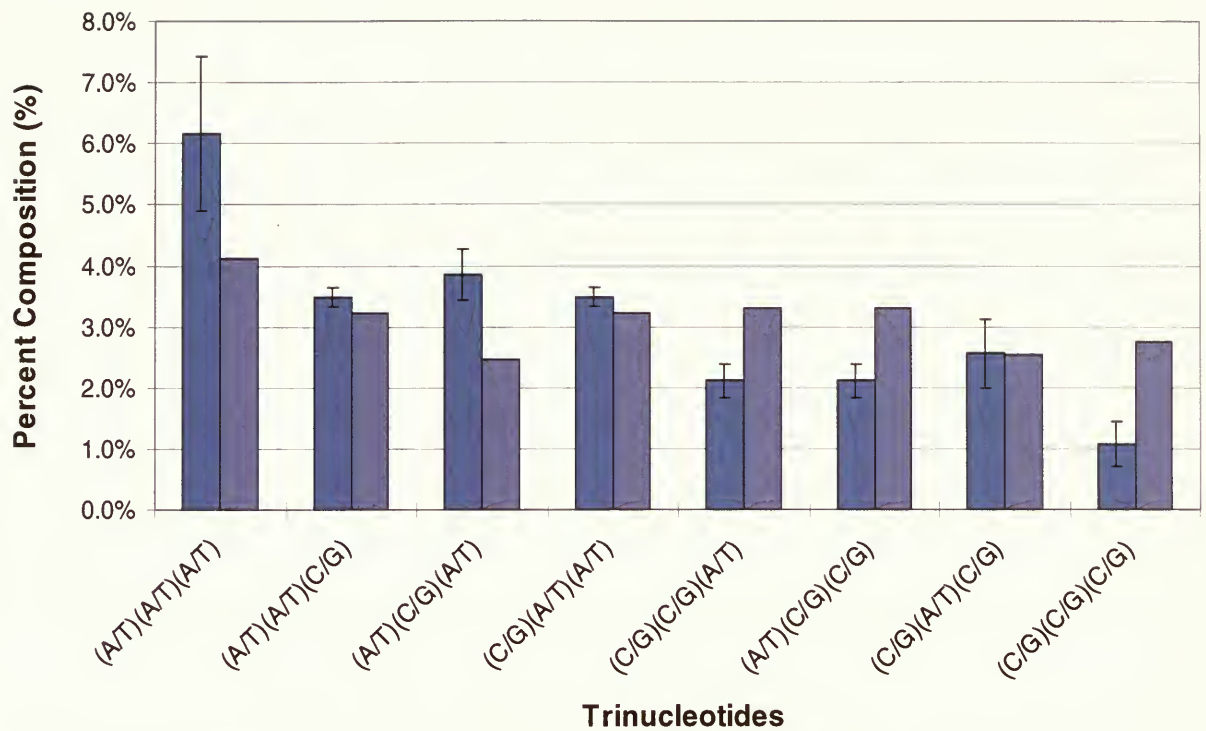


Figure 3.20: Average Trinucleotide content of the 12 sequenced genomic DNA inserts. Trinucleotide composition of the inserts (■) is in comparison to the average trinucleotide composition of the randomly generated DNA sequence (■). The confidence interval was determine for the composition of the insert sequences with $\alpha = 0.05$.

3.5.3: Local and Global Alignment of the 12 sequenced inserts

Alignment of the 12 genomic insert sequences has revealed the presence of a duplicate sequence which was confirmed by gel analysis (Figure 3.18A). The duplicate sequence had no significant effect on the previous nucleotide analysis; however the average insert size has changed slightly to 373 bp, when the two sequences were taken as one (leaving a total of 11 sequences). From here on out the duplicated sequences will be referred to as one sequence (pE059-16/17).

A graphical representation of the global alignment (Figure 3.21) of the 11 sequences using ClustalW (Higgins *et al.*, 1994) showed that the most conserved nucleotides were either A's or T's and that these were usually seen as part of an array of 6 to 8 A-T rich nucleotide clusters (indicated by the red circles).

Local alignment using Dialign (Brudno *et al.*, 2003) and visualized with the Application for Browsing Constraints (ABC; Cooper *et al.*, 2004) software showed that the region with the highest relative degree of similarity (Figure 3.22A) aligned approximately half the sequences. Local alignment occurred primarily in regions of high A/T content; particularly in thymine rich regions as seen in the middle of figure 3.22B. Unfortunately both the local and global alignment failed to reveal a consensus sequence that was relevant for all the chosen sequences.

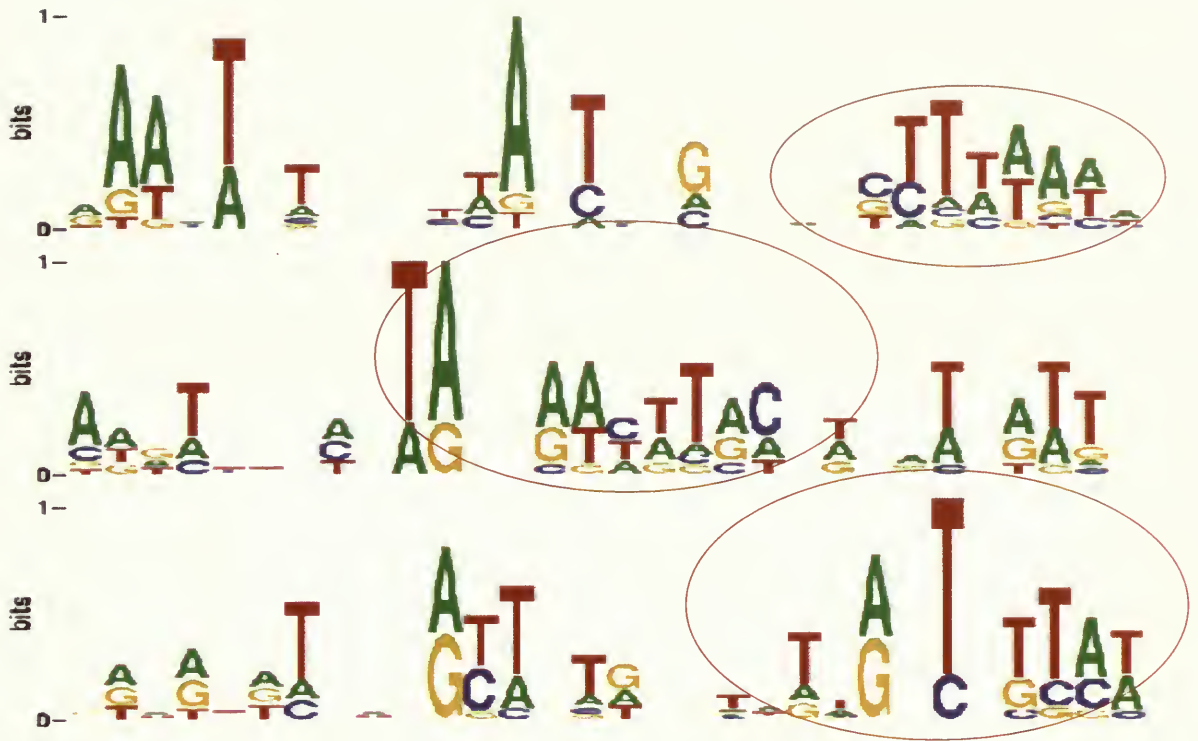


Figure 3.21: Graphical representation of the global alignment of the 11 insert sequences. The region above showed the highest degree of global alignment. The graphical representation was facilitated by the program Web Logo (Crooks *et al.*, 2004). The red circles highlight regions of particular interest which contain adenine & thymine rich regions.

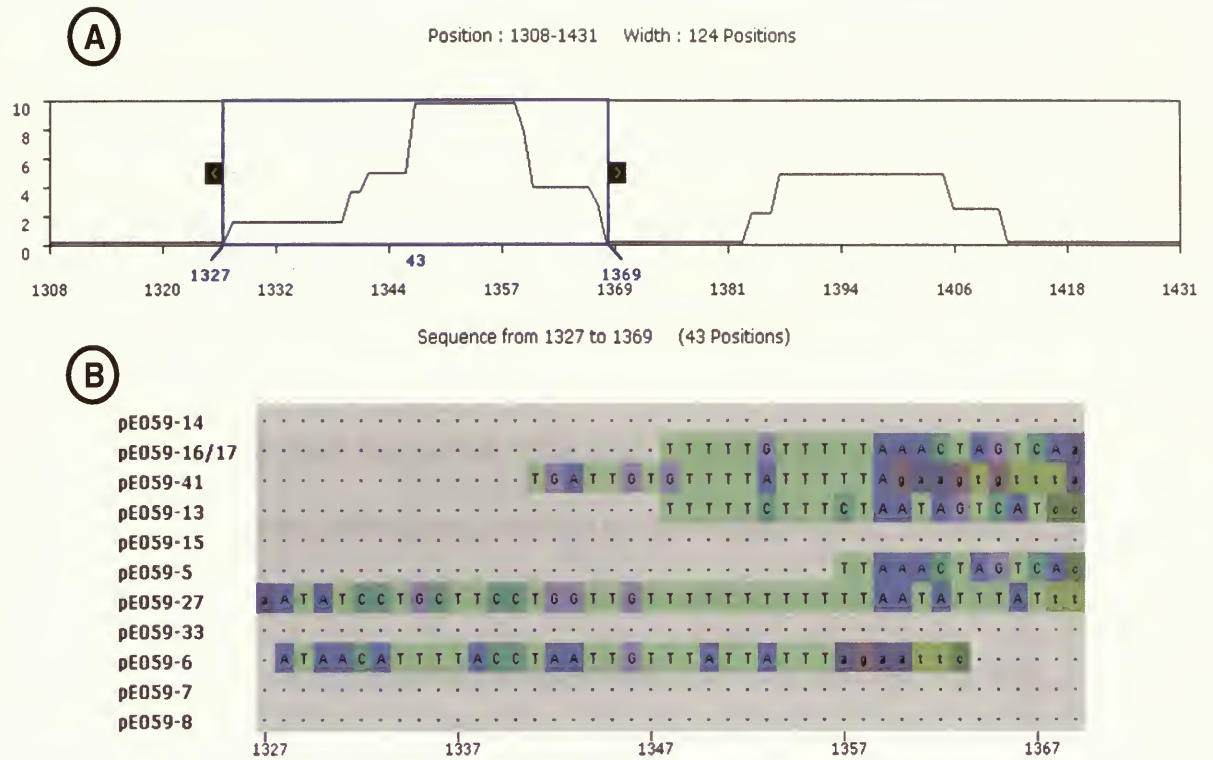


Figure 3.22: Local alignment of the 11 sequences visualized with ABC software (Cooper et al., 2004). The region with the highest relative similarity (A; represented by the blue rectangle) is rich in stretches of thymine and adenine for at least 6 base pairs (B).

3.5.4: Blast Search of the insert sequences and their chromosomal position.

A blast search was performed on the sequenced inserts to determine the possible origin of the insert and also if there was some known function associated with the sequence. Inserts originating from the 3Mix (Table 3.9) & 3Lig509 (Table 3.10) genomic DNA library usually had no known function associated with the DNA sequence. The insert sequences came from many different chromosomes with some overlap on chromosomes 2, 11 and 16. There was also no overlap of inserts sequences originating from the same BAC or Cosmid, therefore each sequence is distinctive from one another with the exception of pE059-16 and pE059-17 which was found to be the same sequence (section 3.5.3).

Table 3.9: Blast search results of inserts originating from the 3Mix genomic DNA library. Multiple hits resulted from these searches however the hit with the highest score and the lowest expected (E) value was used.

Plasmid	Chromosome	Source	E Value	Features	Accession Number
pE059-5	18	Homo sapiens	3e-06	None	AC007998
pE059-6	11	Homo sapiens	0.007	Repeat region (L1MC4a)	AC099842
pE059-7	16	Homo sapiens	0.61	None	AC093511
pE059-8	1	Homo sapiens	1e-22	None, in the region of multiple genes	AL513485
pE059-27	12q	Homo sapiens	4e-46	None	AC090503
pE059-33	2	Homo sapiens	7e-07	None	AC013410

Table 3.10: Blast search results of inserts originating from the 3lig509 genomic DNA library. Multiple hits resulted from these searches however the hit with the highest score and the lowest expected (E) value was used.

Plasmid	Chromosome	Source	E Value	Features	Accession Number
pE059-13	16	Homo sapiens	8e-119	None	AC018767
pE059-14	2	Homo sapiens	0.0	None	AC091485
pE059-15	X	Homo sapiens	5e-104	None	AL445306
pE059-16/17	4	Homo sapiens	1e-42	Repeat region (TTTTTG) _n	AC110805
pE059-41	11p13	Homo sapiens	4e-52	none	AL049635

3.5.5 Analysis of Scaffold Matrix attachment Regions (S/MARs) by MAR-Finder (Singh et al., 1997).

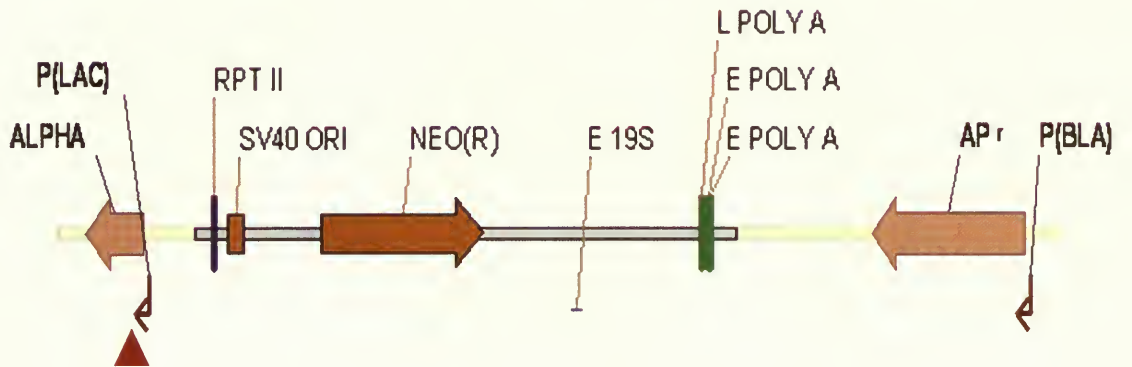
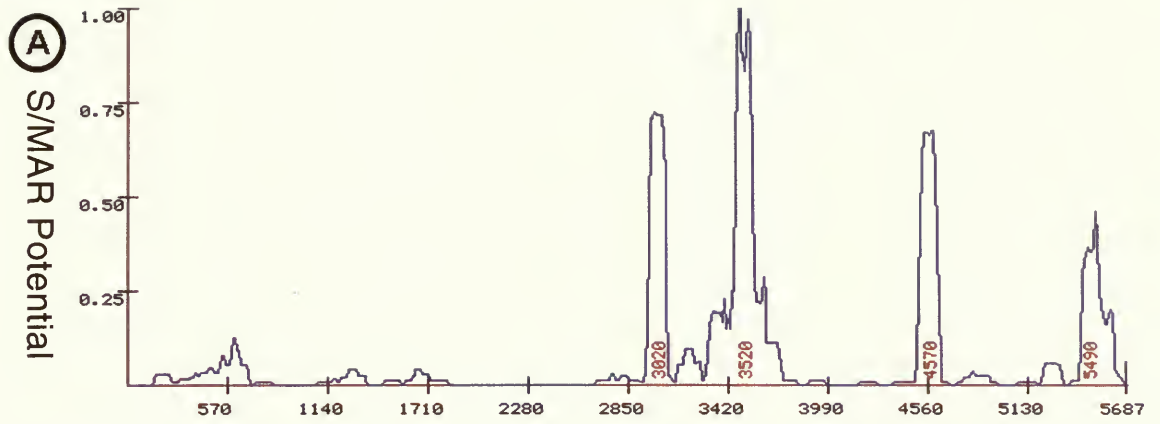
The high thymine content (section 3.5.2), the lack of a consensus sequence (section 3.5.3) and origins from many different parts of the genome (section 3.5.4) suggested that the insert sequences could be S/MARs. Using MAR-Finder as an *in silico* method for identifying S/MARs, we analyzed both the control plasmids pUC19neo (Figure 3.23A) and pUC19neo with the random sequence insert (Figure 3.23B). Although there are no potential S/MAR sites within the vicinity of the viral origin of replication on pUC19neo, S/MAR sites are present on other parts of the plasmid (Figure 3.23A). Potential S/MAR sites are indicated by a S/MAR potential (blue line) which exceeds the threshold value.

Using the results from the negative control (pUC19neo with the random sequence insert) the threshold for the presence of a S/MAR was reduced from 0.6 (default) to 0.3. It also became apparent that the default settings of the program were meant for the analysis of large pieces of DNA, therefore both the slide distance and the window width were decreased by 10-fold in order to examine the smaller inserts more thoroughly.

A positive control plasmid was produced *in silico* using vector NTI and inputted into the MAR-Finder program. The virtual plasmid contained the S/MAR from the 5' upstream region of the human interferon- β gene (Ac. M83137). Naturally the resulting figure (Figure 3.23C) contained multiple potential sites for a S/MAR within the inserted region.

S/MAR potential has been associated with half of the insert sequences originating from the 3Mix genomic DNA library (Figure 3.24) and over half

originated from the 3Lig509 library (Figure 3.25). The curves within the S/MAR containing insert regions shows that each curve is different with different maximum S/MAR potentials. The difference in the S/MAR potential is not limited to the insert regions, as downstream analysis reveals that every insert can change the potential of the S/MARs associated with the vector backbone as well.



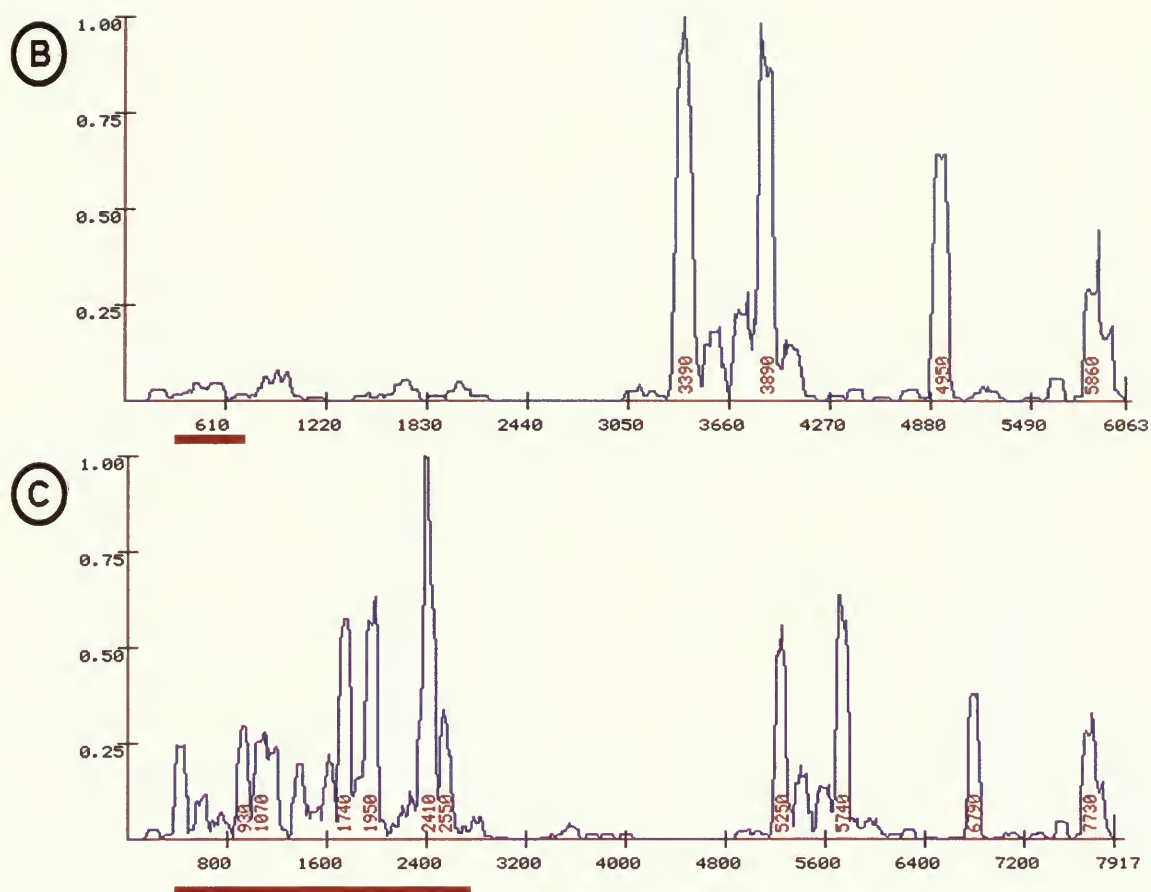


Figure 3.23: S/MAR potential (y-axis) with respect to relative base pair position of control plasmids. The plasmid pUC19neo (A) also has the linear form of the plasmid map for reference with the large arrow (▲) indicating the relative position of the *EcoRI* restriction enzyme site (397 bp). Using vector NTI the random sequence generated (373 bp) was inserted into the *EcoRI* position and analyzed. The red line (—) indicates the relative position of the insert sequence (B). The positive control plasmid contains a known S/MAR region. The region is indicated by the red line (—) (C; 2201 bp).

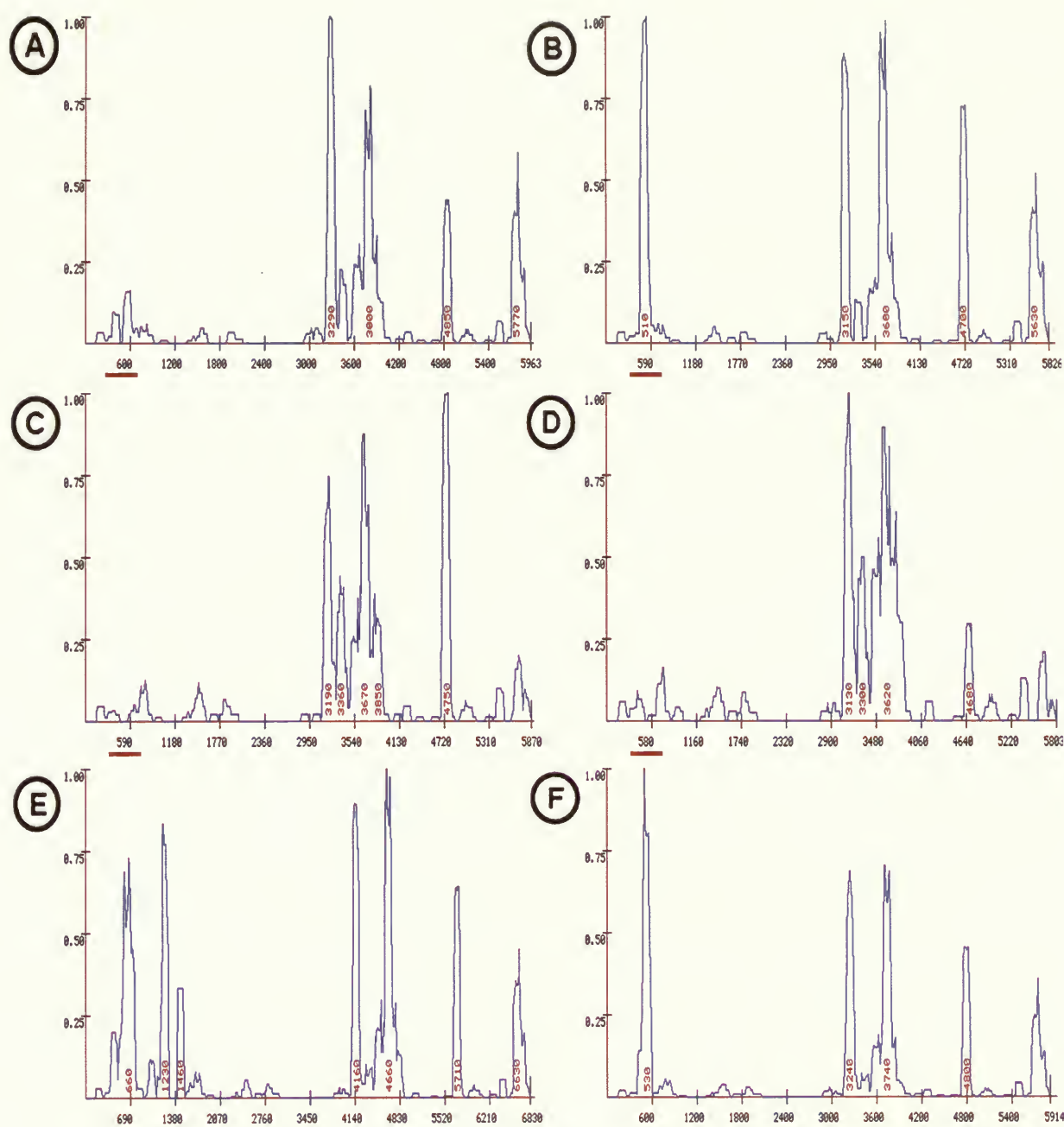


Figure 3.24: S/MAR potential (y-axis) with respect to base pair position of plasmids originating from the 3Mix genomic DNA library. The plasmid names are as follows: pE059-5 (A), pE059-6 (B), pE059-7 (C), pE059-8 (D), pE059-27 (E) and pE059-33 (F). The red line (—) indicates the relative position of the genomic DNA insert sequence.

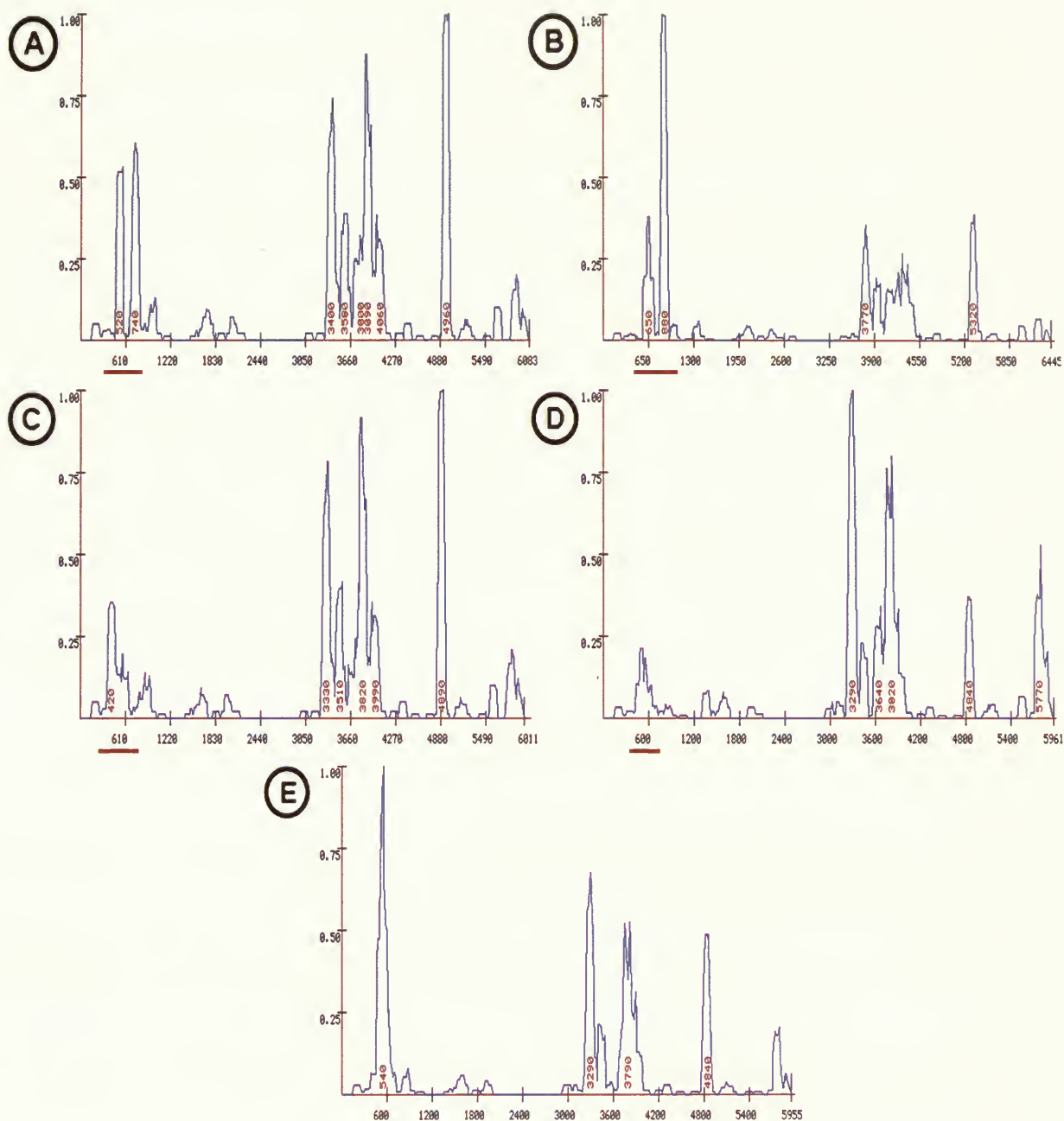


Figure 3.25: S/MAR potential (y-axis) with respect to base pair position of plasmids originating from the 3Lig genomic DNA library. The plasmid names are as follows: pE059-13 (A), pE059-14 (B), pE059-15(C), pE059-16/17(E) and pE059-41 (F). The red line (—) indicates the relative position of the genomic DNA insert sequence.

4.0: DISCUSSION

4.1: The Genomic DNA library

The objective of this study was to identify genomic DNA sequences which enhanced the stability of plasmid DNA in mammalian cells. To accomplish this goal, a plasmid vector (pUC19neo, Figure 3.2) was produced to carry a genomic DNA library which would be used to select for these specific sequences. Both the 3Mix & 3Lig genomic DNA libraries covered parts of the human genome with an average coverage of 51.4% (7.09 GEs) for the 3Mix library and 5.37% (0.73 GEs) for the 3Lig libraries. The plasmid maintaining sequences are believed to be plentiful in the human genome since it is unlikely that one sequence can maintain 3 billion base pairs. Therefore taking these two pieces of information we believe that the genomic DNA libraries together, allows for the efficient screening of the human genome and the successful capture of the sought after sequences.

Theoretical determination of the average insert size based on the restriction enzyme used; with *EcoRI* having an average insert size of 4096 bp and *Tsp509I* was an average size of 256 bp; allowed for the estimation of the percent coverage of the human genome (seen above). Although the original purpose of figure 3.4 was to observe whether the library contained randomly sized insets, subsequent examination identified the average insert size of the first *EcoRI* library (3Mix.1) to be 2550 bp, not the estimated insert size of 4096 bp (assuming a GC content of 50%).

This difference between the theoretical and experimentally determined average insert size has uncovered an important, yet overlooked addition to the results. This information would have allowed us to definitively determine the average insert size and therefore give a more accurate perspective into the scope of the genomic DNA libraries. Secondly, these results could also have been used to strengthen the position of sequence based selection in the mammalian cell. For example, after the production of the genomic DNA library an average size of 2550 bp was observed from the first 3Mix library, this average dropped to approximately 1000 bp after transfection and incubation in mammalian cells. The average size of the inserts originating from the 3lig libraries was approximately 800 bps in size (Figure 3.18). The average insert size from the two different libraries varies by only 200 bps after selection in mammalian cells; however the theoretical difference is approximately 3600 bp.

4.2: Isolation & Percent Recovery

Traditionally, isolation of low molecular weight DNA from mammalian cells is done using the Hirt extraction method (Hirt, 1967). Four low molecular weight extraction methods were tested, with the Hirt extraction method being the most efficient (Figure 3.8). Due to the cumbersome and consuming nature of the Hirt method a modified version of the Norgen Plasmid DNA isolation protocol was used instead. Using the modified Norgen plasmid isolation kit the percent recovery of known amounts of plasmid DNA from mammalian cells was determine to be $5.77 \pm 1.12\%$ (Section 3.3.4). Theoretical extrapolation of the linear range to were $y = 1$ (one colony obtained) resulted in a lower

detection limit of $3.43 \times 10^{-4} \mu\text{g}$. This value corresponds to 0.566 plasmids per cell in 1 million cells.

In an attempt to increase the detection sensitivity, aspects of the lysis and transformation process were probed. It was found that the majority of the DNA is lost during the lysis step of the extraction procedure (Figure 3.10). Modification of the lysis step would only result in small differences in the detection sensitivity. Increases in the transformation efficiencies theoretically would result in greater enhancement of the detection limit and therefore increase the probability of detecting plasmids. We decided to concentrate our efforts on improving the transformation step of the extraction assay.

Electroporation is known to increase transformation efficiencies up to 10^{10} CFU/ μg , however samples must be highly concentrated with a small volume (www.bio-rad.com). In our protocol, samples emerging from the extraction assay prior to transformation had a large volume and were dilute. The chemo-competent method allows for a larger volume of dilute sample and is therefore more amendable to the transformation step of the plasmid extraction assay. Using the Inuoé method (Inuoé *et al.*, 1990) of competent cell preparation efficiencies of 10^7 - 10^8 was observed.

A second concentration step was adopted so that an entire sample could be screened in one transformation. It was also found that the 2nd concentration step allowed the detection of smaller amounts of plasmid during the later stages of a longitudinal assay (Figure 3.11). Even with the 2nd concentration step, carry over of fragmented DNA and proteins from the plasmid isolation procedure would decrease the resistance of the

electroporation medium and reduce the efficiency of the procedure (Bio-rad Instruction Manual). The extraction procedure (Sections 2.18.1 & 2.18.2) and transformation procedure (Sections 2.3.2 & 2.4) as outlined in the materials and methods was deemed to be adequate for the purposes of this study and used in subsequent experiments.

4.3: Isolation Variability

Calcium phosphate transfections are inherently variable (Jordan *et al.*, 1996), with the greatest amount of variability observed between experiments done on two different days (Figure 3.6). Smaller amounts of variability are also been seen between samples from different tubes which were transfected in tandem on the same day (Figure 3.7). This variability can beneficially randomize the plasmids which are taken up by the cells. However it can be extremely detrimental for parametric statistical analysis which assumes equal variance between samples and a normal distribution (Zar, 1984).

This inherent variability is apparent especially on day 32 (figure 3.15). Parametric statistical analysis using the one way ANOVA indicated that the experimental treatments were not significantly different from the control NS($p=0.50$, $df=9$, $F=0.743$). The same result was found using a 2-way ANOVA NS($p=0.2477$, $df=2$, $F=1.78$). However the results from the 2-way ANOVA also indicated that the variance between the replicates was significantly different ($p=0.0421$, $df=3$, $F=5.18$). Also, due to the number of zeros in the data it was clear that the data were not normally distributed. Violation of the assumptions associated with parametric data analysis renders the results of these tests inconclusive.

Use of a nonparametric analysis is possible; however it is not as powerful as a parametric analysis (Zar, 1984). Using the Mann-Whitney Test with tied ranks it was found that the null hypothesis is rejected; therefore the control and the 3lig509 sample are not equal ($p = 0.05$, $U=1$). It is important to note that no parameters or parameter estimates are employed in the statistical hypothesis or in the calculation of the U value (Zar, 1984).

The variability between samples is not conducive to determining statistical differences between the means, especially with parametric assumptions. It is clear that direct comparison to the control plasmid pUC19neo will not decisively determine if the experimental plasmid is maintained. It is possible that this view of the problem is too simplistic and that variables which are induced by the cell should also take part in discriminating the difference between a non-maintained and a maintained plasmid.

4.4: Plasmid Maintenance

Once in the mammalian nucleus, an episomal plasmid can be lost by (a number of cellular mechanisms such as) simple dilution of the plasmid during mitosis (Pauletti *et al.*, 1990), cellular degradation by nucleases (Wilson *et al.*, 1992) or homologous recombination (Bollag *et al.*, 1989). Plasmid DNA can also be lost or dissociate during the physical separation of the sister chromatids during telophase. Analysis of the total population of cells with respect to the variety of plasmid classifications presents a better explanation of the events seen in the longitudinal study of plasmid maintenance (Figure 3.14).

The total amount of plasmid DNA found in the population is dictated by the relative rates of plasmid increase (due to cellular and subsequent plasmid replication) and plasmid loss (due to cell death and cellular mechanisms). These rates will lead to 5 theoretical patterns (Figure 4.1). In the simplest form the amount of plasmid in the population will increase to a limit set by the cell and the number of cells will increase exponentially. This will result in a pattern which resembles traditional exponential growth (Figure 4.1; —).

The other 4 scenarios take into account the carrying capacity of the cellular environment. The carrying capacity of tissue culture cells is determinant upon the presence of sufficient amounts of nutrients and is therefore dictated by the medium. Logistical increase of plasmid DNA can be theoretically attributed to a plasmid which can both replicate and maintain itself in a cell (Figure 4.1; —). This particular pattern assumes that maintenance and replication is 100%, in actuality it has been experimentally determined that 6-8% of EBV derived vectors are lost from a cell population every cell cycle (Van Craenenbroeck *et al.*, 2000). Therefore a non-viral derived plasmid in mammalian cells should also have similar forces affecting its maintenance and replication.

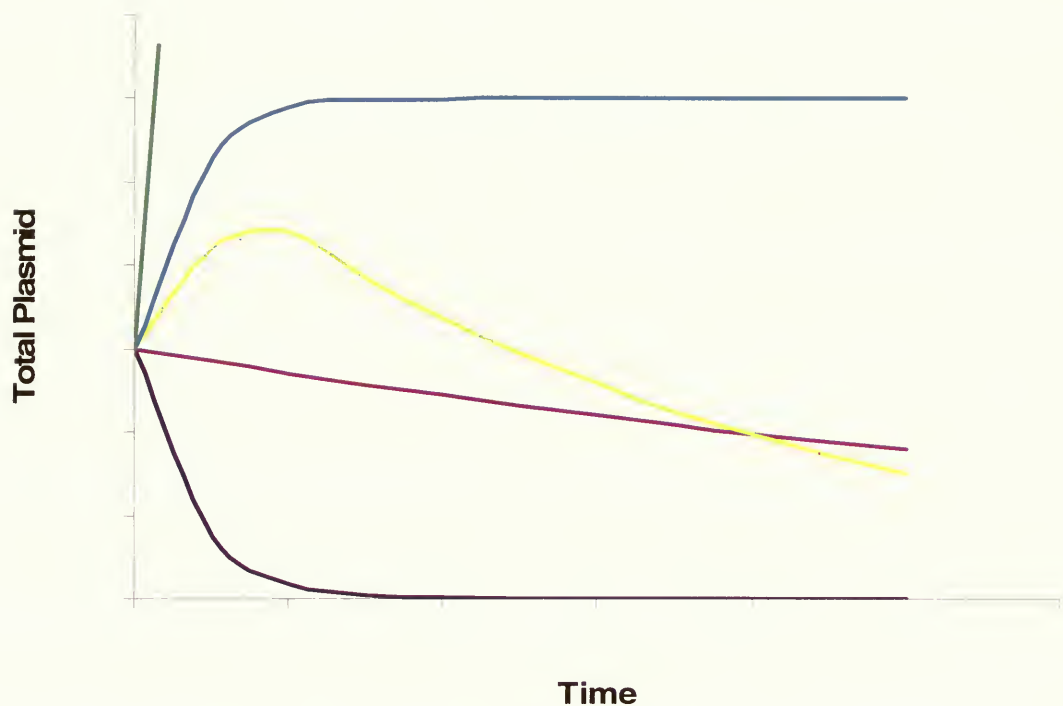


Figure 4.2: Theoretical effect of maintenance and replication elements on the total amount of plasmid DNA present in a cell population. Plasmid DNA will increase exponentially as dictated by the cell (pattern 1; —). In the presence of a limiting factor (carrying capacity) such as medium, the amount of plasmid with both a replication and maintenance (pattern 2; —) element will follow a logistic growth pattern. A plasmid with no replication or maintenance capacity would be lost quite readily (pattern 3; —). A flawlessly maintained plasmid (pattern 4; —) would also generally decrease over time. Finally a plasmid that contains a mammalian origin (pattern 5; —) will allow for an initial increase in total plasmid amount; however a general decreasing trend is seen due to plasmid loss. Though we have depicted patterns 4 and 5 as distinct lines in reality the patterns would fall somewhere between pattern 2 and 3.

In the third pattern both plasmid maintenance and replication is not present (Figure 4.1; —). Therefore no new plasmid is produced and plasmid DNA which is present is lost due to cell death and other cellular mechanisms (described above). The fourth pattern of plasmid maintenance is dictated primarily by mitosis (Figure 4.1; —). In this scenario no new plasmid is created by replication however the plasmid which is present is maintained over time. Both of these patterns result in a decrease number of plasmid copies per cell and therefore a reduction in the total amount of plasmid over time.

The fifth and most complex pattern represents plasmids that are replicated but not maintained in mammalian cells (Figure 4.1; —). As described before, the amount of plasmid DNA in the cell population changes over time and is determined by the relative rates of plasmid increase (due to replication) and the rate of plasmid loss (cellular death and plasmid degradation). In the beginning the rate of plasmid increase is greater than the rate of plasmid loss and therefore results in an increase in the total amount of plasmid DNA in the cell population. Later, the rate of plasmid loss increases due to the limiting factors in the environment until the rate of plasmid increase is equal to plasmid loss. After this time point the rate of plasmid increase is less than the rate of plasmid loss and therefore the total amount of plasmid DNA decreases. The presence of an origin, but lack of a maintenance sequence is similar to the control plasmid pUC19neo (Figure 3.12 & Figure 3.14) and therefore we may extrapolate, assuming this model is correct, the continued reduction of the plasmid in the cell population past day 32.

The information gained from figure 4.1 can be used to describe the 4 stages of the longitudinal assay in section 3.4.2 (Figure 3.14). During the first stage (between days 0 and 5) the rate of plasmid loss is much greater than rate of plasmid increase. This could be attributed to the extensive degradation of the extracellular plasmid DNA and the intracellular degradation of plasmid which did not enter the nucleus (Figure 1.6). During stage 2 (days 5-24) G418 selection removes cells that do not contain the *neo^R* gene. This allowed cells which contained the *neo^R* gene to grow logistically and therefore boosted the rate of plasmid increase (due to cellular replication). At the same time the rate of plasmid loss is reduced due to a phenomenon known as transient drug resistance (reviewed in Krysan *et al.*, 1989). Transient drug resistance can artificially reduce the rate of plasmid loss by occupying the plasmid with the job of transcribing the neomycin resistance gene. Together this would result in plasmids, which are replicated and maintained, to eventually reach detectable levels in the cell population.

During stage 3 (days 24-28) selective pressure is removed and therefore the rate of plasmid loss increases. Up until this time point (Day 28) the plasmids which contain a maintenance element have been overshadowed by background vectors and therefore a general decreasing amount of plasmid DNA was observed. Between days 28 and 32 (stage 4) an increasing amount of plasmid DNA was observed in plasmids which contain genomic DNA inserts. The lack of any sort of increasing amount of plasmid over the 28 days of analysis was attributed to the fact that the original population of plasmid which contains both a maintenance and replication element was initially

extremely small in comparison to the other plasmids in the library. It therefore took time for these plasmids to replicate in the cell population because unlike bacterial cells, mammalian cells only undergo mitosis every 22-24 hours (Fields *et al.*, 1996).

Assuming the fundamentals of the model produced in figure 4.1 are true (that patterns 4 & 5 fall between 2 and 3) it is possible to conclude that the insert containing samples obtained on day 32 include both maintenance and non-maintenance elements.

4.5: Sequence Analysis:

Day 32, although not significant according to statistical analysis (section 3.4.2) day 32 (Figure 3.15) is still an important timepoint, assuming the model (Figure 4.1) is correct. Screening of the white colonies from day 32 resulted in the isolation of 21 plasmids from the 3Mix genomic DNA library (Figure 3.16 & Table 3.6), and 24 plasmids from the 3Lig509 genomic DNA library (Figure 3.17 & Table 3.7). From these 45 plasmids, 12 plasmids (6 from each library) were arbitrarily selected for sequencing.

A blast search of the insert sequences was performed (section 3.5.4) to determine the origin and putative function of the sequences. These analyses indicated that the sequences were derived from different parts of the genome and that there was generally no known function related to these sequences. The presence of “Junk” DNA within the genome with functionality is well established. However the efficient screening of this DNA for functionality is laborious without the aid of bioinformatics (Bejerano *et al.*, 2005).

Determination of the nucleotide composition of the inserts revealed significantly higher thymine content (Figure 3.19) in comparison to the average thymine content of the human genome (Karolchik *et al.*, 2003; Kent *et al.*, 2002). Significantly higher A or T content has been previously associated with ARS regions in yeast and S/MAR sequences in humans (reviewed in Price *et al.*, 2003; Clyne & Kelly, 1997).

It was found that two of the insert sequences selected for sequencing were the same. There are three possible scenarios for this to occur. Firstly this could have resulted from cross contamination between samples. Secondly the initial amount of this particular plasmid in the genomic DNA libraries before introduction to the mammalian cells was extremely high, and therefore is still present in day 32 samples. Finally it is possible that this plasmid was present in greater quantity within the cell population and therefore created two independent events which were screened separately. Unfortunately the first scenario is more plausible and therefore the two sequences were treated as one during the alignment analysis (pE059-16/17).

The final analysis was a global and local alignment of the 11 selected insert sequences. Global alignment (Figure 3.21) revealed a region of moderate conservation. Further examination using weblogo (Crooks *et al.*, 2004) showed strings of 6-8 conserved nucleotides containing primarily adenine and thymine base pairs. The local alignment (Figure 3.22) also revealed a highly conserved region containing mainly A and T nucleotides. These A/T-rich regions are recognized by proteins containing A/T hook motifs (Aravind & Landsman, 1998). In yeast, proteins which contain these motifs

are typified by the ORC4p, a protein that belonging to the origin recognition complex (Lee *et al.*, 2001; Chuang & Kelly, 1999). In humans, the A/T hook motif is characteristic of high mobility group proteins which are associated with transcription and replication (reviewed in Lee *et al.*, 2001; Chuang & Kelly, 1999; Aravind & Landsman 1998).

The high A/T content (section 3.5.2), the lack of a consensus sequence (section 3.5.3) and origins from many different parts of the genome (section 3.5.4) suggests that these sequences could be S/MARs. S/MARs have been used to maintain plasmids episomally for up to 100 cell generations (Piechaczek *et al.*, 1999). It is hypothesized that this is facilitated through the interaction of the scaffold attachment factor A (SAF-A) and the residual nuclear matrix surrounding the chromosome during anaphase (Kipp *et al.*, 2000). Using Mar-Finder (Singh *et al.*, 1997) as a computational method to recognize S/MARs, we identified S/MAR regions present in approximately 60% (7 out of 11) of the screened insert genomic DNA sequences.

The presence of S/MAR sequences in this screen is interesting because it has been proven in the literature that S/MARs can replace the maintenance function of the large-T antigen (Piechaczek *et al.*, 1997). It is also important to note the relative proximity to the SV40 origin in relationship to the genomic insertions, and the fact that S/MAR sequences enhance or facilitate in some way the activity of mammalian origins (Bode *et al.*, 2001).

We have already stated that samples from day 32 will contain both maintenance and non-maintenance sequences. The fact that S/MAR and non-S/MAR sequences are present in an approximately equal ratio within the

screened genomic DNA inserts only strengthens the conclusion that S/MAR sequences facilitate the maintenance of plasmid DNA in mammalian cells.

4.6: Future Research & Questions

In this report we have demonstrated a method for the selection and isolation of sequences which enhances the stability of plasmid DNA in mammalian cells, resulting in the isolation of 45 plasmids. The next step would be to transfect cells with individual plasmids rather than the heterogeneous mixture used in this report. This experiment could be enhanced by producing the plasmid pUC19neo with a S/MAR sequence and using this as a positive control for comparison.

Also undetermined is the functionality of the sequences; are they dependent upon the vector, or can they function in other plasmids. It has been shown that certain sequences can affect other downstream sequences by changing the structural integrity of the DNA (Bode *et al.*, 2001). This effect could also be responsible for the actual maintenance property seen within the plasmids.

Finally, the models which explain the distribution of plasmids in the cell population were based only on the mechanisms of mitosis, and simple population ecology. Experimental verification of these models could then be used to determine the dynamics of a plasmid in the human body, especially when used in gene therapy against cancer cells.

5.0: CONCLUSION

The primary goal of this project was the isolation and characterization of sequences which enhanced the stability of plasmid DNA in mammalian cells. This was accomplished by developing a means to extract and isolate low molecular weight DNA from mammalian cells in an efficient manner. We were able to achieve this using a modified Norgen Mini-Prep protocol coupled with an enhanced transformation step (Inoue *et al.*, 1990). Using six independent genomic DNA libraries we selected for plasmids that remain episomal and unaltered in the mammalian cells for up to 32 days.

We found that these sequences:

1. Were more Thymine rich than the human genome.
2. Contained no consensus sequence as determined by both a local and global alignment.
3. Originated from difference regions of the genome.

Bioinformatical analysis revealed that approximately 60% (7 out of 11 sequences) contain S/MARs, a known region for nuclear cytoskeleton binding (reviewed in Bode *et al.*, 2001; Piechaczek *et al.*, 1999). These regions have also been associated with DNA replication in mammalian cells (reviewed in Bode *et al.*, 1998). Therefore the implication of these sequences with the episomal maintenance and replication of plasmid DNA is not unreasonable. Further research into the utility of these sequences will allow for the engineering of safe and reliable vectors. This would result in fields such as gene therapy to become a viable option for the treatment of human disease.

6.0 LITERATURE CITED

- Aiyar A., Tyree C. and Sugden B. (1998) The plasmid replicon of EBV consists of multiple *cis*-acting elements that facilitate DNA synthesis by the cell and a viral maintenance element. *The European Molecular Biology Organization Journal*, **17(21)**: 6394-6403.
- Ak P. & Benham C.J. (2005) Susceptibility to Superhelically Driven DNA Duplex Destabilization: A Highly Conserved Property of Yeast Replication Origins. *PLoS Computational Biology*, **1**:41-46.
- Aravind L. & Landsman D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Research*, **26(19)**: 4413-4421.
- Ascenzioni F., Donini P. & Lipps H.J. (1997) Mammalian artificial chromosomes--vectors for somatic gene therapy. *Cancer Letters* **118(2)**:135-42.
- Baiker A., Maercker C., Piechaczek C., Schmidt S.B.A., Bode J., Benham C. and Lipps H.J. (2000) Mitotic stability of an episomal vector containing a human scaffold/matrix-attached region is provided by association with nuclear matrix. *Nature Cell Biology*, **2**: 182-184.
- Bejerano G., Siepel A.C. Kent W.J & Haussler D. (2005) Computational screening of conserved genomic DNA in search of functional noncoding elements. *Nature Methods*, **2(7)**: 535-545.
- Benham C., Kohwi-Shigematsu T. & Bode J. (1997) Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. *Journal of Molecular Biology*, **274(2)**:181-96.
- Benham C.J. (1993) Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci. *Proceeding in the National Academy of Science*, **90**: 2999-3003.

- Birnboim H.C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Research*, **7(6)**:1513-1523.
- Blaber J. 1998 BCH5425 Molecular Biology and Biotechnology <
<http://wine1.sb.fsu.edu/bch5425/lect28/lect28.htm>> Accessed 2005 Aug. 23.
- Bode J., Bartsch J., Bouliskas T., Iber M., Mielke C., Schubeler D., Seibler J. and Benham C. (1998) Transcription-promoting genomic sites in mammalia: their elucidation and architectural principles. *Gene Therapy and Molecular Biology*, **1**: 551-580.
- Bode J., Fetzer C.P., Nehlsen K., Scinteie M., Hinrichs B., Baiker A., Piechaczek C., Benhan C., and Lipps H.J. (2001) The Hitchhiking Principle: Optimizing episomal vectors for the use in gene therapy and biotechnology. *Gene Therapy and Molecular Biology*, **6**: 33-46.
- Bode J., Kohwi Y., Dickinson L., Joh T., Klehr D., Mielke C. and Kohwi-Shigematsu T. (1992) Biological Significance of Unwinding Capability of Nuclear Matrix-Associating DNA. *Science*, **225**:195-197.
- Bollag R.J., Waldman A.S. and Liskay R.M. (1989) Homologous Recombination in mammalian cells. *Annual Reviews of Genetics*, **23**: 199-225.
- Bowers M.T., Ramachandran S. and Yu B.W. (1999) Modified Miniprep Method for the Rapid Recovery of Episomes from Transfected Breast Epithelial Cells. *BioTechniques*, **26**: 276-282.
- Brudno M., Chapman M., Götting B., Batzoglou S. & Morgenstern B. (2003) Fast and sensitive multiple alignment of large genomic sequences. *BMC Bioinformatics*, **4**: 66.
- Calos M.P. (1996) The potential of extrachromosomal replicating vectors for gene therapy. *Trends in Genetics*, **12(11)**: 463-466.
- Calos M.P. (1998) Stability without a centromere. *Proceedings in the National Academy of Science*, **95**: 4084-4085.

- Carter M.J & Milton I.D. (1993) An inexpensive and simple method for DNA purifications on silica particles. *Nucleic Acids Research*, **21**(4): 1044.
- Chang Y., Cheng S.-D. and Tsai C.-H. (2002) Chromosomal integration of Epstein-barr virus genomes in nasopharyngeal carcinoma cells. *Head & Neck*, **24**: 143-150.
- Chow C-M., Athanassiadou A., Raguz S., Psiouri L., Harland L., Malik M., Aitken M.A., Grosveld F. and Antoniou M. (2002) LCR-mediated, long-term tissue specific gene expression within replicating episomal plasmid and cosmid vectors. *Gene Therapy*, **9**: 327-336.
- Chuang R.Y. & Kelly T.J. (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proceedings in the National Academy of Science*, **96**(6):2656-61.
- Clemson M. and Kelly W.J. (2003) Optimizing alkaline lysis for DNA plasmid recovery. *Biotechnology Applied Biochemistry*, **37**: 235-244.
- Clyne R.K. and Kelly T.J. (1997) Identification of Autonomously Replicating Sequence (ARS) Elements in Eukaryotic Cells. *Methods: A Companion to Methods in Enzymology*, **13**: 221-233.
- Collaco R.F., Cao X. & Trempe JP. (1999) A helper virus-free packaging system for recombinant adeno-associated virus vectors. *Gene*, **238**(2):397-405.
- Cooper G.M., Singaravelu S.A.G. & Sidow A. (2004) ABC: software for interactive browsing of genomic multiple sequence alignment data *BMC Bioinformatics*, **5**:192.
- Cooper M.J., Lippa M., Payne J.M., Hatzivassiliou G., Reifenberg E., Fayazi B., Perales J.C., Morrison L.J., Templeton D., Piekarz R.L., Tan J. (1997) Safety-modified episomal vectors for human gene therapy. *Proceedings in the National Academy of Science*, **94**: 6450-6455.
- Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. (2004) WebLogo: A sequence logo generator, *Genome Research*, **14**:1188-1190.

- Dale J.W. (1998) Molecular Genetics of Bacteria (3rd ed.), *Nucleic Acid Structure and Function*, Surrey, UK: School of Biological Sciences, University of Surrey.
- Dean D.A. (1997) Import of Plasmid DNA into the Nucleus is Sequence Specific. *Experimental Cell Research*, **230**: 293-302.
- Dean D.A., Strong D.D. & Zimmer W.E. (2005) Nuclear entry of nonviral vectors. *Gene Therapy*, **12**: 881-890.
- Fields B.N., Knipe D.M. & Howley P.M. (1996) Fundamental Virology (3rd ed.), *Viral Genetics*, Boston, USA: Harvard Medical School.
- Freshney R.I. (2000) Culture of animal Cells; a manual of basic technique, 4th edition. Wiley-Liss, Inc., New York, NY.
- Friard, O. (2002) Annhyb. *Free Software Foundation, Inc.* 675 Mass Ave, Cambridge, MA 02139, USA.
- Frisch M., Frech K., Klingenhoff A., Cartharius K., Liebich I. and Werner T. (2001) *In Silico* Prediction of Scaffold/Matrix Attachment Regions in Large Genomic Sequences. *Genome Research*, **12**: 349-354.
- Frouin I., Montecucco A., Spadari S. & Maga G. (2003) DNA replication: a complex matter. *European Molecular Biology Organization Reports*, **4(7)**: 666-670.
- Fujita M., Ishimi Y., Nakamura H., Kiyono T. and Tsurumi T. (2002) Nuclear Organization of DNA Replication Initiation Proteins in Mammalian Cells. *The Journal of Biological Chemistry*, **227(12)**: 10354-10361.
- Graham F.L. & van der Eb A.J. (1973) Transformation of rat cells by DNA of human adenovirus 5. *Virology*, **54(2)**:536-9.
- Grimes B. and Cooke H. (1998) Engineering mammalian chromosomes. *Human Molecular Genetics*, **7(10)**: 1635-1640.
- Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G. & Gibson T.J.(1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment

- through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. **22**:4673-4680.
- Hirt B. (1967) Selective Extraction of Polyoma DNA from Infected Mouse Cell Cultures. *Journal of Molecular Biology*, **26**: 365-369.
- Houk B.E., Hochhaus G. and Hughes J.A. (1999) Kinetic Modeling of Plasmid DNA Degradation in Rat Plasma. *American Association of Pharmaceutical Scientists Pharmsci*, **1**(4): 1-6.
- Huberman J.A. (2003) Editorial: DNA replication, the cell cycle and genome stability. *Mutation Research*, article in press.
- Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**: 23-28.
- Jenke A.C.W., Scinteie M.F., Stehle I.M. and Lipps H.J. (2004a) Expression of a transgene encoded on a non-viral episomal vector is not subject to epigenetic silencing by cytosine methylation. *Molecular Biology Reports*, **31**: 85-90.
- Jenke A.C.W., Stehle I.M. Herrmann F., Eisenberger T., Baiker A., Bode J., Fackelmayer F.O. and Lipps H.J. (2004b) Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. *Proceedings in the National Academy of Science*, **101**(31): 11322-11327.
- Jenke B.H.C., Fetzner C.P., Stehle I.M., Jonsson F., Fackelmayer F.O., Conradt H., Bode J. and Lipps H.J. (2002) An episomally replicating vector binds to the nuclear matrix protein SAF-A *in vivo*. *European Molecular Biology Organization Reports*, **3**(4): 349-354.
- Jordan M., Schallhorn A. & Wurm F.M. (1996) Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Research*, **24**(4): 596-601.

- Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D & Kent WJ (2003) The UCSC Genome Browser Database. *Nucl. Acids Res* 31(1): 51-54
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM & Haussler D (2002) The Human Genome Browser at UCSC. *Genome Res.* **12**(6): 996-1006.
- Kipp M., Gohring F., Ostendorp T., VanDrunen C.M., VanDriel R., Przybylski M. and Fackelmayer F.O. (2000) SAF-Box, a Conserved Protein Domain That Specifically Recognizes Scaffold Attachment Region DNA. *Molecular and Cellular Biology*, **20**(20): 7480-7489.
- Krysan P.J., Haase S.B. and Calos M.P. (1989) Isolation of Human Sequences that Replicate Autonomously in Human Cells. *Molecular and Cellular Biology*, **9**(3): 1026-1033.
- Lee J-K., Moon K-Y., Jiang Y. & Hurwitz J. (2001) The *Schizosaccharomyces pombe* origin of recognition complex interacts with multiple AT-Rich regions of the replication origin DNA by means of the AT-Hook domains of the spOrc4 protein. *Proceedings in the National Academy of Science*, **98**(24):13589-13594.
- Liebich I., Bode J., Reuter I. and Wingender E. (2002) Evaluation of Sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Research*, **30**(15): 3433-3442.
- Lipps H.J., Jenke A.C., Nehlsen K., Scinteie M.F., Stehle I.M. & Bode J. (2003) Chromosome-based vectors for gene therapy. *Gene*, **304**:23-33.
- Lodish H, Berk A, Zipursku LS, Matsudaira P, Baltimore D & Darnell J (2000) *Molecular Cell Biology: Recombinant DNA Genomics*. 4th ed. New York. NY. W. H. Freeman and Company. p 207-253.
- Lupton S. and Levine A.J. (1985) Mapping Genetic Elements of Epstein-Barr Virus that Facilitate Extrachromosomal Persistence of Epstein-Barr Virus-Derived Plasmids in Human Cells. *Molecular and Cellular Biology*, **5**(10): 2533-2542.

- Magin-Lachmann C., Kotzamanic G., D'Aiuto L., Wagner E. and Huxley C. (2003) Retrofitting BACs with G418 resistance, luciferase, and *oriP* and *EBNA-1* – new vectors for *in vitro* and *in vivo* delivery. *BMC Biotechnology*, **3**: 2-13.
- Mielke C., Christensen M.O., Westergaard O., Bode J., Benham C.J. and Breindl M. (2002) Multiple Collagen I Gene Regulatory Elements Have Sites of Stress-Induced DNA Duplex Destabilization and Nuclear Scaffold/Matrix Association Potential. *Journal of Cellular Biochemistry*, **84**: 484-496.
- Nielsen T.O., Cossons N.H., Zannis-Hadjopoulos M. and Price G.B. (2000) Circular YAC Vectors Containing Short Mammalian Origin Sequences Are Maintained Under Selection as HeLa Episomes. *Journal of Cellular Biochemistry*, **76**: 674-685.
- Pauletti G., Lai E. and Attardi G. (1990) Early appearance and long-term persistence of the submicroscopic extrachromosomal elements (amplisomes) containing the amplified *DHFR* genes in human cell lines. *Proceedings in the National Academy of Science*, **87**: 2955-2959.
- Piechaczek C., Fetzter C., Baiker A., Bode J. and Lipps H.J. (1999) A vector based on the SV40 origin of replication and chromosomal S/MARs replicate episomally in CHO cells. *Nucleic Acids Research*, **27(2)**: 426-428.
- Price G.B., Allarakhia M., Cossons N., Nielsen T., Diaz-Perez M., Friedlander P., Tao L. & Zannis-Hadjopoulos M. (2003) Identification of a cis-element that determines autonomous DNA replication in eukaryotic cells. *Journal of Biological Chemistry*. **278(22)**:19649-59.
- Reeves R., Gorman C.M. & Howard B. (1985) Minichromosome assembly of non-integrated plasmid DNA transfected into mammalian cells. *Nucleic Acids Research*, **13**: 3599-3615.
- Remus R., Kammer C., Heller H., Schmitz B., Schell G. & Doerfler W. (1999) Insertion of foreign DNA into an established mammalian genome can alter the methylation of cellular DNA sequences. *Journal of Virology*, **73(2)**:1010-22.

- Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) Molecular cloning: a laboratory manual, 2nd edition. Cold Springs Harbour Laboratory Press, Plainsview NY.
- Schaarschmidt D., Baltin J., Stehle I.M., Lipps H.J. and Knippers R. (2004) An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *The European Molecular Biology Organization Journal*, **23**: 191-201.
- Schaffer D.V. and Laufenburger D.A. (1998) Optimization of Cell Surface Binding Enhances Efficiency and Specificity of Molecular Conjugate Gene Delivery. *The Journal of Biological Chemistry*, **273(43)**: 28004-28009.
- Schubeler D., Mielke C., Maass K. and Bode J. (1996) Scaffold/Matrix-Attached Regions Act upon Transcription in a Context-Dependent Manner. *Biochemistry*, **35**: 11160-11169.
- Scimmenti C.R. and Calos M.P. (1998) Epstein-Barr virus vectors for gene expression and transfer. *Current Opinion in Biotechnology*, **9**: 476-479.
- Shimizu N., Miura Y., Sakamoto Y. and Tsutsui K. (2001) Plasmids with a Mammalian Replication Origin and a Matrix Attachment Region Initiate the Event Similar to Gene Amplification. *Cancer Research*, **61**: 6987-6990.
- Siebenkolten G., Leyendeckers H., Christine R. and Radbruch A. (1995) Isolation of plasmid DNA from mammalian cells with QIAprep. *QIAGEN News*, **2**: 11-12.
- Singh G.B., Kramer J.A. & Krawetz S.A. (1997) Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Research*, **25(7)**: 1419-1425.
- Stehle I.M., Scinteie M.F., Baiker A., Jenke A.C.W. and Lipps H.J. (2003) Exploiting a minimal system to study the epigenetic control of DNA replication: the interplay between transcription and replication. *Chromosome Research*, **11**: 413-421.
- Stenlund A. (2003) Initiation of DNA Replication: Lessons From Viral Initiator Proteins. *Nature Reviews Molecular Cell Biology*, **4**: 777-785.

- Stoll S.M. and Calos M.P. (2002) Extrachromosomal plasmid vectors for gene therapy. *Current Opinion in Molecular Therapeutics*, **4(4)**: 299-305.
- Stothard P. (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* 28: 1102-1104.
- Thomas C.E., Ehrhardt A. and Kay M.A. (2003) Progress and Problems with the use of viral vectors for gene therapy. *Nature Reviews Genetics*, **4**: 346-358.
- Thomas M. and Klivanov A.M. (2003) Non-viral gene therapy: polycation-mediated DNA delivery. *Applied Microbiological Biotechnology*, **62**: 27-34.
- Thompson C.T. and Dvorak J.A. (1989) Quantitation of total DNA per cell in an exponentially growing population using the diphenylamine reaction and flow cytometry. **177(2)**:353-7.
- Tolmachova T., Simpson K. and Huxley C. (1999) Analysis of YAC with human telomeres and *oriP* from Epstein-Barr Virus in yeast and 293 cells. *Nucleic Acids Research*, **27(18)**: 3736-3744.
- Van Craenenbroeck K., Vanhoenacker P. & Haegeman G (2000) Episomal vectors for gene expression in mammalian cells. *European Journal of Biochemistry*, **267**: 5665-5678.
- van Driel R. & Otte A.P (1997) Nuclear Organization, Chromatin Structure, and Gene Expression (1st ed.), *Nuclear matrix-associated regions*, Amsterdam, E.C. Slater Instituut, University of Amsterdam.
- Vashee S., Cvetic C., Lu W., Simancek P., Kelly T.J. & Walter J.C. (2003) Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes & Development*. 2003 Aug 1;17(15):1894-908.
- Verma I.M. & Weitzman M.D. (2005) Gene Therapy: Twenty-First Century Medicine. *Annual Reviews of Biochemistry*, **74**: 711-38.

- Wade-Martines R., White R.E., Kimura H., Cook P.R. & James M.R. (2000) Stable correction of a genetic deficiency in human cells by an episome carrying a 115 kb genomic transgene. *Nature Biotechnology*, **18**:1311-1314.
- Wendl MC, Marra MA, Hillier LW, Chinwalla AT, Wilson RK, and Waterston RH (2001) Theories and Applications for Sequencing Randomly Selected Clones. *Genome Research*, **11**(2): 274-280.
- Wilson J.M., Grossman M., Cabrera J.A., Wu C.H. and Wu G.Y (1992) A Novel Mechanism for Achieving Transgene Persistence *in Vivo* after Somatic Gene Transfer into Hepatocytes. *The Journal of Biological Chemistry*, **267**(16): 11483-11489.
- Woodcock, D.M., Cowther, P.J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S.S., Michael, M.Z., Graham, M.W. (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucl. Acid Res.* **17**: 3469-3478.
- Yallop C.A. and Svendsen I. (2001) The effects of G418 on the growth and metabolism of recombinant mammalian cell lines. *Cytotechnology*, **35**: 101-114.
- Zar J.H., editor. 1984 Biostatistical Analysis. 2nd ed. New Jersey: Prentice-Hall. 100-140p.

