Analysis Of S/MAR Vectors For Gene Therapy In Muscle

A Thesis Submitted For The Degree Of Doctor Of Philosophy
In The University of London

April 2011

By SAMAH FAKHRO

Centre for Biomedical Sciences, School of Biological Sciences, Royal Holloway, University of London

Declaration of Authorship

I, Samah Fakhro, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed: Samah Fakhro

Date: 1 April 2011

Abstract

Muscular Dystrophy (MD) is a progressive muscle wasting disease which currently has no cure, and is caused by the mutation of the dystrophin gene. A multitude of approaches for the improvement of the muscular pathology caused by this condition are being investigated, one of which is gene therapy. This approach is used to deliver vectors containing therapeutic transgenes such as dystrophin to target muscle cells. One method of gene delivery utilises viral vectors, and although this has resulted in systemic delivery and efficient transgene expression, there are many safety implications which have led to the development of non-viral approaches, such as the direct delivery of naked plasmid DNA. However, the shortcomings of these vectors include an inability to replicate within host cells, resulting in the loss of vector as cells replicate, and the silencing of transgene expression. In an effort to overcome such limitations, a novel system called the 'pEPI vector' has been developed. Here, the inclusion of the β-IFN scaffold/matrix-attachment region (S/MAR) element into the open reading frame of an actively transcribed transgene has been found to lead to sustained, long term transgene expression, and to allow the episomal propagation and maintenance of the vector in dividing cells over many generations. The aim of this thesis was to investigate the potential of this vector for use as a gene therapy vector in muscle cells in order to treat MD.

In this study, the long-term expression of the eGFP reporter transgene inserted into the pEPI vector was evaluated, and the pEPI vector's episomal/integrant status was investigated, in C2C12 murine myoblasts, HeLa, and HepG2 cell lines. 60 days after transfection the vector was found not to have integrated into the host genomes of any of the cell lines. Transgene expression had declined to nearly undetectable levels in fast-replicating C2C12 and HeLa cells, but was at high levels in the relatively slow-dividing HepG2 cells. An attempt to improve long term transgene expression in C2C12 cells by changing the promoter from CMV to CAGG still led to low transgene expression after 60 days.

To address this issue, this study focused on the development of a novel approach to ameliorate long-term transgene expression, based upon the origin of replication and nuclear matrix attachment properties of the S/MAR element, as well as the results obtained from testing the vector in the HeLa and HepG2 cells. It involved the arrest of C2C12 cells in the G0/G1 phase of the cell cycle post transfection with the pEPI vector in order to allow these fast-dividing cells an extended period of time to epigenetically mark pEPI prior to selection. The findings indicated that this novel method of pEPI vector establishment was superior to that which utilises selection alone. However, in spite of the improvement in long-term episomal transgene expression observed using this novel method of establishment of the pEPI vector, transgene expression levels were still relatively low after 35 days of cell proliferation, which led to the conclusion that further development of this vector is essential in order for it to be able to elicit a significant restoration of muscle function in MD patients.

Additionally, two other S/MAR vectors were tested in C2C12 cells. One vector contained the 'mini-S/MAR', a shorter version of the β-IFN S/MAR, and the other contained a novel S/MAR derived from the *c-myc* proto-oncogene. Transgene expression by either vector was nearly undetectable after several weeks of proliferation, and both were found to integrate into the C2C12 host genome, leading to the conclusion that not all S/MAR elements inserted within a plasmid vector can lead to long-term transgene expression, nor confer protection from vector integration.

Acknowledgements

A big thank you to my parents, without whom none of this would ever be possible. I'd also like to thank my supervisor Jon Beauchamp for all his support and guidance and for sharing his knowledge and expertise, and to Professor Peter Bramley and Professor George Dickson. Through all the great times, hard times, and insane times, thank you to friends Hui Low (and her family for long-distance support!) and Taeyoung Koo, who's never ending faith made every step of the way lighter, easier, happier, and full of laughter. A big thank you to the brilliant friends who always helped with a big smile, Linda Popplewell, Takis Athanasopoulus, Capucine Trollet, Ian Graham, Rob from Stores, Alberto Malerba, Julain Harris, Safina Khan, and Laslo Bogre. A special thank you goes to Pavlos Alifragis and Alessandra Devoto, I have not met kinder people. Thank you to to Adam Abadi for making everything lighter. An exclusive and immense thank you to Dalal Al Safar, who never let me give up.

To my father, the greatest inspiration of all, who always shows me the light, I dedicate this thesis to you.

Glossary of Abbreviations

2'OMe 2'-O-Methoxyethyl
AAV Adeno-Associated Virus
Ac Acetylation
ADP Adenosine Diphosphate
ATF Activating Transcription Factor
ATP Adenosine Triphosphate
BrdU Bromodeoxyuridine
BUR Base-Unpairing Region
CAF-1 Chaperone Chromatin Assembly Factor
CAGG CMV enhancer/chicken β-actin fusion promoter
CaM Calmodulin
CBP CREB Binding Protein
cc coiled coil
CMV Cytomegalovirus
c-myc c-myelocytomatosis oncogene
CREB cAMP-Response Element-Binding
CUE Core-Unparining Element
DAPC Dystrophin Associated Protein Complex
DMD Duchenne Muscular Dystrophy
DNA Deoxyribonucleic Acid
EF-1 α Elongation Factor-1 α
FACS Fluorescent Activated Cell Sorting
FCS Foetal Calf Serum
FISH Fluorescent In-Situ Hybridisation
G0 Growth phase 0/quiescence
G1/2 Growth phase 1/2
HAT Histone Acetyltransferase
HDAC Histone Deacetylase
HMG High Mobility Group
Kb Kilobase
LNA Locked Nucleic Acids
M Mitosis
MBD1 Methyl CpG-Binding protein
MBF-1 Multiprotein Bridging Factor 1
MCM Minichromosome Maintenance proteins
MD Muscular dystrophy
Me Methylation
Myf-5 Myogenic Factor 5
MyoD Myoblast Determination Protein
NFκB Nuclear Factor Kappa B
nNOS neuronal Nitric Oxide Synthase
NuMA Nuclear Mitottic Apparatus protein
ORC Origin Recognition Complex
Ori Origin ofreplication
P Phosphorylation
PCNA Proliferating Cell Nuclear Antigen
PCR Polymerase Chain Reaction
pDNA plasmid DNA
PMO Phosphorodiamidate Morpholino
PNA Pepetide Nucleic Acids
Pre-RC Pre-Recognition Complex
rAAV recombinant Adeno-Associated Virus
RNA Ribonucleic Acid
S Synthesis phase
S/MAR Scaffold/Matrix Attachment Region
SAF Scaffold Attachment Factor
SATB1 Special AT-rich Binding Protein
SDF-1 Stromal Cell-Derived Factor-1
SIN Self-Inactivating
SP Side Population
TLR9 Toll-Like Receptor 9
TNF-α Tumor Necrosis Factor α

TABLE OF CONTENTS

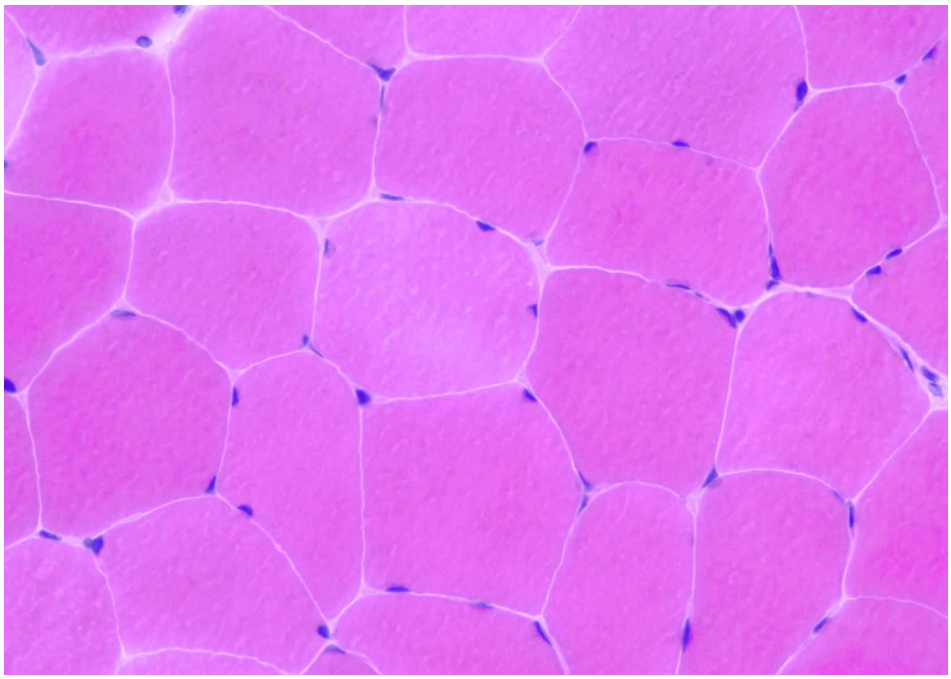
List Of Figures & Tables

Chapter 1: Introduction

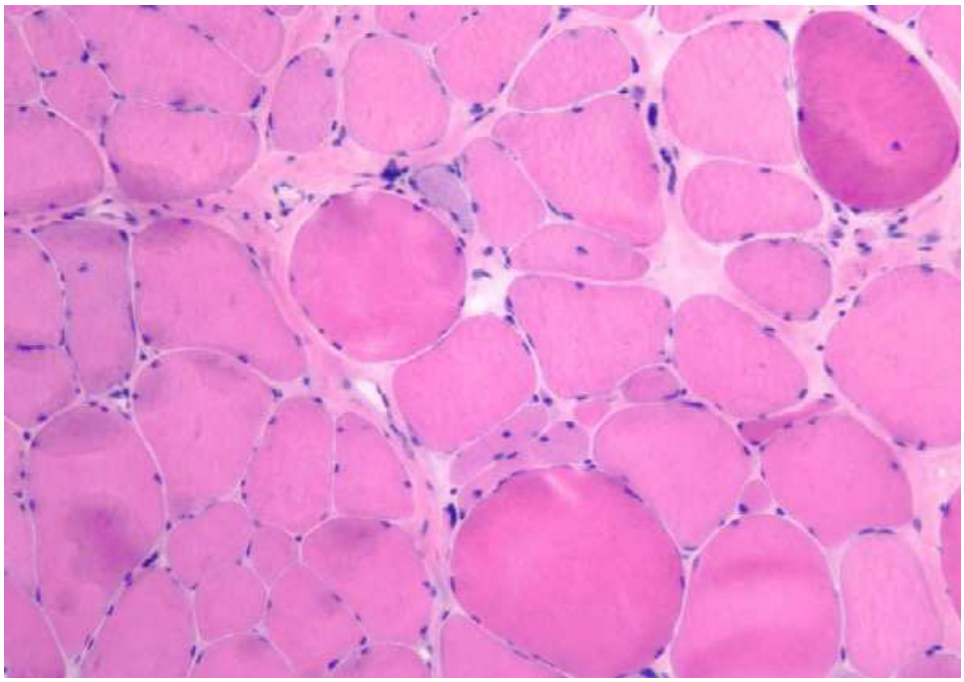
1.1 Muscular Dystrophy

Duchenne muscular dystrophy (DMD) is an X-linked recessive genetic disease whereby the dystrophin gene is mutated. This disease can also occur as a result of spontaneous mutations seen in 1/10000 gametes (Nowak and Davies, 2004). There are over 40 types of muscular dystrophy (Muir and Chamberlain, 2009) that occur as a result of different mutations occurring in dystrophin or other associated proteins (Blake *et al*., 2002), which include Duchenne, Becker, Limb- girdle, Congenital, and Oculopharyngeal muscular dystrophies, amongst others (Trollet *et al*., 2009). An estimated 60% of these mutations are a result of large insertions or deletions, and 40% from point mutations (Hoffman and Dressman, 2001). Dystrophin is a subsarcolemmal protein found in muscle that is required for the dystrophin-glycoprotein transmembrane complex, essential for the prevention of exercise-induced muscle damage (Nowak and Davies, 2004). The more severely the protein is mutated, the more susceptible the sarcolemma is to damage, of which repeated cycles of damage/regeneration leads to tissue necrosis and myofibre replacement with fat and connective tissue (Grounds and Davies, 2007) (Figure 1.1). The severity of this type of muscular dystrophy is due to mutations which lead to out of frame transcription resulting in a dystrophin protein that is non-functional.

(A)



(B)



(Seidman, RJ, 2006)

Figure 1.1 Adapted from Seidman, RJ, 2006. Hematoxylin and eosin (H&E) staining of frozen cross-section of normal muscle (A) indicating similar sized fibres (pink), with thin endomysial connective tissue between the fibres (white), and the nuclei present at the peripheries (purple). Image (B) shows H&E staining of a frozen cross-section muscle from a patient with muscular dystrophy and indicates an increase in fibre size variability, atrophied rounded fibres, and an increase in the amount of endomysial connective tissue between and around the fibres denoting fibrosis.

This disease affects 1/3500 young boys and eventually claims their lives due to cardiac or respiratory failure (Grounds and Davies, 2007). At birth these boys appear normal, however abnormally high levels of the enzyme creatine kinase can be detected in the bloodsteam, indicating muscle damage (Blake *et al*., 2002). The first symptons are usually evident between the ages of 2 and 5 (Dubowitz *et al*., 1978; De Matteis and Morrow, 2000), which include a 'waddling gait', pseudohypertrophy of the calf muscles, Gower's sign where the use of the arms is required to stand up from a lying position, and a general weakness in the limbs (Sakuta, 2009). Dystrophin has also been seen to be expressed in the brain. While its mutation can lead to abnormal brain function in some patients (Blake and Kroger, 2000; Mehler, 2000), other patients have been found to be of normal, or above normal, intelligence. By their early teenage years the boys are wheelchair-bound, and most die by their early 20s.

In the case of Becker's muscular dystrophy, a mutation occurs without disrupting the reading frame, where a semi-functional truncated protein is created. The disease is far less severe or debilitating than Duchenne's (Nowak and Davies, 2004), and a later onset of pathology is observed in addition to longer survival time (Blake *et al*., 2002). Over 90% of those affected are still alive by their early 20s, and many can live to remain mobile until an old age (reviewed in Blake *et al*., 2002).

To fully understand the disease in order to amerliorate pathologies and find treatments, it is essential to study the molecular mechanisms behind muscle function, in addition to genetic mutations which lead to impaired function thus leading to the onset of muscular dystrophy.

1.2 Skeletal Muscle Function

1.2.1 General Muscle Structure

Healthy muscle is composed of fascicles, which are each composed of bundles of fibres. Each fibre is enveloped in a membrane called the sarcolemma, is multinucleated, and is formed by the fusion of many myoblasts with one another. These nuclei are found directly beneath the sarcolemma. Within the cytoplasm of these fibres is a high concentration of mitochondria, present in order to supply energy, and myofibrils, which contain the contractile components of muscle. Each myofibril is

surrounded by a membrane called the sarcoplasmic reticulum, and the components that make up each myofibril are thick filaments composed of myosin, and thin filaments made of actin. In skeletal muscle, these filaments run parallel to the long axis of the fibre and are found in repeating units along each myofibril called sarcomeres (Germann and Stanfield, 2001).

Myosin is a molecule of two subunits, each with a long tail and a globular head, both intertwined with one another. It has an actin-binding site located on the globular head, in addition to an ATPase site. Thick filaments are composed of many of these myosin filaments that are bound at the tail end, with the head ends jutting out in opposite directions, arranged in a staggered pattern (Germann and Stanfield, 2001).

Actin is made up of monomeric globular protein arranged as two helical polymer strands. On each monomer is a myosin-binding site. In close proximity to the actin and within the thin filaments are two other proteins essential for muscle contraction: tropomyosin and troponin. Tropomyosin is a long protein that extends over a number of actin molecules and blocks their myosin-binding sites. Troponin is composed of three proteins, each with a specific function. The first is bound to actin, the second to tropomyosin, and the third contains a calcium binding site. When calcium binds at this site, troponin undergoes a conformational change which allows it to move the tropomyosin from its resting position to expose the myosin binding sites on the actin (Germann and Stanfield, 2001).

Contraction occurs through an interplay between all of these proteins following an influx of calcium into the cytosol. Muscle receives an impulse, or an action potential, from nerves at the neuromuscular junction. This impulse is in the form of the chemical acetylcholine released from the motor neuron. Receptors at the motor end plate region of the sarcolemma bind the chemical, which in turn depolarises the membrane. This depolarisation triggers a release of calcium ions from the sarcoplasmic reticulum where they are stored, through voltage-gated channels, and into the cytosol. When calcium enters it binds to others channels triggering the release of more calcium ions. This increase in calcium concentration leads to the binding of calcium to troponin, which moves tropomyosin, allowing actin and myosin to bind. They bind and allow muscle contraction by what is called a ‘crossbridge cycle’(Germann and Stanfield, 2001).

An influx of calcium leads troponin to displace tropomyosin and expose actin's myosin-binding sites. ATP bound to the ATPase site on the myosin head is hydrolysed into ADP and P_i where the energy released leads to a myosin conformational change, allowing it to bind to actin. The P_i is then released which makes myosin revert back to its previous conformation, pulling the thin filament along with it and releasing the ADP. This pulling action between actin and myosin in every sarcomere along the fibres causes a contraction. Finally, ATP binds to the ATPase site on the myosin head and triggers another conformational change which leads to the detachment of myosin from the actin.

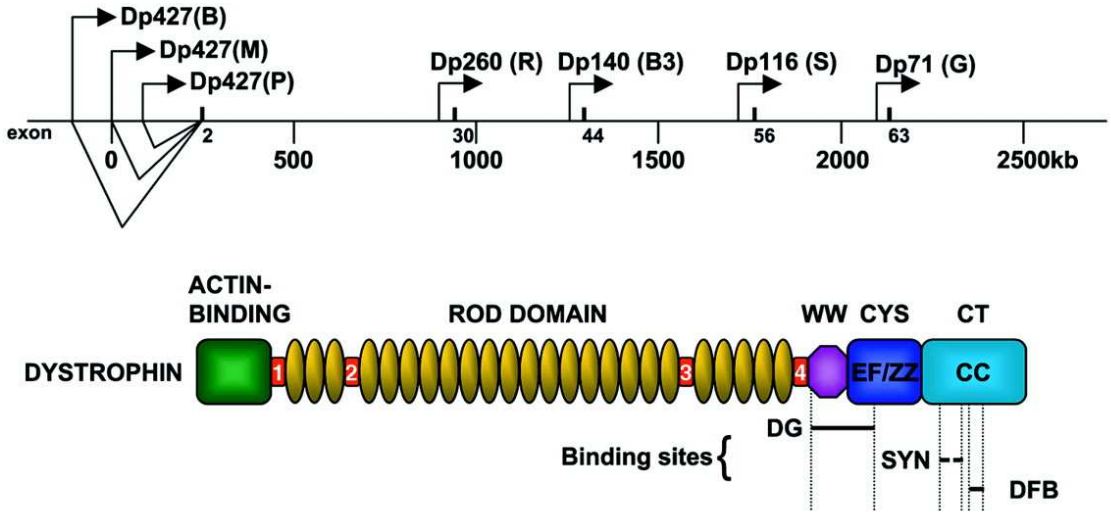
Once the calcium levels are high they bind to sites on the SR channels to which they have a lower affinity to than those sites that lead the channels to open. This closes the gates and the calcium is then actively transported back out of the cytosol. The decrease in calcium leads it to dissociate from troponin where the myosin-binding sites on actin are once again blocked (Germann and Stanfield, 2001).

1.2.2 The Role of Dystrophin Protein In Skeletal Muscle Function

There are many other proteins that are involved in skeletal muscle contraction besides actin and myosin that perform crucial structural and/or signalling roles. Dystrophin is one such protein, which interacts with a series of proteins that form a complex and are collectively named the Dystrophin-Associated Protein Complex (DAPC). Dystrophin is bound to the actin-based cytoskeleton within the muscle fibre, then spans the membrane and binds the sarcolemma via the DAPC complex. The dystrophin gene is the largest gene in the genome, composed of approximately 2.5Mb, or 79 exons, and has a molecular weight of 427 kDa (Blake *et al*., 2002). Its expression is controlled by three different promoters, each named according to the location in which it is active: B promoter expresses mostly in the brain (Barnea *et al*., 1990; Chelly *et al*., 1990; Gorecki *et al*., 1992), P promoter in Purkinje cell and skeletal muscle (Gorecki *et al*., 1992; Holder *et al*., 1996), and M promoter in skeletal and cardiac muscle (Barnea *et al*., 1990; Chelly *et al*., 1990). The dystrophin gene also has four internal promoters that initiate the expression of several isoforms of the protein by unique exon splicing mechanisms that leave the protein truncated at the COOH terminal (Blake *et al*., 2002). Again, the promoters' names are based on their locations of expression: R promoter

expresses in the retina, B3 expresses in brain-3 cells, S in the Schwann cells, and G expresses generally in several tissue types (Figure 1.2). More alternative splicing also occurs in these truncated versions at the 3' ends of the transcripts (Bies *et al*., 1992; Feener *et al*., 1989). Although it is unclear what the functions are of these isoforms it is believed they may bind to dystrophin-like complexes found outside muscle tissue, or that they may regulate the binding of dystrophin to its associated protein complex. In skeletal muscle, the dystrophin protein resulting from the translation of the full length transcript is 427kDa and is a cytoskeletal protein that has a structural function in muscle contraction (Blake *et al*., 2002). It is composed of four different regions (Acsadi *et al*., 1991) which are: the NH_2 terminal, the central rod domain, the cysteine-rich domain, and the COOH-terminal domain (Figure 1.2). As a member of the β-spectrin/α-actinin family (Koenig *et al*., 1988), dystrophin can bind actin at its NH_2 terminal. After this structure lies the central rod domain which is composed of a series of 24 spectrin-like repeats. Spectrin repeats are structures of three-helix bundles assembled as antiparallel dimers that allow it to interact with many different proteins (Djinovic-Carugo *et al*, 2002). These α-helical coiled coil repeats make up most of the protein and give it flexibility and a rod shape similar to that of β-spectrin (Koenig and Kunkel, 1990). There are four proline-rich regions within the rod domain that act as hinges. The fourth hinge is located at the end of the last spectrin-like repeat, after which the WW domain can be found. This domain binds to proline-rich ligands (Macias *et al*., 1996) and it is believed to interact with the cytoplasmic region of a protein called β-dystroglycan which is very proline rich. The third domain which is the cysteine-rich domain has two types of motifs that define its function, which is believed namely to be its binding to the protein calmodulin and allowing its interaction with other protein members associated with dystrophin. The first is an EF-hand motif, which is where two helices E and F are bound by a loop that can directly bind intracellular calcium (Koenig *et al*., 1988). The second is a ZZ domain composed of a calcium-dependent zinc finger motif that binds to the protein calmodulin (Anderson *et al*., 1996). Calmodulin (CaM) is an important calcium-binding protein that mediates calcium binding for proteins that are unable to do so themselves and is involved in muscle contraction. Once bound to calcium it changes its conformation and binds to other proteins in order to elicit a specific response (www.ebi.ac.uk). The fourth and final domain of dystrophin, the COOH terminus, is believed to be made of α-helical coiled coils such as those found in

the rod domain (Blake *et al*., 1995), each composed of a 'conserved repeating heptad' called the CC (coiled coil) domain. This is a region that can bind to other proteins such as dystrobrevin, and is thought to mediate interaction between members of the DAPC and syntrophin (Blake *et al*., 1995; Sadoulet-Puccio *et al*., 1997).



(Davies *et al*, 2002)

Figure 1.2 Adapted from Davies *et al*, 2002. The dystrophin gene is the largest gene in the genome, composed of approximately 2.5Mb, or 79 exons, and has a molecular weight of 427 kDa. Its expression is controlled by three different promoters, each named according to the location in which it is active and the protein molecular weight (kDa): Dp427(B) promoter expresses mostly in the brain, Dp427(P) promoter in Purkinje cell and skeletal muscle, and Dp427(M) promoter in skeletal and cardiac muscle, as indicated in the top image of the figure above.The dystrophin gene also has four internal promoters that initiate the expression of several isoforms of the protein by unique exon splicing mechanisms that leave the protein truncated at the COOH terminal. Again, the promoters' names are based on their locations of expression and protein molecular weight (kDa): Dp260(R) promoter expresses in the retina, Dp140(B3) expresses in brain-3 cells, Dp116(S) in the Schwann cells, and Dp71(G) expresses generally in several tissue types. More alternative splicing also occurs in these truncated versions at the 3' ends of the transcripts

The lower figure indicates the structure of the dystrophin protein which includes the actin-binding domain at the NH_2 terminal (in green), the rod domain composed of 24 spectrin-like repeats giving the protein its rod shape and flexibility (in yellow), the four proline-rich hinges (1-4 in red), the WW domain (purple) which binds proline-rich ligands such as the cytoplasmic region of β-dystroglycan (DG), the cysteine-rich (CYS) domain (in dark blue) composed of an EF-hand motif (EF) that can bind intracellular calcium, and a calcium-dependent zinc finger motif (ZZ) that binds to the protein calmodulin , and the COOH terminus (CT) (light blue), made of α-helical coiled coils (CC), binds dystrobrevin (DFB), and may mediate interaction between members of the DAPC and syntrophin (SYN).

1.2.3 The Role of Dystrophin-Associated Protein Complex (DAPC) In Skeletal Muscle Function

The DAPC complex shows the interrelated nature of protein-protein interaction in muscle contraction. It is a complex that allows dystrophin to connect the inside of a myofibre to the outside. Yoshida *et al* found that the complex can be separated into three entities: the dystroglycan complex, the sarcoglycan-sarcospan complex, and the cytoplasmic dystrophin-containing complex (Yoshida *et al*., 1994) (Figure 1.3).

The dystroglycan complex is believed to be involved in organising the extracellular matrix (McDearmon *et al*., 1998; Pall *et al*., 1996). It is composed of several dystroglycan isoforms that are produced by the action of a protease (Ibraghimov-Beskrovnaya *et al*., 1993), which are different molecular weights in different tissues. In muscle the α-dystroglycan protein is found in the extracellular matrix and is believed to bind β-dystroglycan (Figure 1.3), which is a transmembrane protein with its COOH terminus located within the cytoplasm. This terminus is proline-rich and is the region that binds to the cysteine-rich region of dystrophin (James *et al*., 2000) at the WW domain and the EF hands region (Huang *et al*., 2000). β-dystroglycan is also capable of binding another transmembrane protein whose function it is to recruit signalling molecules and is called Caveolin-3 (Sotgia *et al*., 2000). Like dystrophin it has a WW domain with which it competes with dystrophin at the COOH terminal of β-dystroglycan. β-dystroglycan also binds other proteins such as laminins, agrins, and perlecan (Hohenester *et al*., 1999).

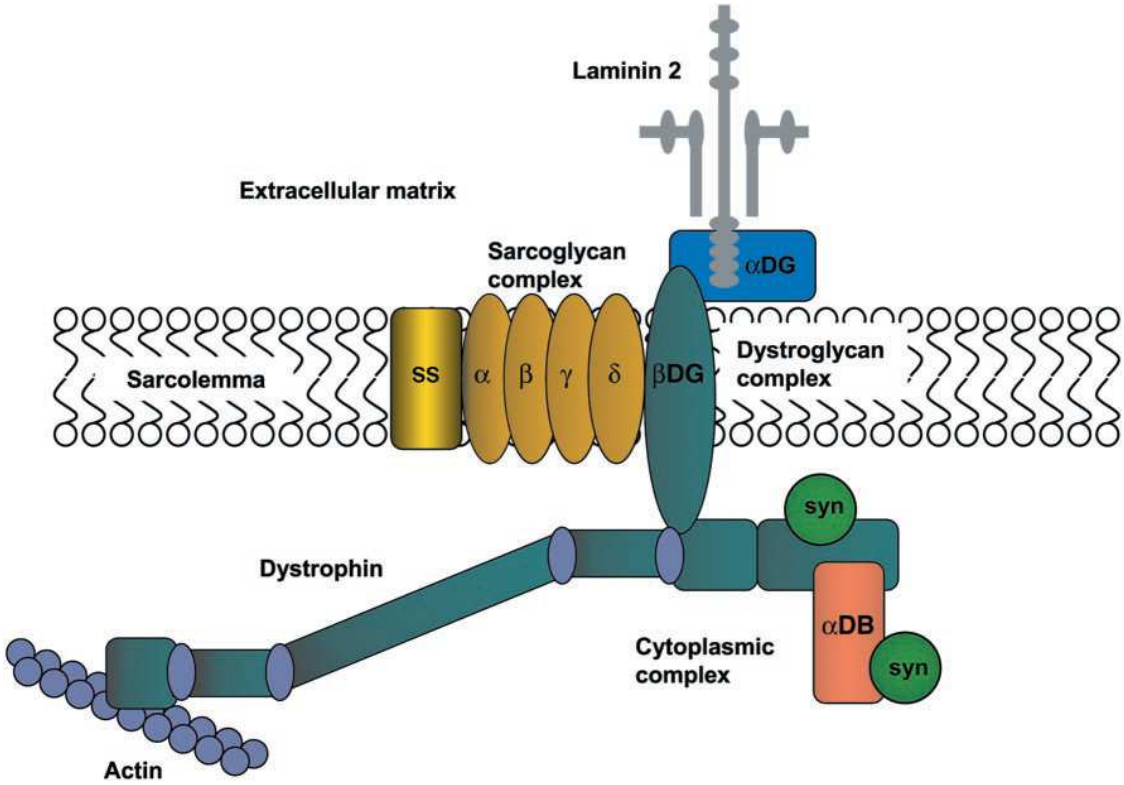
The sarcoglycan-sarcospan complex anchors the dystroglycan complex (Figure 1.3) and is essential for its stability within the sarcolemma (Blake *et al*., 2002) and may also mediate signalling functions. The complex is made up of α, β, γ, and δ- sarcoglycans which are all transmembrane glycoproteins, as well as the protein sarcospan (Crosbie *et al*., 1997; Lim and Campbell, 1998). α-sarcoglycan is found only in skeletal muscle and is a Type I membrane protein which is a transmembrane protein whose COOH terminus is located in the cytoplasmic side. ε-sarcoglycan is the same as α-sarcoglycan and replaces it in smooth muscle (Barresi *et al*., 2000; Roberds *et al*., 1993; Straub *et al*., 1999). β, γ, and δ- sarcoglycans are found in skeletal, smooth, and cardiac muscle and are Type II membrane proteins (Barresi *et al*., 2000; Roberds *et al*., 1993; Straub *et*

al., 1999), which are transmembrane proteins whose COOH terminals are in the extracellular matrix rather than the cytoplasm. It has been shown by Chan *et al* that β, γ, and δ- sarcoglycans are closely associated and that δ- sarcoglycan binds dystroglycan (Chan *et al*., 1998). γ and δ- sarcoglycan are believed to be paralogs and have been found to interact with a protein that is involved in transducing signals as well as reorganising actin filaments, called filmin-2 (Thompson *et al*., 2000). This interaction is evidence that the DAPC has a role in signalling functions within muscle (reviewed in Blake *et al*., 2002).

The cytoplasmic dystrophin-containing complex is composed of the syntrophins and the dystrobrevins (Figure 1.3). The syntrophin family is composed of α, β_1, β_2, γ_1, and γ_2 –sarcoglycan (Adams *et al*., 1993; Bradley *et al*., 1972; Piluso *et al*., 2000). They all have a similar structure in that they have one split Pleckstrin Homology (PH) domain, one intact one, and one PSD95/DLG/ZO-1 (PDZ) domain. PH domains are defined by their affinity to ligand binding due to their electrostatic 'pocket' and their hydrophobic core that provides stability (www.ebi.ac.uk). PDZ domains are protein binding domains that can bind peptide sequences or COOH terminal sequences (www.ebi.ac.uk). The PH and PDZ domains in syntrophins allow them to bind to dystrophin as well as dystrobrevin at a similar location in both proteins, which is the first coiled coil in the COOH termini. There are two syntrophin binding sites in dystrophin and two in dystrobrevin (Newey *et al*., 2000), meaning four syntrophin molecules are bound to the DPC at once (Newey *et al*., 2000).

The other protein family dystrobrevin is believed to have a role in intracellular signalling (reviewed in Blake *et al*., 2002). The family is composed of at least five different isoforms (Blake *et al*., 1996; Sadoulet-Puccio *et al*., 1996), with the gene for α-dystrobrevin located in a different location to the β-dystrobrevin gene. The genes have different promoters for expression in different locations for each of the various isoforms (reviewed in (Blake *et al*., 2002). These proteins can be found at the neuromuscular junction and at the sarcolemma (Wagner, 2008) where they are believed to bind to dystrophin directly and to the sarcoglycan complex (Yoshida *et al*., 2000) which keeps them anchored to the DAPC (Figure 1.3). They also bind to three other proteins: dysbindin (Benson *et al*., 2001), syncoilin (Newey *et al*., 2001), and desmuslin (Mizuno *et al*., 2001). From their interaction with the predicted intermediate filament proteins syncoilin and desmuslin (Mizuno *et al*., 2001; Newey *et*

al., 2001) it is thought that the DAPC is linked to the cytoskeletal network within muscle (reviewed in Blake *et al*., 2002). The role of dysbindin has not been elucidated yet, and no protein binding domains have been found, therefore it is theorised that it is a protein that recruits other proteins to the DAPC (reviewed in Blake *et al*., 2002).



(Davies *et al*, 2002)

Figure 1.3 Adapted from Davies *et al* (2002). The DAPC is a complex that allows dystrophin to connect the inside of a myofibre to the outside. The complex can be separated into three entities: the dystroglycan complex, the sarcoglycan-sarcospan complex, and the cytoplasmic dystrophin-containing complex. The dystroglycan complex is composed of several dystroglycan isoforms and organises the extracellular matrix. The α-dystroglycan (αDG, in blue in the figure above) protein is found in the extracellular matrix and is believed to bind β-dystroglycan (βDG, in teal), which is a transmembrane protein with its COOH terminus located within the cytoplasm, bound to dystrophin (also in teal). β-dystroglycan is also capable of binding another transmembrane protein, Caveolin-3, and other proteins such as laminins, agrins, and perlecan. The sarcoglycan-sarcospan complex (in yellow) anchors the dystroglycan complex and is essential for its stability within the sarcolemma. The complex is made up of α, β, γ, and δ- sarcoglycans which are all transmembrane glycoproteins, as well as the protein sarcospan (SS). β, γ, and δ- sarcoglycans are closely associated and δ- sarcoglycan binds dystroglycan. The cytoplasmic dystrophin-containing complex is composed of the syntrophins (syn, in green) and the dystrobrevins (DB, in pink). The syntrophin family is composed of α, β_1, β_2, γ_1, and γ_2 –sarcoglycan. The PH and PDZ domains in syntrophins allow them to bind to dystrophin as well as dystrobrevin. Dystrobrevin is believed to have a role in intracellular signalling, and at least five different isoforms exist. These proteins can be found at the neuromuscular junction and at the sarcolemma where they bind to dystrophin directly and to the sarcoglycan complex, which keeps them anchored to the DAPC.

1.2.4 Utrophin Protein in Skeletal Muscle

Utrophin is a protein that has been found to be a paralog of dystrophin, and also plays a role in muscle function. The full length transcript encodes a 395kDa protein and its structure is believed to be similar to that of dystrophin (Tinsley *et al*., 1992). Its exons are also spread out over a large area of the genome (Pearce *et al*., 1993), and it is also controlled by several different promoters (Burton *et al*., 1999; Dennis *et al*., 1996) leading to the creation of several isoforms (Pearce *et al*., 1993) altered at the COOH terminus as a result of variable splicing (Blake *et al*., 1995; Lumeng *et al*., 1999; Wilson *et al*., 1999). Its structure at the COOH terminus suggests that it can bind to protein members of the DAPC such as β-dystroglycan (Matsumura *et al*., 1992), α-dystrobrevin-1 (Peters *et al*., 1998), the syntrophin proteins (Peters *et al*., 1998; Kramarcy *et al*., 1994), and also has an affinity for a section of the sarcoglycan complex (Matsumura *et al*., 1992). The NH_2 terminus has been found to contain an actin binding site similar to that in dystrophin (Tinsley *et al*., 1992; Winder *et al*., 1995) and can bind F-actin (Winder *et al*., 1995) which is the filamentous protein composing thin filaments in muscle, and β-actin (Moores and Kendrick-Jones, 2000; Morris *et al*., 1999; Winder *et al*., 1995) which is a cytoskeletal protein. Utrophin has been found in many different tissue types (reviewed in Blake *et al*., 2002) and in adult skeletal tissue it is mostly found in the vascular structures and associated with nerves (Khurana *et al*., 1991; Vater *et al*., 1998) and localised in the myotendinous and neuromuscular junctions (Nguyen *et al*., 1991; Ohlendieck *et al*., 1991). It is only found in the sarcolemma during embryonal development or after muscular injury and regeneration (Chelly *et al*., 1990; Gramolini *et al*., 1999) where the levels of utrophin are higher than in normal healthy muscle.

1.2.5 Therapeutic Approaches to Muscular Dystrophy

It is clear that there are many different proteins involved in the correct function of muscular contraction, and that their roles are highly interrelated. Therefore, it can be seen that a mutation in a gene encoding one of these factors can have a knock-on effect on the function of other related factors, thus affecting muscle performance, and leading, in the case of dystrophin mutations, to the onset of muscular dystrophy.

The treatment of muscular dystrophy is a challenging one where several factors must be considered. There are several different approaches to the treatment of the muscular dystrophies that have emerged such as gene therapy or cell therapy which have promising potential. The cell therapy approach includes the delivery or engraftment of allogeneic or treated autologous myogenic precursors with the ability to fuse with myofibres and generate new ones. The viral gene therapy approach utilises a virus such as Adeno-Associated Virus (AAV) or Lentivirus to deliver DNA to the muscle fibres where the therapeutic protein is subsequently expressed. The non-viral gene therapy approach involves the delivery of naked plasmid DNA, or the systematic delivery of oligonucleotides resulting in exon skipping, leading to in-frame, albeit truncated, transcript expression of the mutated gene. In order to describe the cell therapy treatments that have thus far been developed for muscular dystrophy, muscle stem cells, which are used for muscle cell therapy, will first be described. Following this, the general principles of viral and non-viral gene therapy will be covered, before reviewing the gene therapy approaches taken in the development of treatments for muscular dystrophy.

1.3 Muscle Stem Cells

1.3.1 Satellite Cells

Muscle fibres are post-mitotic tissue and therefore are unable to replicate. However, when injury occurs a mechanism must exist where the damaged tissue can be repaired or necrotic tissue can be replaced with healthy new fibres. These fibres are created by the fusion of myoblasts which are believed to originate from a precursor stem cell called the satellite cell. These cells are located in 'niches' just underneath the basal lamina of muscle fibres. They are classified as stem cells due to their ability to divide and give rise to differentiated cells as well as new satellite cells (Collins *et al*., 2005). After division, they can re-enter their state of quiescence until recruited once more. Collins *et al* locally irradiated muscle in mice and grafted a single fibre estimated to have 7 satellite cells (Collins *et al*., 2005). It was found that these cells were able to generate new muscle fibres and satellite cells as the number of new ones exceeded

the number originally transplanted. Blaveri *et al* and Heslop *et al* found that these precursor cells are even capable of occupying the satellite cell 'niche' (Blaveri *et al*., 1999; Heslop *et al*., 2001). It was proposed by Yablonka-Reuveni and Rivera that these satellite cells, once activated, commit to the myogenic lineage, and are replenished from another source (Yablonka-Reuveni and Rivera, 1994; Blaveri *et al*., 1999; Heslop *et al*., 2001). Several different sources have been proposed, such as endothelial-associated cells by (De Angelis *et al*., 1999), interstitial cells (Tamaki *et al*., 2002; Polesskaya *et al*., 2003; Kuang *et al*., 2006), or side population (SP) cells (Gussoni *et al*., 1999; Asakura *et al*., 2002). Primary cells isolated from the endothelium as well as cells isolated from vasculature such as mesoangioblasts were able to create myogenic progeny (Graves *et al*., 2000; Cusella De Angelis *et al*., 2003). Satellite cells and primary endothelial cells even have markers in common such as CD34 (De Angelis *et al*., 1999; Beauchamp *et al*., 2000) indicating the possibility of a common precursor cell (Kardon *et al*., 2002). SP cells are derived from bone marrow and can differentiate into the hemopoietic or myogenic lineage, and can even occupy the satellite cell niche on myofibres (Gussoni *et al*., 1999; Asakura *et al*., 2002).

Other stem cells may be able to contribute to muscle fibre growth or regeneration. A study by Sherwood *et al* indicated that some precursor cells that are able to create myogenic progeny are also able to occupy the satellite cell niche (Sherwood *et al*., 2004). However, the niche is unable to direct these cells to behave as satellite cells, nor to have the same characteristics (Sherwood *et al*., 2004). They are able to have myogenic precursors and fuse with myofibres as is the case with cells from bone marrow (LaBarge and Blau, 2002; Dreyfus *et al*., 2004), and do not have to go through a satellite cell step in order to do so (Grounds *et al*., 2002). But, they are unable to repopulate muscle tissue to any significant degree, even when the tissue is sustaining continuous damage as was observed in dystrophic mice (Ferrari *et al*., 2001). The contribution by these circulating precursor cells is low, and most of muscle fibre is generated from satellite cells (Zammit *et al*., 2002; Collins *et al*., 2005), which is evident through studies by Wakeford *et al* and Heslop *et al* where regeneration was found to be insignificant by these cells after muscle was irradiated and satellite cells destroyed (Wakeford *et al*., 1991; Heslop *et al*., 2000).

1.3.2 Satellite Cell Identifiable Markers and Transcription Factors During Quiescence, Activation, and Differentiation

Studies have uncovered several markers by which satellite cells can be identified or isolated. These include Pax7 (Seale *et al*., 2000), M-cadherin (Irintchev *et al*.,1994), saliomucin CD34 (Beauchamp *et al*., 2000), Myf5 (Tajbakhsh *et al*., 1997; Beauchamp *et al*.,2000; Shefer *et al*.,2006), lysenin which binds to sphingomyelin found on the cell membrane (Nagata *et al*.,2006), and caveolin I (Volonte *et al*.,2005) (Figure 1.4). Markers such as CD34 are not exclusive to satellite cells but are useful as markers when identifying satellite cells on isolated myofibres (reviewed in Zammit, *et al*.,2006) (Figure 1.5).

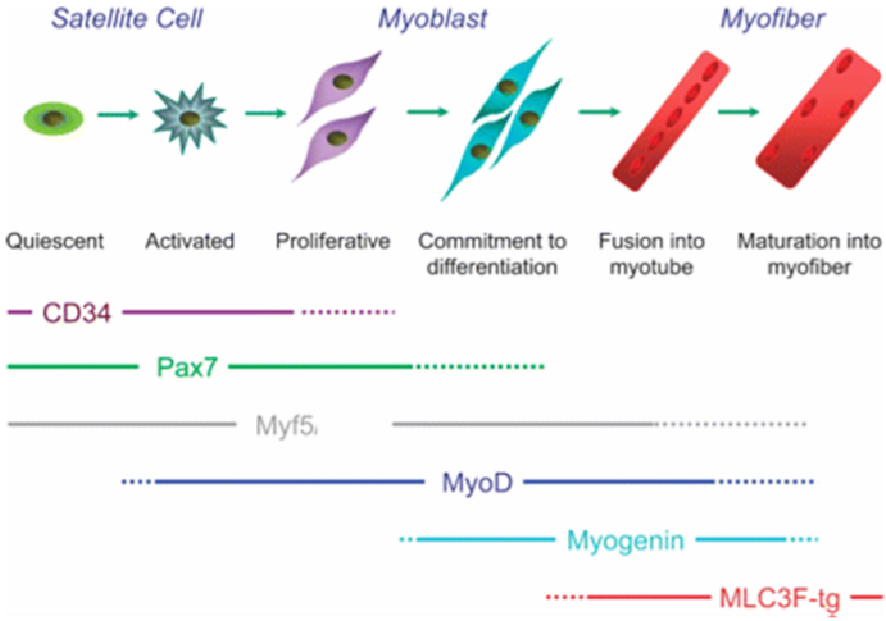
MyoD, Myf-5, and myogenin, amongst others, are transcription factors which are important regulators that activate the transcription of genes required for myogenesis to occur in muscle (Weintraub, 1993). Throughout quiescence, MyoD protein expression is downregulated whereas Myf-5 and all the markers mentioned above are expressed (Yoshida *et al*., 1998; Kitzmann *et al*.,1998).

The expression of certain genes in satellite cells during quiescence changes once these cells are activated (Figure 1.4). It is believed that activated satellite cells divide asymmetrically (Zammit *et al*., 2004) where the fate of each cell diverges. Before the cell divides, Pax7 and MyoD are co-expressed. One of the daughter cells suppresses MyoD, expresses Pax7, and goes into quiescence (Zammit *et al*., 2004) (Figure 1.5). The Notch signalling pathway also exerts a form of control where it directs the satellite cell to stop cycling (Conboy *et al*., 2002). Shinin *et al* found that when satellite cells divide asymmetrically Numb, a Notch inhibitor, segregates into the cells that will keep cycling (Shinin *et al*., 2006).

In those cells that have been activated, MyoD expression is rapidly upregulated (Fuchtbauer and Westphal, 1992; Grounds *et al*., 1992; Yablonka-Reuveni and Rivera, 1994) and the expression of Myf-5 decreases as the cells re-enter the cell cycle (Yoshida *et al*., 1998; Kitzmann *et al*., 1998). There is a change in the CD34 isoform expressed, and sphingomyelin amounts decrease (Nagata *et al*., 2006). After 24hrs in cell culture, the cells are then able to co-express MyoD and Myf-5. More genes that are characteristic of proliferating cells begin to be expressed such as PCNA (proliferating cell nuclear antigen) found in cells at the DNA synthesis, or S phase, stage

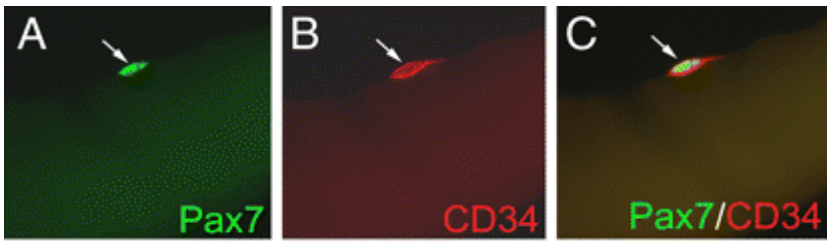
of the cell cycle. Other regulatory and structural genes expressed in myocytes are also switched on. Upon differentiation, the cells express myogenin, leading to the fusion of myoblasts to create myofibres (Andres and Walsh, 1996).

Immunostaining experiments by Kitzman *et al* revealed that the ratio of expression of MyoD and Myf-5 is highly affected, and is in fact regulated, by the cell cycle (Kitzmann *et al*., 1998). MyoD expression varies in proliferating myoblasts but is found to be evenly high in myotubes (Tapscott *et al*., 1988). MyoD expression peaks twice in the cell cycle, once in the middle of G1 phase and again at the end of G2. Myf-5 shows the opposite to MyoD, where it is maintained mostly in G0, S, mitosis, and G2 phases, where MyoD expression is minimal.



(Zammit *et al*, 2006)

Figure 1.4 Adapted from Zammit *et al*, 2006. This diagram represents myogenesis from satellite cell activation, through to proliferation, commitment, and finally fusion into myotubes and the formation of myofibres. It also indicates the markers that can be used to identify the cells at each stage, and the duration of which they are expressed.



(Yablonka-Reuveni *et al*, 2006)

Figure 1.5 Adapted from Yablonka-Reuveni *et al*, 2006. Image of muscle satellite cell stained for quiescent satellite cell markers Pax7 and CD34, located on a muscle fibre.

1.3.3 The C2C12 Cell Line As A Model For Quiescent Satellite Cells, Proliferating Myoblasts, And Myofibres

The C2C12 murine muscle cell line has been used in many studies as a model for muscle cells. An important feature which makes this cell line an excellent model is that some of the cells can be induced into quiescence, where they express many of the same markers as those which quiescent satellite cells express, such as Myf5 and CD34. The induction of C2C12 cells has been conducted by Kitzman *et al* by two different methods (Kitzmann *et al*., 1998).

The first method by which quiescent myoblasts were generated was to differentiate a culture for 3 days using low-serum medium. This resulted in a culture containing approximatly 60-70% of fused, differentiated myoblasts, now called myotubes, and 30-40% of myoblasts which had ceased to proliferate, but had not differentiated. These are called reserve cells. These quiescent cells could be separated from the myotubes and isolated by trypsinising with a low percentage of trypsin (Kitzmann *et al*.,1998). Myotubes are more sensitive to such treatment and detach from the dish first, leaving the undifferentiated quiescent myoblasts attached to the dish. These quiescent cells were found in both C2C12 differentiated cultures and in isolated primary mouse myoblasts put into culture. They could be made to re-enter the cell cycle by the addition of proliferation medium. The second method involved the use of low serum methionine depleted medium to induce myoblasts artificially into a reversible state of G0/G1 quiescence. Methionine is an amino acid that is vital for the growth of a cell, but not necessary for it to stay alive (Nadal-Ginard, 1978).

These quiescent cells, as well as reserve cells, were found to express the same factors and regulators during quiescence, proliferation, and differentiation, and were able to differentiate into myotubes just as myoblasts derived from adult skeletal muscle differentiate into myofibres (Yoshida *et al*., 1998; Kitzmann *et al*., 1998). An analysis of C2C12 cells' MyoD, Myf-5, and myogenin markers showed the same pattern of expression during quiescence, proliferation, and differentiation (Yoshida *et al*., 1998; Kitzmann *et al*., 1998; Andres and Walsh, 1996) where in G0 the cells expressed Myf-5 but not MyoD. In mid-G1 phase, the cells expressed MyoD and were able to

differentiate. During proliferation, MyoD and Myf-5 were co-expressed during S and G2 phases. Furthermore, upon differentiation, myogenin was found to be expressed.

1.4 Muscular Dystrophy Treatment Using Cell Therapy

Cell therapy involves the transplantation of stem cells into the body, derived from one of two ways: the first is from allogeneic precursor cells which requires immunosuppression of the patient and also runs the risk of graft rejection, and the second is the extraction of autologous precursors, their modification *ex vivo*, followed by their re-transplantation.

Stem cells derived from blood vessels are called mesoangioblasts and have been found to have the ability to migrate to damaged tissue and repopulate cells more efficiently than other types of stem cells (Sampaolesi *et al*., 2006). Galvez *et al* found that stem cells migrate to a specific location within the body based upon cytokines circulating in the target tissue as well as receptors expressed on the surface of the stem cells and endothelial cells which allow adhesion of the cells and their transmigration into muscle tissue (Galvez *et al*., 2006). In their study, Galvez *et al* pre-treated wild type mesoangioblast stem cells with cytokines Stromal cell-Derived Factor 1 (SDF-1) and Tumor Necrosis Factor α (TNF- α), which led to the expression of L-selectin and α4 integrin proteins (Galvez *et al*., 2006). These proteins, not normally expressed on mesoangioblast cell surfaces, dimerised with integrins already being expressed on these cells such as β1, β7, αL, and αm, which in turn allowed 15 times more endothelial adhesion and hence transmigration in vitro and over 5 times more *in vivo*, without affecting myogenic differentiation. Where a mere 10% of intraarterially injected untreated non- SDF-1/TNFα expressing mesoangioblasts managed to transmigrate to damaged muscule tissue in the control experiment, with the rest being presumably trapped in filter organs such as the liver or the lungs, an impressive 50% were now reaching targeted downstream muscle. This was confirmed by the detection of modified mesoangioblasts using immunohistochemistry within muscle fibers, which were also found as satellite cells.

Though this experiment shows that mesoangioblasts have enormous potential, there are of course a few parameters that need further investigation. For example, the excess of the cells injected that are trapped by filter organs, such as the lungs, may

have some negative effects on the organism that may as yet have not been detected, or may not manifest in mice but may become apparent in humans. Furthermore, mesoangioblasts don't have an infinite ability to divide and self-renew, which could be limiting in that dystrophic phenotypes may be reversed or prevented until these cells have exhausted their replicative potential and then another injection may be required. And in that light, Goetz *et al* had found that age was a significant factor in transmigration efficiency, due to altered microcirculation, as is disease progression, due to much of the muscle having been replaced by fat, leading to the absence of key cytokine excretion from damaged muscle tissue essential for mesoangioblast recruitment (Goetze *et al*., 2005). However, despite these potentially limiting factors, mesoangioblasts appear to offer the best alternative to viral vectors in terms of their safety, and can compete with their ability to significantly restore muscle integrity and function to muscle tissue throughout the body.

The second method using genetically modified autologous precursors does not run the risk of graft rejection and can be conducted without immunosuppression of the patient. However, there still remains the challenge of collecting enough precursor cells from patients at more progressive stages of the disease, transforming the cells *ex vivo*, and efficiently engrafting the cells back into the host after expansion (Skuk and Tremblay, 2003).

1.5 Gene Therapy

Gene therapy is defined as the delivery of a therapeutic gene into a cell in order to treat a disease. There are two main systems used for the delivery of such a gene: viral and non-viral. Viruses have been the prime candidates to be used as vectors as they exhibit high rates of transduction and offer general rather than localized delivery of the expressing transgene. Additionally, retroviruses such as MLV and lentiviruses are able to give sustained expression of the transgene by integrating their DNA into the host's genome (Schroder *et al*.,2002; Wu *et al*.,2003). Unfortunately, however, there are safety implications associated with the use of viruses which have not been so far surpassed thus limiting their potential as suitable candidates for gene therapy. For example, it has been found that retoviruses and lentiviruses preferentially integrate

their DNA into regions of their host's DNA that are actively being transcribed (Schroder *et al*.,2002; Wu *et al*., 2003). Also, viral proteins expressed that are essential for the propagation of virus and/or the replication of their DNA can be immunogenic and cytotoxic to the host, thus leading to inflammatory responses (Glover *et al*., 2005). Additionally, viral proteins are also capable of disrupting normal gene expression either by activating oncogenes or disrupting the expression of tumour suppressor genes (Glover *et al*., 2005), thus leading to cancer. As a result, some of the focus has been shifted to the development of non-viral vector alternatives.Non-viral vectors have several advantages over viruses, such as a lower ability for insertional mutagenesis. Another advantage is that they do not directly induce host aquired inflammatory responses (Glover *et al*., 2005), however they have been found to be able to stimulate the innate immune response which could, in itself, lead to an inflammatory reaction. There have been several types of non-viral approaches, such as Yeast Artificial Chromosomes (YACs), which helped define the requirements for the creation of a replicating chromosome which include a centromere, telomeres at both ends, and origins of replication (Murray and Szostak, 1983). YACs led to the development of Human Artificial Chromosomes (HACs) but have unfortunately not been very popular as vectors due to their large size, thus leading to inefficient *in vivo* delivery. Furthermore, instability as a result of concatamer formations and unforeseen recombination events leading to variation in sizes make it even less of a candidate. Another alternative is to send plasmid DNA into cells, as plasmids are smaller in size and are more stable. Plasmid DNA can be delivered to cells as naked DNA, or with the use of a carrier such as cationic lipids, polymers, or pepetides. The benefits to using naked plasmid DNA include the fact it allows a large capacity for transgene size which is not limited as is the case with packaging DNA into viral vectors (Gill *et al*., 2009). Furthermore, plasmid DNA can be constructed and produced cheaply and with relative ease in large amounts to a quality standard sufficient for *in vivo* delivery. This means that it does not contain bacterial contaminants such as DNA, RNA, endotoxins, or other bacterial proteins that could trigger an immunogenic reaction, and that the majority of the pDNA is supercoiled, a form that allows higher transfection efficiency into target cells (Gill *et al*., 2009). pDNA used in gene therapy is made of two components. The first is the expression cassette which includes the promoter, transgene, and polyadenylation signal. The second is the segment that contains the elements required

for its propagation and production in bacteria which includes a bacterial origin of replication and an antibiotic resistance marker.

Transgene expression resulting from the delivery of plasmid DNA does not always reach the threshold required for a significant functional improvement of a disease, thus efforts are being directed towards the improvement of pDNA delivery to, and expression in, target cells. When pDNA is sent into target tissue it must first be able to escape its delivery vehicle or carrier, if used, and then avoid cellular endosomes or cytosolic nucleases that degrade DNA before reaching the nucleus (Gill *et al*., 2009). A major drawback to using plasmid DNA is that its expression is often lost with time in dividing cells as it cannot replicate with the host DNA. Plasmids contain antibiotic resistance markers in order to maintain the plasmid within the host cell by using an initial period of selection for establishment. However, when selection is removed, the episome is gradually lost at a rate of 4% per cell generation (Glover *et al*., 2005). If the plasmid is delivered to non-dividing cells where it can remain episomal within the nucleus expression can still often be silenced with time for several different reasons, and this varies according to the target cell type (Gill *et al*., 2009).

Often when persistent expression of a pDNA transgene is observed it is due to an integration event occurring as a result of the use of an initial period of selection or due to highly recombinogenic sequences present within the plasmid. Integration of DNA into the host genome is unsafe as has already been mentioned. Attempts have been made at directing DNA integration into 'safe sites' (Glover *et al*., 2005) by recombination (Kapsa *et al*., 2001) but the frequency of such an event is too low and thus cannot be used for therapeutic purposes. Several Site-Specific Recombinases (SSRs) have been identified that can direct integration to a specific site within the genome. These act very efficiently, however further studies are still required in order to be certain that integration occurs strictly and solely at identified and known safe sites (Glover *et al*., 2005).

There are many factors that can influence the expression of a pDNA transgene and its persistence over time. Daughter cells of actively dividing cells do not merely inherit DNA but also inherit the markings placed upon the DNA, such as histones and sites of methylation. Indeed there is no doubt that foreign DNA introduced into cells is also subject to such markings that will also be passed down to their progeny. The study of such markings is termed 'epigenetics' and is essential in the study of cells' mechanisms

of expression and/or silencing and the passing down of plasmid DNA down the generations.

1.6 Muscular Dystrophy Treatment

1.6.1 Viral Approaches for Gene Delivery

Recombinant Adeno-Associated Virus (rAAV) is a viral vector that carries single stranded DNA and the wild type requires a helper virus in order to replicate (Atchison *et al*., 1965). It has been engineered to be able to replicate on its own (rAAV) and is able to infect both dividing and post-mitotic cells (Podsakoff *et al*., 1994) which allows it to target a wide variety of tissues. Its DNA does not integrate into the host genome but has been found to persist episomally (Duan, 1998; Schnepp *et al*., 2009) unless high doses of the virus are administered where integration has been shown to occur (Chamberlain *et al*., 2004; Inagaki *et al*., 2007). Disadvantages of using this vector include the packaging capacity which is approximately 5kb, and the fact that the episomes delivered are eventually lost with cell division in mitotic tissue (Muir and Chamberlain, 2009). In treating Duchenne muscular dystrophy where a correct form of the dystrophin gene is required the rAAV vector system would be unsuitable due to packaging issues. However, several methods have been devised to circumvent this issue. Mini- and micro-dystrophins have been engineered which yield truncated yet functional forms of the dystrophin protein and are within the carrying capacity of AAV. However, the problem with using truncated versions of the original protein is shown by several studies where dystophin's function of localising the signalling molecule nitric oxide's precursor nNOS (neuronal nitric oxide synthase) is lost (Thomas *et al*., 1998; Kobayashi *et al*., 2008). This molecule is important for supplying blood to active muscles and vasodilation (Thomas *et al*., 1998; Kobayashi *et al*., 2008).

Another way the rAAV packaging problem was resolved was by the use of a 'split-vector' approach where the full length cDNA of the dystrophin gene is carried by two AAV vectors. This then involves trans-splicing, recombination, or the use of overlapping sequences in order to combine the transcripts into one, thus yielding a full length dystrophin transcript that can be expressed. The trans-splicing method is a discontinuous group II intron trans-splicing method that occurs in mammalian cells and

is used to join two coding sequences together. It requires the insertion of a splice donor sequence at the tail end of the first transcript carrying the 5' end of the sequence, derived from a natural dystrophin intron. A splice acceptor sequence from the same intron is inserted at the start of the 3' transcript (La *et al*.,. 2005). The sequences are then spliced together within the cell creating a full length transcript yielding the dystrophin protein. When recombination is the method to be employed, the transcripts contain fragments that overlap and undergo homologous recombination within the cell.

Lentiviral vectors, like AAV, can infect a wide range of tissues that are mitotic or post-mitotic, but it has a greater carrying capacity than rAAV vectors (approximately 9kb) (Muir and Chamberlain, 2009). However, it integrates its DNA into the host genome (Muir and Chamberlain, 2009). Although integration could provide longer lasting expression, there is the risk of insertional mutagenesis in addition to the hazard of viral promoters and enhancers activating proto-oncogenes, and conversely disruption of gene expression that may inactivate a tumour suppressor gene locus (Hacein-Bey-Abina *et al*., 2003; Ciuffi *et al*., 2006; Beard *et al*., 2007). A safer lentiviral vector has been referred to as SIN (self-inactivating) vectors that do not contain any of these promoters or enhancers and has been used to deliver minidystrophin to muscle (Li *et al*., 2005).

Although viral gene transfer methods may result in greater gene expression and more systemic delivery, the use of non-viral vectors has many benefits such as the reduced risk of insertional mutagenesis (Glover *et al*., 2005), and a carrying capacity for transgenes which is not limited by a packaging capacity. There are several different approaches that are included as non-viral gene therapy methods. These include oligonucleotide-mediated exon skipping, oligonucleotide-mediated genome repair, cell therapy, and naked plasmid DNA (pDNA) delivery amongst others (reviewed in Trollet *et al*., 2009; Muir and Chamberlain, 2009)

1.6.2 Muscular Dystrophy Treatment Using Non-Viral Gene Transfer

An estimated 10-15% of cases of Duchenne muscular dystrophy are caused by nonsense mutations where a base pair is altered leading to a frameshit mutation, a premature stop codon, or an inappropriately placed splice site that results in a non-

functional protein (Trollet *et al*., 2009). The exon skipping approach aims to target such mutations by allowing the transcription machinery to effectively 'skip' the exon where the mutation is located and continue reading the rest of the gene in-frame. This results in a truncated yet still functional dystrophin protein, thus alleviating the severe symptoms of Duchenne and changing it to the milder form of Becker's.

Antisense Oligonuceotide-Induced exon skipping is an RNA-based approach where an oligonucleotide complementary to the exon targeted is designed which then modifies the pre-mRNA splicing leading to the restoration of the reading frame (Trollet *et al*., 2009). The oligonucleotides must be designed specifically for each exon that is to be skipped. The skipping can occur either by physically blocking the enhancer sequences at a specific location, or by altering the secondary structure of the pre-mRNA in order to prevent splicing, thus 'skipping' the targeted exon (Trollet *et al*., 2009) by the removal of the mutated exon together with its flanking introns (Muir and Chamberlain, 2009). The stronger the binding of the oligo and the more easily accessible it is to the target sequence, the more effective exon skipping will be (Popplewell *et al*.,2009; Wilton and Fletcher, 2007).

There are several different chemistries of oligonucleotides that have been designed for exon skipping. These include Peptide Nucleic Acids (PNAs), Locked Nucleic Acids (LNAs), 2'-O-Methoxyethyl (2'OMe), and Phosphorodiamidate Morpholinos (PMOs) (Trollet,C. 2009). In terms of safety the PMOs and 2'OMe chemistries have proven the safest (van Deutekom *et al*., 2007). PMOs, also called morpholinos, have a similar chemistry to 2'OMe, except that the ribose rings were replaced with morpholino rings thus replacing the phosphodiester bonds with phosphorodiamidate (reviewed in (Muir and Chamberlain, 2009). This gives the PMOs a better ability to resist nuclease degradation within a cell, and they have proven to be the most efficient at maintaining skipping that is both reliable and persistent (Trollet *et al*., 2009). This was shown in studies utilising in the mdx mouse model (Alter *et al*., 2006) and human muscle explants (McClorey *et al*., 2006). It was also taken into clinical trials (Kinali *et al*., 2009). 2'OMe was also in clinical trial (Arechavala-Gomeza *et al*., 2007) showing that there were also no apparent adverse effects resulting from this treatment (van Deutekom *et al*., 2007). Another advantage of using PMOs is that their delivery to target cells can be further improved by their conjugation to cell-penetrating peptides (Fletcher *et al*., 2007).

The drawback with exon skipping, however, is the requirement for repeated oligonucleotide administration due to their short half-life (Vacek *et al*., 2003). The major risk with this method is non-specific binding of the oligo which may cause unknown adverse effects (Trollet *et al*., 2009). Even if the non-specific binding is very low, administration of high doses of oligos in order to reach an acceptable therapeutic index would magnify such harmful effects (Muir and Chamberlain, 2009). Furthermore, the cost of creating these oligos is high (Muir and Chamberlain, 2009).

Another non-viral method involves the delivery of naked DNA such as a plasmid containing the correct form of the gene. There are several methods that have been devised for the delivery of naked plasmid DNA into muscle tissue *in vivo*. Although these methods are less efficient than the use of viral vectors, there is no doubt that some of them are safer for use in gene therapy (Gao *et al*., 2007). One is the direct injection of the plasmid DNA however the DNA does not enter the cells easily and with great efficiency, and it may be degraded by intra-/extra-cellular nucleases before reaching the target nucleus. A second method is the use of a gene gun. This involves coating gold particles with pDNA and penetrating cells of a tissue with them thus releasing the DNA into these cells. A third and popular method is electroporation, where the pDNA is injected into the tissue first before electrodes are inserted and moves the DNA along an electric field through the cell membrane which is reversibly permeabilised and into the cell. Vilquin *et al* (Vilquin *et al*., 2001) found that both healthy and dystrophic mouse muscles were transfected with the same efficiency indicating that fibrosis does not interfere with the efficiency of electroporation-mediated gene transfer. Therefore this method could be useful and applicable to gene therapy of muscular dystrophy. However, several issues are posed by the use of this method. The first is that the expression in the study by Vilquin *et al* was mostly found near the sites of injection and was not spread out over the whole muscle area (Vilquin *et al*., 2001). It is further limited in that it cannot be applied to deep seated tissue within the body without surgery being required, nor for large areas of tissue (Gao *et al*., 2007). Furthermore, it has safety implications in that it could considerably damage the muscle. A fourth method involves the use of ultrasound where it creates temporary holes within the cell membrane through which the DNA can diffuse into the cell, however this method has a low efficiency and needs improvement. A fifth method is the hydrodynamic injection via the tail vein of mice where a temporary overflow of

fluid within the mouse's system results in a overflow within the liver thus efficiently deliver the DNA into the liver. This method is the most efficient non-viral method, however it may be more complicated to apply to humans. Chemical methods have also been used to deliver pDNA to target cells where particles are taken up by the cells by methods such as endocytosis or phagocytosis into the cytoplasm, and then find their way to the nucleus. The most used chemical method is the complexing of pDNA with cationic lipids to form lipoplexes. The benefit of using this method is that the pDNA is protected until it is released into the cell. However, these lipoplexes can be toxic to some cells according to the cell type (Gao *et al*., 2007).

A limiting factor to the use of plamid DNA is that plasmids contain antibiotic resistance markers used to maintain the plasmid within the host cell by using an initial period of selection for establishment. When selection is removed, the episome is gradually lost at a rate of 4% per cell generation (Glover *et al*., 2005) and if expression is maintained it is due to an integration event occurring.

Furthermore, these plasmids do not replicate within their host genomes. A study by Vilquin *et al* confirmed this (Vilquin *et al*., 2001). They attempted the use of notexin, a snake venom that destroys muscle fibres, after transduction with the pDNA carrying the β-Gal reporter gene in healthy mice, and then assessed the expression after the muscles had regenerated. It was found that the β-Gal reporter protein could hardly be detected. This indicated that the plasmid did not transduce muscle satellite cells efficiently, and if it did then the plasmid was unable to replicate with the cell divisions and the plasmid was subsequently diluted out and lost.

Considering the safety implications involved with the use of viral vectors, non-viral vectors have been utilised and manipulated in many studies in an effort to improve the levels and duration of transgene expression. In order to develop these vectors, it is necessary to understand the epigenetic factors that led these vectors to be less than ideal to begin with. Therefore, a description of epigenetics, the mechanisms of DNA marking, and gene silencing will be discussed, before reviewing the improvements that have been made to conventional non-viral vectors to circumvent their shortcomings thus far, and how this can be applied to the development of a novel therapeutic approach to the treatment of muscular dystrophy.

1.7 Epigenetics

'Epigenetics' is defined as "the study of heritable changes in gene function that occur without alterations to the DNA sequence" (Probst *et al*., 2009). Epigenetic marks are things that can define the reading of DNA by the cell and can be passed down to daughter cells. Epigenetic markers are both stable and flexible modifications to DNA and include the methylation of cytosine residues, the packaging of DNA with histones, of which there can exist several variants, the modification of these histones, such as acetylation, the binding of non-histone proteins to DNA or to histones/modified histones, higher-order chromatin organisation, and positional information. All of these marks contribute to and influence the way in which DNA is accessed and therefore read and transcribed (Probst *et al*., 2009). Certain modifications are short term and can occur as a result of a signal due to specific stimuli, whereas others play a role in cellular memory and remain longer-term (Turner, 2002).

1.7.1 Histones

Histone packaging of DNA is an important function of the cell and is crucial in DNA organisation. DNA is organised by wrapping around a histone octamer that is composed of an H3-H4 tetramer and capped by two H2A-H2B dimers. Histone H1 is a linker histone, and all together this forms a nucleosome. The modifications that can occur to histones include methylation (Me), acetylation (Ac), and phosphorylation (P). The acetylation of histones is conducted by histone acetyltransferase enzymes (HATs) and involves the neutralising of the histone tail's positive charge leading the histone to have less affinity to the negatively charged DNA (Grant, 2001). Acetylated histones are found with DNA that is transcriptionally active such as euchromatin. Deacetylated histones are found to be associated with inactive DNA such as heterochromatin, and are deacetylated by histone deacetylases (HDACs). Phosphorylated histones are also those associated with transcriptionally active DNA and, as with histone acetylation, also have their tail charge neutralised, and here it is by the addition of a negatively charged phosphate group (Grant, 2001). It is believed that histone phosphorylation may actually lead to transcriptional activation, possibly by stimulating the activity of HATs on the same histone tail leading to acetylation as well. These marks create a

'histone code' which can be read alone, in combination if several marks exist, or in addition to marks present on other histones nearby.

In addition to histone markings, the variant of histone present can relate information. Different variants can lead to variable levels of nucleosomal stability which can determine regions of higher or lower degrees of transcriptional activity.

Groth *et al* suggest that there may be a mechanism by which the cell re-uses and re-incorporates parental histones and newly created ones (Groth *et al*., 2007). Histone chaperones would be the factors that carry out functions such as tethering histones and assembling or disassembling nucleosomes. One such factor is the chaperone chromatin assembly factor (CAF1) that associates with H3.1 and H4 in addition to PCNA and histone deacetylases or Lys methyltransferases (Shibahara and Stillman, 1999). Another such factor, also an H3-H4 chaperone, is ASF1 which can also interact with CAF1 and donates newly made histones during replication. Newly made H4 histones do carry acetylation marks (Loyol *et al*., 2006) that are subsequently removed. There are two ways in which these marks are inherited after the passing of a replication fork where such marks are inevitably displaced. One is to inherit histone variants outside of S phase where the inheritance is not coupled to DNA replication. These are histone variants such as H3.3 or CenH3 (CENP-A in humans) that bind to the centromeres of chromosomes (Loyola and Almouzni, 2007; Henikoff *et al*., 2004). H3.3 is found to associate with transcriptionally active DNA and can mark it as such. These histones have many active histone markers (Loyola *et al*., 2006; McKittrick *et al*., 2004) and also make the nucleosomes in which they are contained less stable than those that may have the variant H3.1 incorporated (Jin and Felsenfeld, 2007). The degree of instability varies according to other additional factors which include the amount and type of modification status of the nucleosome itself (Loyola *et al*., 2006), or the possible incorporation of other variants (Henikoff, 2008). For example, histone H3 variant is defined as a replicative histone and is expressed and incorporated during DNA replication (for example, H3.1 and H3.2), whereas a replacement variant is expressed throughout the cell cycle and is incorporated in a DNA synthesis-independent manner (for example, H3.3m and the centromere-specific histone H3 variant CenH3) (Grant, 2001). This means that when DNA is not being replicated in the S phase, the exchange of replicative histone variants such as H3.1 and H4 is very low, whereas that for H2A and H2B can easily occur (Kimura and Cook, 2001). This could

point to the histone variants H3.1 and H4 to be the ones most liable to transfer information which can be considered heritable.

The second way in which marks are inherited is by being coupled with replication. After a replication fork passes histones are displaced and there is no template from which the nucleosomes can be reassembled. There are three possible scenarios by which parental histones are re-incorporated into replicating DNA after displacement. The first is considered random, where the marked H3 and H4 histones are depositied randomly and the corresponding gaps where histones are missing are subsequently filled by new H3-H4 dimers. The second proposal is a semi-conservative one, where the parental dimers are distributed evenly between the daughter strands and new ones are placed to complete the nucleosome structures. The final theory is an asymmetric distribution where all the parental histones are transferred to one daughter strand and newly synthesised ones are placed in the other strand based on possible cross-talk between the two strands. It is quite possible that all three proposals could be occurring as well.

It is also a possibility that new histones could inherit their marks from older neighbouring ones. It has been found that PCNA and CAF1 are able to stay on replicated DNA for an amount of time sufficient enough to allow them to modify the new chromatin such as by removing the acetylation markers present on newly synthesised histones (Sporbert *et al*., 2002; Taddei *et al*., 1999; Loyola *et al*., 2006). It is also possible that regions which are highly methylated are able to guide histone modifications via the machinery that maintains these methylation marks that may act as a guide. Sarraf and Stancheva (Sarraf and Stancheva, 2004) found that methyl CpG-binding protein (MBD1) is capable of forming a complex with Lys methyltransferase (SETDB1) and Reese *et al* found that this complex is able to associate with CAF1 during replication (Reese *et al*., 2003). Furthermore, Fuks *et al* found an association between DNMT1 and HDAC1 (Fuks *et al*., 2000) and with HDAC2 (Rountree *et al*., 2000). This indicates that a connection may indeed exist between DNA modifications via marks and the placement of histones.

1.7.2 Plasmid DNA And Epigenetics

The aim of a study conducted by Riu *et al* was to determine whether histones associate with pDNA, and if they are at least in part responsible for silencing if it occurs (Riu *et al*., 2007). Riu *et al* confirmed that plasmid DNA is epigenetically altered within cells (Riu *et al*., 2007). In the study by Riu *et al*, mouse livers were transfected with minicircle, or the minicircle parental plasmid, using RSV or EF1-α promoters to control transgene expression. It was found that the minicircle produced expression that was 100 times that of the parental plasmid and was persistent whereas the parental plasmid was silenced over the period of several weeks (Riu *et al*., 2007). This was not due to any difference in plasmid copy number within the livers as this was tested and the numbers were found to be similar. Antibodies were used against certain modified histones and chromatin immunoprecipitation (ChIP) analysis was performed. The results indicated that the minicircle had histones associated with it whose modifications were those found to be associated with euchromatin. This shows that pDNA does undergo chromatinisation within cells and that the sustained expression observed can be attributed to the association of euchromatin-associated modifications of the histones associated. The analysis was conducted at several time points. On day 1 no differences were observed between the two indicating that chromatinisation had not yet fully occurred, or that the DNA had not yet reached the correct nuclear compartment. Similar results were observed by day 7. At day 35 however, it was shown that the parental samples contained 2-5 fold more heterochromatin markers than the minicircle samples, indicating transcriptional repression. These markers included di- and tri-methylation of lysine 9 on histone 3 (H3K9me2 and H3K9me3), di- and tri-methylation of lysine 20 on histone 4 (H4K20me2 and H4K20me3), histone deacetylase 2 (HDAC2), HP1α, and H3K9 methylase SUV39H1. Low levels of these markers were found within the minicircle lysates, however the authors attributed this to a low level of contaminating parental plasmid within the minicircle precipitation solution. In the minicircle lysates, 2-6 fold more euchromatin markers were present than in the parental plasmid lysates, which indicated transcriptional activation. These included the acetylation of histones 3 and 4 (H3Ac and H4Ac), the di- and tri-methylation of lysine 4 in histone 3 (H3K4me2 and H3K4me3), the di-methylation of

lysine 79 in histone 3 (HeK79me2), the acetylation of lysine 9 in histone 3 (H3K9Ac), and the phosphorylation of serine 10 on histone 3 (H3S10P). This indicates that chromatin in an open configuration can lead to long term transgene expression.

These results were obtained when the RSV and the EF1-α promoters were used. Interestingly, when a third promoter was tested, the UbC promoter, sustained transgene expression resulted from both the minicircle and the parental plasmid, without silencing. Both plasmids showed high levels of serum hAAT, which was the transgene being expressed, over the period of 35 days. ChIP analysis at day 35 indicated that both types of lysates had low amounts of heterochromatin-associated markers and large amounts of euchromatin markers in both the promoter and cDNA regions. This indicates that sustained long term expression is a result of an open and active state of DNA that has been marked with euchromatin-associated markers.

1.7.3 Methylation

Silencing of DNA occurs by histone modifications and changes in chromatin structure, both of which are easily reversible (Shi *et al*., 2004), in addition to methylation, which is a more permanent modification. DNA methylation is the alteration of the cytosine residue by the covalent addition of a methyl group by methyltransferases with S-adenosyl-methionine as the methyl donor (Miranda and Jones, 2007). Methylation of CpG residues takes place in mammalian cells and is usually a marking of heterochromatin (Miranda and Jones, 2007). Methylation as a silencing mechanism may have evolved due to the foreign elements present within the human genome such as transposons and parasitic elements which are potentially dangerous if not silenced (Kochanek *et al*., 1995; Robertson and Wolffe, 2000). Furthermore, methylation is important for other heritable epigenetic markings such as X-chromosome inactivation and imprinting. X-chromosome inactivation occurs during embryogenesis is where one allele is methylated and silenced leading to expression resulting only from the other allele (Chang *et al*., 2006). Imprinting is also important in order to control which parental allele will be expressed. DNA methylation also occurs in somatic cells during differentiation where certain genes are silenced thus allowing the cell to progress down a particular lineage (Webe *et al*., 2007).

Methylated DNA recruits methyl-binding proteins (MBPs) to the methylation sites. There are five members in this protein family, all of which contain a homologous domain that binds methylated CpGs (Sansom *et al*., 2007). They are called MBD1-3, MeCP2, and Kaiso. These proteins can bind to histone deacetylases and repressors that make the chromatin inactive by changing its structure (Harikrishnan *et al*., 2005). De novo methylation occurs by DNA methyltransferases recognising chromatin structures within the genome. Euchromatin is DNA that is actively transcribed and is defined by the di- and tri-methylation of histone H3 at its lysine 4 site, and the acetylation of the histones H3 and H4 that are bound to the DNA. Kouzarides states that this is a flexible and reversible state of DNA (Kouzarides, 2007). Heterochromatin is DNA that is silenced and compacted and is characterised by the trimethylation of histone H3 at the lysine 27 or the lysine 9 site, or the methylation of histone H4 on lysine 20. These histone modifications in heterochromatin may possibly be targets for methyltransferases.

Methylation can lead to silencing by two proposed theories. The first is that the methyl groups bound block the binding of transcription factors to the DNA. Comb and Goodman (Comb and Goodman, 1990) found that many transcription factors bind to sequences that contain CGs and therefore methylation could be interfering with the binding. The second proposed theory was that methylation alters the structure of the chromatin by affecting histone modifications and therefore the structure of the nucleosomes. This in turn could block the binding of transcription factors by disallowing activation markings to occur or to be propagated. When methylation occurs in promoters gene expression is silenced and the genes cannot be reactivated, even when the silencing histone markers are removed (McGarvey *et al*., 2007).

The lack of methylation markers on CpG motifs of foreign DNA can also trigger an immune response. Cells have an innate immune response against forgein DNA (Vilaysane and Muruve, 2009). Toll-like receptor 9 (TLR9) is a protein that is bound to the membranes of endosomes and acts as a DNA sensor (Vilaysane and Muruve, 2009). Hemmi *et al* showed that TLR9 was able to recognise foreign unmethylated bacterial DNA but not its own self-DNA (Hemmi *et al*., 2000). TLRs are able to activate the transcription of cytokines that initiate an inflammatory response.

1.8 Improving Plasmid Transgene Expression

One approach involved the development of a novel kind of plasmid called the 'minicircle', which is as a conventional plasmid but does not contain an antibiotic resistance gene nor any bacterial sequences but solely the gene transcriptional unit (Chen *et al*., 2003). Its creation is based upon the recombination of a parent plasmid yielding the minicircle and a 'maxicircle' which is then digested and degraded leaving only the minicircle. Its benefits include the circumvention of the safety issues that come with the use of a conventional plasmid in addition to better transfection efficiency as a result of its smaller size and a smaller likelihood of silencing from CpG methylation. Minicircles are also relatively easy to produce. Chen *et al* developed a new plasmid vector devoid of bacterial backbone based upon the *Streptomyces* temperate phage integrase ØC31 –mediated site-specific intramolecular recombination technology (Chen *et al*., 2003). The parent plasmid undergoes recombination at two specifically inserted sites resulting in a minicircle containing the transcription unit, and a maxicircle containing the bacterial backbone. The principle is based on the destruction of the maxicircle and parent plasmid after recombination has taken place by the insertion of the endonuclease I-SceI gene and I-SceI recognition sequences outside the minicircle unit in order to linearize maxicircle and parent plasmid, which would eventually be digested by exonucleases within the cell, leaving the minicircle intact. Chen *et al* conducted a study which compared the human FIX expressing minicircle to the unrecombined plasmid in mice (Chen *et al*., 2003). The mice infused with the minicircle had high levels of human FIX in their serum initially, which then stablilized to a concentration of 12µg/ml, being over two fold normal levels, which remained constant for the 7 weeks that were the length of the experiment. The mice infused with the unrecombined plasmid showed high expression at first and then dropped more than 45 times the initial amount within three weeks and continued to decrease thereafter. When tested with another expression cassette using human AAT, minicircle transgenes gave a 560 fold higher level of expression than that of the unrecombined parent plasmid transgene. This study shows that the minicircle can provide high levels of transgene expression and persist within mouse liver. However, it is unable to be replicated in dividing tissue and would best be employed in post-mitotic tissue.

The second approach was to change the promoter. Many studies have utilised ubiquitous promoters, often of viral origin, such as the cytomegalovirus (CMV) promoter, which drive high levels of transgene expression in a wide variety of cell types. However, expression is often short-lived and is shut down after a period of time. The use of non-viral ubiquitous promoters can circumvent such issues (Gill *et al*., 2009). In a study by Pringle *et al* for example, the use of the human polyubiquitin C (UbC) promoter drove the long term expression of the transgene when delivered as naked plasmid DNA to mouse lungs by electroporation (Pringle *et al*., 2007).

Fusion promoters have also been tested with promising results, such as the CMV enhancer/chicken β-actin (CAGG) fusion promoter in the liver that allowed sustained long-term expression (Pringle *et al*., 2007). Several studies have compared the CMV and CAGG promoters in plasmid constructs and found the resulting expression to be superior when driven by the CAGG promoter. One such study was conducted by Niwa *et al* where it was tested in several cell lines and was found to have superior expression over the CMV promoter construct (Niwa *et al*.,1991). A second study by Garg *et al* used the CAGG promoter to drive the expression of the influenza-hemagglutinin (HA) gene, to be used as a DNA vaccine against influenza (Garg *et al*., 2004). The expression of this construct was compared with one driven by the CMV promoter in mice. It was found that there was a significantly higher immune response from the vaccine with the CAGG construct than the CMV one, leading to the conclusion that its expression better than that of the CMV. Another study by Alexopoulou *et al* used vectors expressing eGFP driven by CAGG that were stably transfected in undifferentiated and differentiated murine embryonic stem cells (CCE cell line) (Alexopoulou *et al*., 2008). This vector was directly compared to one driven by the CMV promoter, and one by the chicken β-actin promoter. Under the CAGG promoter it was found that 50% of the stable transfectants of undifferentiated and differentiated cells were positive. Under the CMV promoter it was found that less than 50% of cells were eGFP positive and this declined further in differentiated cells. The same was seen under the chicken β-actin promoter, leading to the conclusion that the CAGG promoter drove higher expression than the other two promoters tested in these cells. Yet another study, by Ngyen *et al* aimed to compare four promoters driving the luciferase transgene in Hek 293 cells and in mouse liver by hydrodynamic injection through the tail vein (Nguyen *et al*., 2008). These promoters were CMV, CAGG, EF1α (elongation

factor-1 α), and PGK (phosphoglycerokinase). It was found that the CAGG promoter produced the highest levels of transgene expression both *in vitro* and *in vivo* when assessed 24hours post transfection.

Tissue-specific promoters are also useful tools for driving expression as, although they may not lead to initial expression levels as high as promoters of viral origins, they can allow for more specific expression in target cells, leading to increased safety. They may also lead to more sustained, long-term gene expression as was observed in the study by Argyros *et al* in comparing the AAT liver-specific promoter to CMV, where the expression resulting from AAT was slightly higher than that of CMV 6 months post transgene delivery (Argyros *et al*., 2008).

A third approach to improve expression was to reduce or eliminate CpG sequences from the pDNA that can be methylated. In a study by Hyde *et al* the aim was to improve the expression of a plasmid carrying the Cftr transgene complexed with cationic lipids to the lungs of cystic fibrosis mouse model by aerosolization (Hyde *et al*., 2008). The CMV promoter was used to drive expression, which lasted approximately 2 weeks before being silenced. The promoter was then changed to UbC, and a small increase in the duration of the expression was observed. The aim was then to determine whether the CpG motifs within the plasmid sequence were the cause of the silencing observed. It was found that even a single CpG sequence was enough to trigger an immune response within these animals leading to silencing, and that the eradication of these motifs resulted in continuous long term transgene expression free from an immune response, and the level was adequate enough to functionally improve the condition of cystic fibrosis (Hyde *et al*., 2008).

A fourth approach to improve transgene expression and its duration was to include a Scaffold/Matrix Attachment Region (S/MAR) element within the open reading frame of the transgene in the plasmid. The inclusion of this element has led, in some cases, to prolonged expression without being subject to silencing, in addition to preventing integration and thus maintaining the plasmid in an episomal state in actively dividing cells (Jenke *et al*., 2004). Such a vector appears to be a promising gene therapy vector as it overcomes the integration issues that hold back other potential vectors whilst at the same time offering stable, long-term transgene expression. A plasmid containing an S/MAR element has not, as yet, been tested in muscle cells. Considering the properties the S/MAR element can potentially confer to its plasmid, this could be a

useful vector in the treatment of muscular dystrophy if it can be directed to and successfully transfected into muscle satellite cells.

In order to understand the mechanism by which S/MAR elements can confer such properties to a plasmid, it is important first to evaluate what S/MARs are, what defines them, and what their natural functions and characteristics have been shown to be.

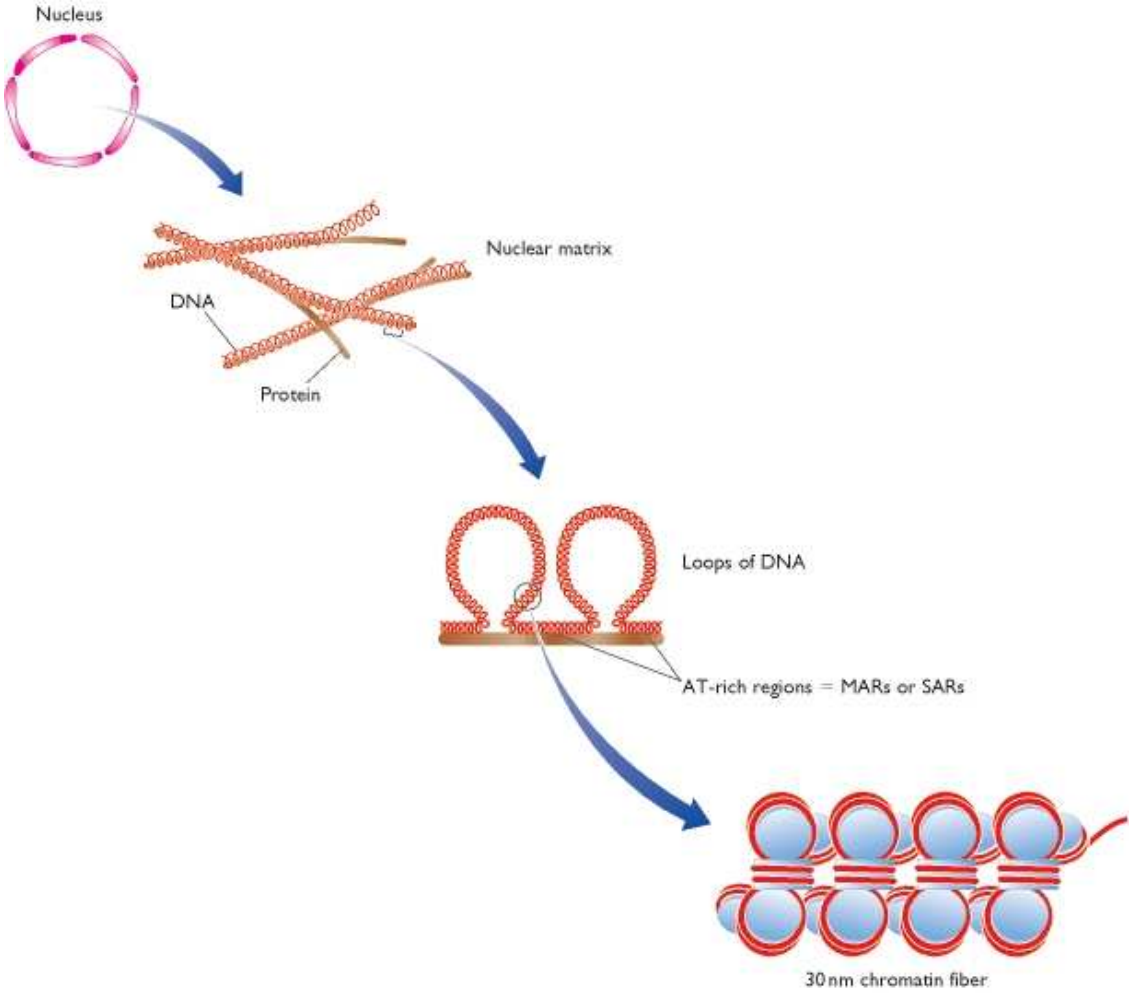
1.9 Scaffold/Matrix Attachment Region (S/MAR) Elements, DNA Organisation, and The Nuclear Matrix

Scaffold or matrix attachment regions (S/MARs) are DNA sequences that are considered to be chromatin domain borders (Bode *et al*., 2003). S/MAR elements are interspersed throughout the genome and are typically found at intervals of between 5-200kb (Bode *et al*., 2003). S/MARs do not have a consensus sequence, but all are AT-rich and have an unusual inclination for strand separation, which occurs most easily at the site of a core-unpairing element (CUE), where the DNA destabilizes the most, and strand separation spreads out from there throughout the base-unpairing regions (BURs) (Bode *et al*., 2006). It is believed that the 25 million nucleosomes in mammalian cells are arranged by S/MARs into around 60,000 chromatin loops (Heng *et al*., 2004) (Figure 1.6). These loops are dynamic, as the S/MARs move along the matrix, attaching as needed (Heng *et al*., 2004).

Eukaryotic chromosomes have a quarternary structure with many levels of packaging. DNA associates with histones that are compacted into structures called nucleosomes, which then fold into chromatin fibers. Chromatin is organized into independent loops of DNA called replicons (Heng *et al*., 2004). These loops are anchored at the S/MAR sites via proteins to the nuclear matrix (Bickmore and Oghene, 1996) (Figure 1.6) which is made up of protein, RNA (Bode *et al*., 2003), and nuclear lamina (Wan *et al*., 1999). Several proteins that enable the S/MAR-matrix association have been identified and are involved in nuclear architecture. These include two isoforms of the scaffold attachment factor A (SAF-A), two isoforms of SAF-B, and the factor E1B-AP5 (Kipp *et al*., 2000). Kipp *et al* found that these proteins bind S/MARs via a SAF-Box which is a domain that has been evolutionarily conserved (Kipp *et al*., 2000). This domain does not recognise specific sequences, considering S/MARs do not contain consensus

sequences, however the interactions occur as a result of the S/MARs' structural features and possibly a series of sequence motifs common to these elements (Kipp *et al*., 2000).

The loops created as a result of S/MAR-matrix association are important structures for packaging of DNA, DNA replication, and transcription (Bode *et al*., 2003). It is believed that the matrix plays two important roles in DNA replication. Firstly, it acts as a support for the replication machinery; and secondly the key proteins required for replication are associated there, with Hozak *et al* establishing that the replication machinery is, in fact, attached to the nuclear matrix (Hozak *et al*., 1993). In addition, Pardoll *et al* proved that newly synthesized DNA moves away from the loop bases, suggesting that DNA replication occurs at the bases of the loops by DNA polymerase complexes attached to the nuclear matrix (Pardoll *et al*., 1980).



(Brown, 2002)

Figure 1.6 Adapted from Brown, 2002. The figure shows DNA organisation within the nucleus as packaged by histones, and indicates loops of DNA attached to the nuclear scaffold/matrix via AT-rich regions of the genome, known as scaffold/matix attachment regions (S/MARs).

1.9.1 DNA Replication Initiation Complex And The Binding of Associated Proteins To S/MAR Elements

DNA replication is associated with the nuclear matrix and it was found that the origin of replication binds to the matrix just before S phase of the cell cycle (Cook, 1999). S/MARs have been found to be located close to origins of replication where they arbitrate function (Bode *et al*., 2003) or even act as origins of replication themselves (Jenke *et al*., 2002). Their AT tracts are highly important for this function as origin recognition complexes (ORCs) preferentially bind to AT rich sequences and do not discriminate between AT-rich tracts and natural Ori sites (Abdurashidova *et al*., 2003). Replication initiation begins with the initiation complex assembling at origins of replication. Cells have a mechanism to regulate genomic replication by limiting it to a once-per-cycle event. The origin recognition complex, which makes up the pre-replication complex, recognizes and binds an origin of replication at early G1 phase. Other proteins are then able to bind such as Cdc6, Cdt1, and MCM2-MCM7 (MiniChromosome Maintenance) proteins. As the cell cycle progresses to S phase protein kinases are activated which lead to the pre-replication complexes to becoming initiation complexes (Bell and Dutta, 2002; Blow and Hodgson, 2002). Once replication is initiated the pre-replication complex loses the associated Mcm proteins and partly dissociates (Schaarschmidt *et al*., 2002) which then leads to the complex's dissociation from the DNA (Tye, 1999). It has been postulated that replication origins are not determined by DNA sequence as much as by epigenetic factors such as the binding of specific factors its position within chromatin of a specific conformation (Mechali, 2001). Studies using BrdU indicated that the episomes replicate in a once-per-cycle manner just as genomic DNA (Price *et al*., 2003). Price *et al* blocked the cells at the G1/S-phase checkpoint and, using BrdU again, determined that the time the episome is replicated is at early S phase (Price *et al*., 2003). Regions containing S/MARs are considered to be areas that have open DNA conformations around active genes and may be replicated earlier (Gilbert, 2001; Lin *et al*., 2003). Using ChIP it was determined that Orc2p and Mcm3p are able to associate with the S/MAR. Price *et al* believe that they were able to determine a very loosely defined 36bp consensus sequence which may be contained in every origin of replication where the ORC binds (Price *et al*., 2003). If 3 mismatches are allowed on this sequence, it can be found to occur 6 times

within the S/MAR sequence. Schaarschmidt *et al* also believe that the pre-RC components bind to the S/MAR at more than one site and are thought to be functional as they also dissociate partially in S phase as they would when bound to genomic DNA (Schaarschmidt *et al*., 2004).

Other proteins that have been found to have an affinity for S/MARs which demonstrate the involvement of some S/MAR elements in DNA replication are Topoisomerase II, nucleolin, and NuMA. Adachi *et al* found that S/MARs have Topoisomerase II cleavage sequences within them (Adachi *et al*., 1989). Topoisomerase II is an enzyme capable of relaxing supercoiled DNA and unknotting DNA by introducing double-stranded breaks and then re-linking them back again. These enzymes are found on the nuclear matrix during metaphase (Adachi *et al*., 1989) and are important in resolving replication. In addition, nucleolin, a protein that is highly abundant in the nuclei of actively dividing cells, also binds S/MARs at the minor groove of AT rich tracts (Dickinson and Kohwi-Shigematsu, 1995). Finally, Luderus *et al* found that the structurally related protein NuMA which has an important role during mitosis (Compton and Cleveland, 1993) also specifically binds to S/MARs (Luderus *et al*., 1994).

1.9.2 The Role of S/MARs In Transcription And Its Related Factors

S/MARs have also been found to exert an effect on gene transcription (Bode *et al*., 2006). For instance, they can act as insulators for genes from positive/negative effects from the surrounding genome (Goetze *et al*., 2005). For example, an enhancer in one domain acts only on the gene within its respective domain and not beyond (Bode *et al*., 2003) (Figure 1.7). They can also increase the rate of transcription depending upon their location (Bode *et al*., 2000). It has been established that if the element is located immediately downstream from a promoter, transcription is hindered, whereas a more distant location can help transcription by allowing RNA polymerase to bind, and by relieving the supercoils in the DNA ahead of the enzyme (Bode *et al*., 2006).

Heng *et al* found that S/MARs are flexibly bound to the nuclear matrix and that this binding is dynamic in nature (Heng, *et al*., 2004) (Figure 1.8). Heng *et al* believe that the changing of the S/MARs that bind to the matrix is not due to a lack of binding

space for them, and that it must be regulated by the cell in order to meet its DNA's structural requirements and the cell's functional needs (Heng, *et al*., 2004). A correlation that was observed between gene expression and the transgene's proximity to the matrix which highlights the significance of DNA's proximity and contact with the matrix for transcription, and it appears to be vital for transcription to take place. (Vassetzky *et al*., 2000) experimented with the cMYC gene and its S/MAR and found that transcription of the gene actively takes place when the S/MAR is anchored and bound to the matrix. When gene expression was analysed it was found that not all the transgenes were being expressed but that the level of expression correlated with the number of transgenes associated with the matrix. (Bode *et al*., 1995) found that the S/MAR immediately downstream of the IFNB1 gene binds to the matrix as the gene is to be transcribed and, simultaneously, the S/MAR region becomes DNaseI insensitive indicating it is bound whereas the gene's coding region becomes sensitive (Figure 1.8). It is also believe that the stronger the S/MAR, the longer it is able to bind, which would directly affect gene expression. Just as proteins with roles in replication associate with S/MARs, so do proteins with an influence on transcription, such as SAF-B, certain lamins, histones, and high mobility group proteins (HMG-I/Y). SAF-B (scaffold attachment factor B) has been found to associate with RNA polymerase II in addition to serine-/arginine-rich RNA processing factors (SR proteins), which in turn associate with factors of the transcription machinery (Nayler *et al*., 1998). Nayler *et al* even proposed that SAF-B creates a base upon which transcription complexes in close proximity to actively transcribed genes are assembled (Nayler *et al*., 1998).

Transcription factors have been found to bind to a meshwork of protein composed of lamins, which are believed to have a role in organizing interphase chromatin. This leads to the assumption that lamins serve as a part of the nuclear matrix that may stablize active transcription complexes (Bode *et al*., 2003). A- and B-type lamins are capable of binding to S/MARs by associating with an open conformation single strand of S/MAR DNA, or by binding to the minor groove (Luderus *et al*., 1994).Additionally, histone H1 binds to S/MAR elements. It is a protein that packages chromatin leading to transcriptional silencing (Garrard, 1991) by cooperative binding (Zhao *et al*., 1993). It was found that HMG-I/Y isoforms prefer binding to A-rich tracts such as those found on the S/MARs, leading them to act as D-proteins (mimicking Distamycin), subsequently dislodging histone H1 and consequently derepressing S/MAR

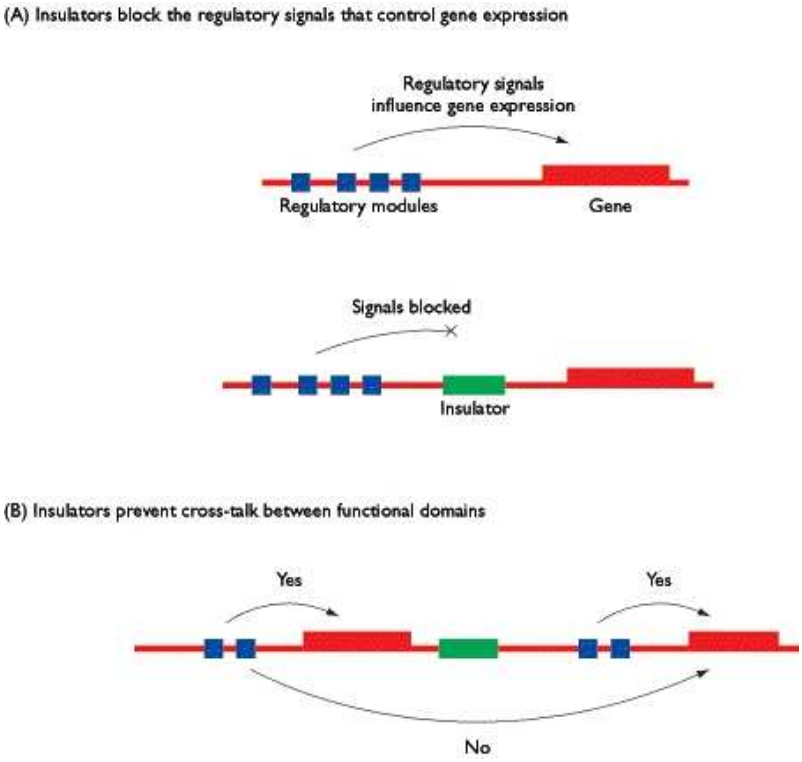
transcription and repressing non- S/MAR transcription by shifting the histones to the non- S/MAR DNA (Zhao *et al*. 1993). Martens *et al* have found that another histone, H3, was found at many S/MAR sites analysed within their experiment and that these were all found to be acetylated, which is an indication of an actively transcribed region of the genome (Martens *et al*., 2002).

Martens *et al* found that the transcriptional coactivator protein p300/CBP binds to S/MAR elements via the SAF-A protein (Martens *et al*., 2002). It also binds to several transcription factors and has been found to be involved in regulating transcription at promoters and enhancers in addition to S/MARs. Further it was shown to have acetyltransferase activity and can associate with other proteins such as p/CAF (Martens *et al*., 2002) which also exhibit such activity. Therefore p300/CBP is capable of acetylating histone tails (Struhl, 1998) which occurs at actively transcribed regions of the genome, and can also acetylate transcription factors which can either lead to transcriptional activation or repression (Sterner and Berger, 2000). Although transcription was not found to occur when p300/CBP is bound it has been suggested that it may make local chromatin ready to be transcribed.

An experiment involving the use of a plasmid containing the β-IFN S/MAR element was conducted by Jenke *et al* (Jenke *et al*., 2002). In it they attempted to co-purify vectors (one containing an S/MAR and the other one without the element) that had been transfected into cells, with the nuclear matrix. They found the S/MAR-containing plasmid in the purified lysate whereas the other plasmid was not, suggesting the association of this S/MAR plasmid to the matrix. A study by Baiker *et al* also confirmed this by the use of nuclear fractionation procedures where this same S/MAR plasmid was associated with the nuclear matrix (Baiker *et al*., 2000). They were also able to demonstrate that this association occurred via proteins through the use of fluorescent in-situ hybridisation, and that this plasmid can also be seen to associate with chromosomes. Furthermore, Jenke *et al* were able to show that if this S/MAR element was linked just downstream of an active transcription unit, this was sufficient for an episome to remain mitotically stable and replicate by associating to the nuclear matrix *in vivo* via the matrix protein SAF-A (Jenke *et al*., 2004).

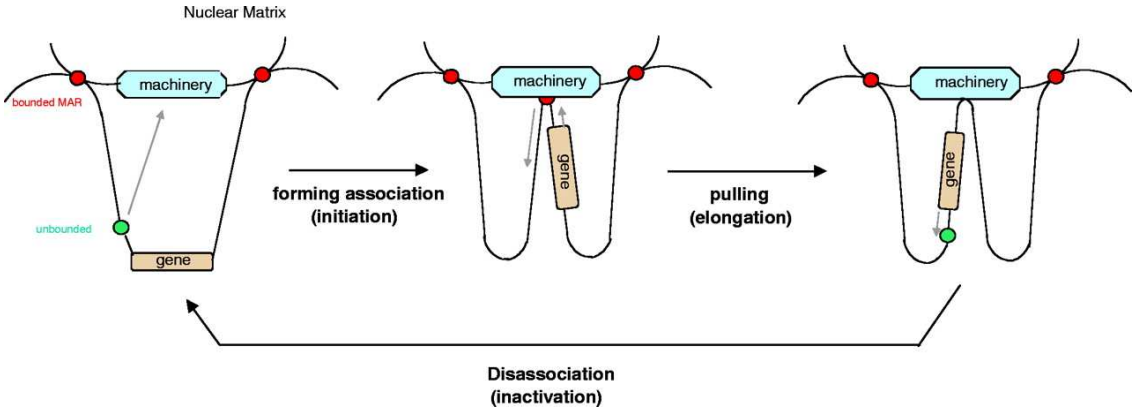
Considering all the roles S/MARs play in nuclear matrix attachment, DNA replication, and gene transcription, it is clear that they may have some useful applications in gene therapy. There have been several studies in the use of S/MAR elements in plasmid

vectors that have proven the element's role in maintaining the vector as an episome for many generations without integration into the genome. By counteracting methylation and allowing histone acetylation (Dang *et al*., 2000) an S/MAR-containing plasmid manages to avoid silencing and exhibits continuous transgene expression. Silencing is a eukaryotic defence system against the expression of foreign DNA, which is achieved by the methylation of CpG dinucleotides found in high abundance in non-mammalian organisms (Miranda and Jones, 2007). Methylated DNA is condensed and takes on a heterochromatin structure leading to transgene silencing, except in the presence of an insulator element such as an S/MAR (Goetze *et al*., 2005) whose properties allow the DNA to maintain an open, non-methylated conformation. Furthermore, the vector is able to bind to the nuclear matrix and replicate as the chromosomes do by using the cell's machinery, and get distributed and passed on to daughter cells in actively dividing cell lines, even without the use of selection, normally used to ensure maintenance of the plasmid and sustained gene expression (Jenke *et al*., 2004; Schaarschmidt *et al*., 2004).



(Brown, 2002)

Figure 1.7 Adapted from Brown, 2002. S/MARs can act as transcriptional insulators for genes from any effects resulting from regulatory modules in the surrounding genome. As shown in the figure above, when no insulator element, or S/MAR, is positioned between regulatory modules such as enhancers, for example, and a gene, the effect of the module can be exerted upon the gene downstream. However, when an S/MAR is present between them, the modules may exert their effects within their domain and not beyond the S/MAR element, thus insulating the gene from the surrounding influences of other regulatory modules.



(Heng *et al*, 2004)

Figure 1.8 Adapted from Heng *et al*, 2004. This figure shows S/MAR elements that are attached to the nuclear matrix (red dots, as indicated in the figure) and those which are not attached (green dots). The S/MAR element located prior to the gene brings the gene to be transcribed in proximity to the transcription machinery bound to the nuclear matrix, in order for transcription to occur.

1.10 S/MAR Studies

A number of studies have been conducted where plasmids containing the human β-IFN S/MAR element were transfected into cells and tested for their ability to replicate and be passed on to daughter cells after many rounds of replication. In these studies the element was inserted within the open reading frame of the reporter gene eGFP and expression of this gene was assessed. Additionally, the status of the plasmid was examined to find whether it remains episomal or integrates into the host genome. Many, but not all, studies have reported the safety of using this S/MAR-based plasmid as a vector for gene transfer due to a lack of evidence of genomic integration. However, success in efficiently expressing a desired transgene has not been reported by everyone and has proven to be variable. The following paragraphs outline the studies conducted testing this vector and its applications.

Of the successful experiments conducted with an S/MAR-based vector where expression proved to be unsilenced is that conducted by Manzini *et al*. The aim of this experiment was to create transgenic pigs expressing the eGFP driven by the CMV promoter using the S/MAR plasmid as a vector (Manzini *et al*., 2006). The vector was transferred into sperm using Sperm-Mediated Gene Transfer (SMGT). The sperm was then used to fertilise sows' eggs by laparoscopic insemination. The tissues were collected from 18 fetuses from 2 different sows at day 70 of the pregnancy, where cell differentiation and organ development was complete and only growth (embryogenesis) remained to take place during the rest of the pregnancy. The negative control in this experiment was 9 fetuses that were produced from sperm that was not incubated with the exogenous DNA. Tissue samples were taken from skeletal muscle, heart, liver, kidney, and lung from each fetus. A PCR conducted on total genomic DNA extracted from the tissues of the fetuses confirmed that the vector sequence was present in 12 of the 18 fetuses in every tissue analysed. A PCR on the HIRT extracts of these samples indicated that the vector was present as an episome in each of these tissues. A Southern Blot was conducted using the total genomic DNA and the HIRT extracts from each tissue of 4 of the positive foetuses. The DNA band patterns were the same in each case and showed episomal retention of the vector with no evidence of integration into the host genome. From the band intensity the copy number was estimated to be less than 10 copies per cell. A plasmid rescue on HIRT extracts of 2

tissues from each of the 4 positive foetuses supported the episomal status of the vector. Restriction digests of the plasmids grown up after the rescue experiment confirmed that the vector was episomal and had not been altered. Expression was assessed by RT PCR in 3-5 tissues of each of the 12 positive foetuses. It was found that the transcript could be detected in each tissue tested for 9 of the foetuses. Expression was also detected by Western Blotting, epifluorescence, and confocal microscopy. They found expression of eGFP in 62-88% of the cells from each sample, and these percentages did not differ significantly between different foetuses or different tissues. These findings strongly suggest no integration had occurred, and Manzini *et al* believe that if it does occur it would be uncommon and sporadic and would take place at the embryogenesis stage of pregnancy, which is the stage after the foetuses were taken out and analysed in this experiment (Manzini *et al*., 2006).

Another successful experiment was undertaken by Jenke *et al* (Jenke *et al*., 2005). In this study the aim was to use a vector containing an S/MAR and carrying a gene coding for a short hairpin RNA (shRNA) controlled by an RNA polymerase III promoter. shRNA is used to suppress gene expression and this study uses the bcr-abl model system to silence the expression of a fusion protein, the gene of which is created as a result of a chromosomal rearrangement, which causes leukemic transformation (Daley *et al*., 1990). This system was tested in K562 cells which are bcr-abl positive and expression was assessed using Western and Northern Blotting. The cells were tranfected with the vector, put under selection for 14 days, then allowed to proliferate for 28 days. As a positive control the vector was linearised and transfected into the cells and was found by Southern Blotting to have integrated. The results showed a significant reduction in the growth of the cells containing the vector episomally, meaning the vector was expressing the shRNA that was successfully inhibiting the translation of the bcr-abl fusion transcript into the protein that leads to cellular transformation and thus overgrowth. However the cells where the vector had integrated in the control did not show a significant difference in comparison to the negative control which contained untransfected cells, possibly due to the expression cassette being deleted as a result of the integration, or silencing possibly as a result of integration position. On the other hand, it is possible that the cells that had the vector integrated and were expressing the shRNA effectively were outgrown by those that did not have the vector sequence or were not expressing it effectively and therefore were unable to be identified. The

cells were tested again after 4 months and were found to still retain the vector episomally, and the effect on gene silencing had continued to be effective and unchanged.

A study by Schaarschmidt *et al* showed similarly positive results in the use of this S/MAR-based vector (Schaarschmidt *et al*., 2004). The study involved the transfection of CHO and HeLa cells with an S/MAR vector containing an eGFP reporter gene driven by the CMV promoter. The control was the same vector without the S/MAR. Following a period of selection post transfection a number of positive clones were picked for each cell line and the cells were allowed to proliferate. Southern Blotting on HIRT extracts confirmed that 9/10 CHO cell clones and 5/5 HeLa clones were positive for the episome. All the control clones using the control plasmid in the positive CHO and HeLa clones were negative for the episome, and the Southern Blot for these samples showed integration. In a previous study conducted by Baiker *et al* the same vectors were used to transfect CHO cells and the same results had already been achieved (Baiker *et al*., 2000). They had additionally found that these cells maintained an average of 4-13 copies per cell with little variation between 20-50 generations after transfection.

While some studies showed sustained expression and maintenance of these S/MAR-based vectors over many cell generations, others showed variable results in terms of the S/MAR vector's expression of a given transgene. A study by Papapetrou *et al* examined the use of the S/MAR vector expressing the eGFP transgene driven by the CMV promoter in several hematopoietic progenitor cells to assess its expression and episomal status generations after transfection (Papapetrou *et al*., 2006). The plasmid was first tested in two progenitor cells lines: K562 which is a human chronic myeloid leukemia blast crisis cell line, and MEL which is a murine erythroleukemia cell line. A population of cells were transfected and and selected with G-418 for 3 weeks. Then each population was divided into two populations, one where selection was maintained and the other where it was removed, and the cells were allowed to proliferate for over 100 generations. The cells of each population remained G-418 resistant and eGFP expression declined only slightly with time. A non-radioactive Southern Blot using 10µg of total genomic DNA showed a single band of the expected plasmid size which suggested episomal status, and this was further confirmed by plasmid rescue. The vector was retained by less than 1% of the cells, which was

calculated by the initial number of cells transfected divided by the number of positive clones established after the selection period. The average plasmid copy number per cell in the K562 population was 2.5. Similar results were obtained using another population of primary fibroblast-like human cells that were isolated from umbilical cord blood where the cells were allowed to proliferate for 30 generations.

The MEL cells, on the other hand, lost eGFP expression by 7 days post-transfection and it could not be detected by FACs or immunofluorescence as it had been with the K562 cells. The use of the more sensitive method RT-PCR showed that eGFP was still being expressed but at very low levels. When $CD34^+$ enriched cells derived from umbilical cord blood were tested the plasmid was not detected using Southern Blotting or plasmid rescue, even when the cells were expanded and $1x10^7$ cells were used for the HIRT extract. The episome was only detected using PCR, and the vector was calculated to only be present in approximately 1% of the cells. Despite the similarity observed between the different cell types where only an average of 1% of the cells contained the vector as an episome, transgene expression was undoubtedly variable. Papapetrou *et al* conclude that cosidering S/MAR elements utilize cellular factors which allow the vector containing them to express transgenes and replicate, it would be logical to assume that the differences found in cellular backgrounds from one cell type to the next could justify the differences observed of the same vector's status, expression and/or episome copy numbers between different cell lines (Papapetrou *et al*., 2006). These could be factors specific to different tissues or between species. However, even though expression in the MEL cells did decline it did not affect vector maintenance and passing down to daughter cells.

Thus far all the studies have reported positive results in their assessment of the S/MAR-based vector being replicated and passed down to daughter cells for many generations. However, one has reported a different conclusion. In a study by Argyros *et al*, the S/MAR vector was tested *in vivo* in mouse liver. It contained the luciferase transgene and was driven by either the CMV promoter or the human AAT promoter (Argyros *et al*., 2008). The vectors were delivered by hydrodynamic injection, and three weeks post transduction hepatectomies was performed on the livers. The livers were allowed to regain their original mass before the animals were sacrificed and the liver cells examined for the presence of the vectors. It was found that expression for all the plasmids had declined to near background levels except for the S/MAR plasmid

containing the AAT promoter which was slightly higher than the other three constructs after 6 months. Southern Blotting and Plasmid Rescue experiments confirmed that the plasmids had remained episomal. However, they discovered that the vector had not been replicated within these cells. Bacterial methylation on a plasmid is lost after being replicated twice in mammalian cells. Restriction digest tests using enzymes that can cut only at sites where the DNA is unmethylated using vector DNA extracted from the liver 6 months post-transduction indicated that the plasmid was not replicated within these cells in the duration of this period. Argyros *et al* concluded that in murine hepatocytes the vectors were able to remain passively but were not being actively replicated. They also concluded that the decline in expression observed over a long period of time was due to silencing of the promoter rather than the loss of plasmid DNA, as just transducing the liver cells without hepatectomies showed that the S/MAR construct, although not replicating episomally, was passively maintained and was able to uphold sustained expression for 6 months.

The studies outlined have shown that an S/MAR-based vector can behave differently as a result of many different factors such as the transgene expressed, the promoter driving it, or the cell type in which it is used which appears to be a major factor. This may be due to species origin or cell/tissue-specific factors which would cause differences in the S/MAR vector's episomal replication and/or expression.

1.11 S/MAR Database and Alternative S/MARs

The S/MAR that has been used in most of the studies to date, where the element has been inserted within a vector for gene delivery, is the human β-IFN S/MAR, or variations of it, and it is the most well characterised S/MAR element to date. But there are other S/MARs that have been identified in different regions of the genome, serving similar insulator/enhancer functions. These S/MARs have been collected in an online database (http://smartdb.bioinf.med.uni-goettingen.de/) and are defined as S/MARs on the basis of several essential characteristics previously discussed, such as containing a base-unpairing region (BUR), the ability to bind to the nuclear matrix, to act as an origin of replication, and to bind S/MAR-binding proteins previously characterised such

as SAF-A or SATB1. The database also provides the references of the studies conducted on each putative S/MAR that led it to be characterised as such an element.

1.11.1 The *c-MYC* Proto-oncogene S/MAR Element

One such S/MAR element that was identified through the database is the murine c-MYC gene S/MAR. It was identified as an S/MAR from the studies conducted by Girard-Reydet and Cai *et al* (Girard-Reydet *et al*., 2004; Cai *et al*., 2003). Girard-Reydet *et al* found that in mice the region upstream of the *c-myc* promoter contains a matrix attachment region, proven by digesting the genomic DNA with DNase I, purifying the nuclear matrix, extracting the DNA still bound to the matrix, radiolabeling it, then using it as a hybridisation probe in order to identify the S/MAR region (Girard-Reydet *et al*., 2004). In this study they also found that an origin of replication site exists within this S/MAR region. Cai *et al* found that the nuclear protein SATB1 (Special AT-rich Binding Protein) binds to the MYC gene locus (Cai *et al*., 2003). This protein has a high affinity to double stranded BUR DNA which are, in themselves, AT-rich regions of DNA. These studies confirmed the presence and site of the *c-myc* S/MAR.

The *c-myc* gene itself is a proto-oncogene whose product is a nuclear transcription factor (Bruckert *et al*., 2000). When mutation occurs this proto-oncogene has been identified as the most frequently amplified, and therefore over-expressed, gene in patients with Acute Myeloid Leukaemia (AML) (Bruckert *et al*., 2000). Yet its over-expression alone has not been found to induce transformation and only leads to tumorigenesis if expressed in addition to other oncogenes (Cole, 1986). In AML it is generally amplified late in tumorigenesis and is often found as a Double Minute (DM) (Bruckert *et al*., 2000). DM's are extrachromosomal pieces of DNA made of chromatin, just as normal chromosomes, but lack a centromere, telomeres, and are found as circular pieces of DNA (Barker, 1982; Thomas *et al*., 2004). In a review of 33 cases of AML, 76% of the cases were found with c-MYC double minutes (Barker, 1982; Thomas *et al*., 2004). This ability to amplify and exist as a double minute may be related to the c-MYC's S/MAR element that allows the propagation of this amplified circular piece of DNA that needs neither a centromere nor telomeres to replicate within tumour cells. As it has been seen that S/MAR elements may act as origins of replication themselves, perhaps it is possible that the S/MAR is the element that is keeping the over-amplified

c-myc region a double minute and allowing expression. These characteristics would make the c-MYC S/MAR an interesting alternative S/MAR to test in replicating cells.

1.12 Aims and Objectives

S/MAR-based plasmid vectors are proposed to lead to improved long-term transgene expression and episomal maintenance of vectors in actively replicating cells, and to protect from vector integration into the host genome. It is also proposed that arresting transfected cells containing the S/MAR-based vectors prior to proliferation improves long-term transgene expression as a direct result of the presence of the S/MAR element.

The main aim of the work presented in this thesis was to investigate the use of S/MAR-based vectors in expressing a transgene and in maintaining the vector episomally for many cell generations in proliferating muscle cells, the comparison of these parameters to its non-S/MAR counterpart control, and the enhancement of the S/Mar based vector in order to further its development for use as a safe and effective vector for gene delivery. Several secondary aims were conducted to complement the main aim of the thesis. These included:

- The assessment of these vectors in different cell lines.
- The assessment of these vectors in myotubes.
- The improvement of transgene expression by means of changing the promoter driving transgene expression.
- The improvement of long term transgene expression by arresting cells in G0/G1 phase of the cell cycle, or quiescence, post-transfection.
- The investigation of the use of other S/MAR elements, besides the β-IFN S/MAR element, in plasmid vectors.
- The construction of the ideal minicircle vector including an S/MAR element.

The objectives for each chapter were created in order to fulfil the aims.

Chapter 3 Objectives:

1) The assessment and comparison of eGFP transgene expression by β-IFN S/MAR-based vector pEPI-eGFP in 3 replicating cells lines C2C12, HepG2, and HeLa in order to elucidate long-term transgene expression and the differences of this expression between different cell lines. The C2C12 cell line was chosen as the skeletal muscle cell line, the HepG2 cell line was chosen as it represented a different tissue type and had a longer doubling time than the other two cell lines chosen, and the HeLa cell line was chosen as a control due to the assessment of pEPI-eGFP within this cell line in previous studies.
2) The comparison of initial and long term eGFP transgene expression between S/MAR-based vector pEPI-eGFP and its non-S/MAR counterpart control vector eGFP-C1 in the three cell lines chosen.
3) The assessment of episomal/integrant status of the transfected vectors pEPI-eGFP and eGFP-C1 at the final day of the experiment in the three cell lines chosen.
4) The assessment of initial and long term eGFP transgene expression from a 'mini'- β-IFN-S/MAR-containing vector pEPI-M18 as compared to pEPI-eGFP and eGFP-C1, and the investigation of its episomal/integrant status, in C2C12 cells.
5) The assessment of long term eGFP expression by pEPI-eGFP, eGFP-C1, and pEPI-M18 in C2C12 cells differentiated into myotubes.

Chapter 4 Objectives:

1) The cloning of vectors CeGFP-C1 and CpEPI-eGFP, whereby the CMV promoter was replaced by the CAGG promoter in the eGFP expression cassette within each construct.
2) The cloning of vector CMYC-pEPI in order to create a vector with an eGFP expression cassette including the c-myc S/MAR element within its open reading frame.
3) The cloning of an eGFP expression cassette including the β-IFN-S/MAR element into a minicircle-producing parental plasmid, in order enable the creation of minicircles.

Chapter 5 Objectives:

1) The assessment of initial and long-term eGFP transgene expression of the S/MAR-based vector and the non-S/MAR control vector using a different promoter (CAGG replaced CMV) to improve transgene expression by these vectors (CpEPI-eGFP and CeGFP-C1) in C2C12 cells.
2) The assessment of long term eGFP expression by CpEPI-eGFP and CeGFP-C1 in C2C12 cells differentiated into myotubes
3) The investigation of episomal/integrant status of CpEPI-eGFP and CeGFP-C1 at the final day of the experiment.

Chapter 6 Objectives:

1) To use the data that was obtained from testing pEPI-eGFP in the HepG2 and HeLa cell lines in Chapter 3 to enhance long term transgene expression in C2C12 cells.
2) The assessment of long term eGFP expression in C2C12 cells after the cell cycle arrest in G0/G1/quiescence of CpEPI-eGFP and CeGFP-C1 transfected C2C12 cells for the maximum amount of time possible, without resulting in total cell death, using two different methods. Transgene expression was compared between the two vectors, and between the two methods used for cell cycle arrest.
3) The assessment of long term eGFP expression by CpEPI-eGFP and CeGFP-C1 arrested in G0/G1/ quiescence in C2C12 cells differentiated into myotubes
4) The investigation and comparison of episomal/integrant status of the vectors in all transfected populations on the final day of the experiment.

Chapter 7 Objectives:

1) The assessment of the requirement for the use of antibiotic selection for enhanced trangene expression after cell cycle arrest was conducted post-transfection with CpEPI-eGFP and CeGFP-C1 in C2C12 cells by the comparison

of long-term eGFP transgene expression between transfected populations put under four different conditions:

a) Transfected eGFP positive cells were sorted by FACs then allowed to proliferate.
b) Transfected eGFP positive cells were sorted by FACs, arrested in G0/G1 for 7 days, then allowed to proliferate.
c) Transfected cells were arrested in G0/G1 for 7 days, selected with antibiotics for 10 days, then allowed to proliferate.
d) Transfected cells were selected with antibiotics for 10 days then allowed to proliferate.

2) The assessment of initial and long-term eGFP transgene expression from a C-MYC S/MAR-containing vector (CMYC-pEPI) and the comparison of eGFP transgene expression resulting from vectors CpEPI-eGFP and CeGFP-C1 under all 4 conditions (a)-(d) as stated above.
3) The investigation of episomal/integrant status of the transfected vectors CMYC-pEPI, CpEPI-eGFP, and CeGFP-C1 in C2C12 cells at the final day of the experiment under all 4 conditions (a)-(d) as stated above.

Chapter 2: Materials and Methods

2.1. Materials

2.1.1. General Buffers and Chemical Reagents

- Boric Acid (Sigma, B6768)
- Dextran sulphate (Sigma, 8906)
- EDTA (Ethylenediaminetetraacetic acid) (Sigma, E-5134)
- Ethanol (Analar, VWR, 10107)
- Glacial acetic acid (Sigma, A9967)
- Glycerol (Sigma, G-2025)
- Isopropanol (propan-2-ol, Analar, VWR, product number 102246L)
- KCl (Sigma, P9541)
- Methanol (Sigma, 322415)
- Na_3citrate (Sigma, S1804)
- NaCl (Sigma, S-7653)
- NaOH (Sigma, S8045)
- Phosphate Buffered Saline (PBS) (PBS tablets, Dulbecco A, OXOID, BR0014G)
- Trizma Base (tris[hydroxymethyl]aminomethane) (Sigma, T-1503)
- Tween 20 (polyoxyethylenesorbitan monolaurate, Sigma, P-1379)

2.1.2. Cell Culture

2.1.2.1. Cell Types

The murine and human derived cell lines used in this thesis are described below:

C2C12 cell line: A myogenic subclone of an adherent skeletal muscle cell line of murine origin (Yaffe and Saxel, 1977). These cells can differentiate rapidly to form contractile myotubes and produce characteristic muscle proteins. This cell line is tumorigenic and variably polyploid.

HeLa cell line: An adherent cervical epithelial cell line derived from an adenocarcinoma and of human origin, first described by Scherer (Scherer, 1954). Its morphological aspects are described in Macville *et al* (Macville *et al*., 1999).

HepG2 cell line: An adherent hepatocellular cell line derived from a hepatocellular carcinoma and of human origin (Aden *et al*., 1979).

Further details of the features of the above cell lines can be found at the European Collection of Animal Cell Cultures (ECACC) www.ecacc.org.uk and the American Type Culture Collection (ATCC) http://www.lgcstandards-atcc.com.

2.1.2.2. General Cell Culture Materials and Reagents

- 6 well dishes (Nunc, 140675)
- Cell Culture Flasks (175 cm^2) (Nunc, 178883)
- Cell Culture Flasks (80 cm^2) (Nunc, 178905)
- Cell lines: C2C12, HeLa, HepG2 (ATCC: http://www.lgcstandards-atcc.com)
- Cryotubes (nunc, 366656)
- Dimethyl Sulphoxide (DMSO) (Sigma, D2650)
- Dulbecco's Modified Eagle Medium (DMEM) (Sigma, D2650)
- Foetal Calf Serum (FCS) (Sigma, F7524)
- Gentamicin (Sigma, G-1397)
- Horse Serum (Sigma, H1270)
- L-Glutamine 100x (200mM) (Sigma, G7513)
- Lipofectamine 2000 (Invitrogen, 11668019)
- Methionine-Depleted Medium (Sigma, D0422)
- Penicillin Streptomycin (10mg) (Sigma, P4333)
- Trypsin/EDTA 10x (Sigma, T4049)

2.1.3. Bacterial Culture

2.1.3.1. General Bacterial Culture Materials and Reagents

- Ampicillin (Sigma, A9518)
- Agarose (Invitrogen, 15510-027)
- Bactoagar (Beckton, Dickinson and Co, 214010)
- Chloramphenicol (Sigma, C0378)
- Electroporation cuvettes, 0.1cm (Invitrogen, P41050)
- Ethidium Bromide (Sigma, E-1510)
- Kanamycin (Sigma, 60615)
- LB Broth (Sigma, L-3022)
- LB SOC provided with OneShot TOP10 Electrocomp *E.Coli* (Invitrogen, C404052)
- NTP mix (Bioline, BIO-39050)
- OneShot TOP10 Electrocomp *E.Coli* (Invitrogen, C404052)
- Plasmid Maxiprep Kit (Endofree) (Qiagen, 12362)
- Plasmid Miniprep Kit (Qiagen, 27104)
- QIAquick Gel Extraction Kit (250) (Qiagen, 28706)
- Quick Ligation Kit (New England Bioloabs, M2200S)
- Restriction Enzymes (New England Biolabs)
- T4 DNA Polymerase (Promega, M421A)
- XL-1 Blue supercompetent Cells *E.coli* (Stratagene, 200236)
- OneShot TOP10 Electrocomp *E.coli* (Invitrogen, C404052)
- JM110 Competent (Dam/Dcm negative) Cells *E.coli* (Stratagene, 200239)

2.1.3.2. Plasmid Vectors Recieved/Purchased

pBACe3.6 RPCI Mouse BAC clone: containing cMYC S/MAR element (Sanger Institute, Cambridge, UK)

peGFP-C1: A 4.7kb eukaryotic expression construct containing the human immediate early cytomegalovirus (CMV) promoter driving the enhanced green fluorescent protein (eGFP) reporter gene terminated by the SV40 polyadenylation signal, a multiple cloning site (MCS) containing restriction sites, the pUC origin of replication for vector propagation in *E.coli*, the SV40 origin of replication for vector propagation in mammalian cells expressing the SV40 T-antigen, a bacterial promoter to drive the expression of the Tn5 Kanamycin resistance gene within *E.coli*, the early SV40

promoter driving the expression of the Neomycin resistance gene within mammalian cells terminated by the Herpes Simplex Virus thymidine kinase (HSV TK) polyadenylation signal (purchased from Clontech).

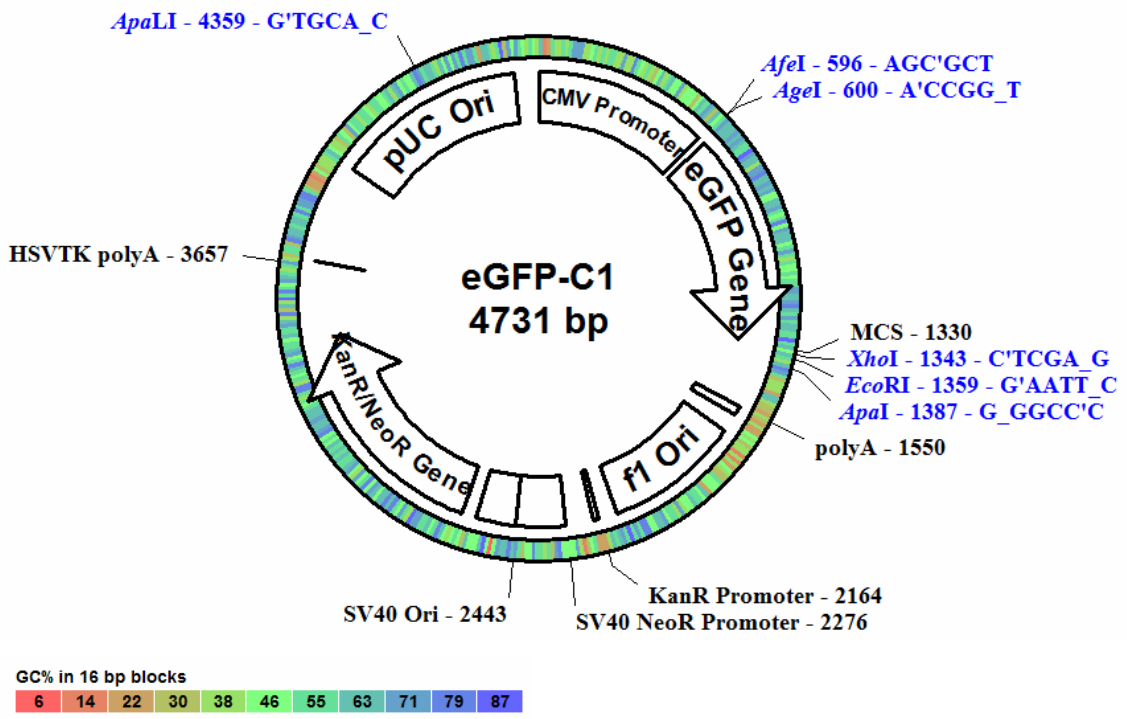


Figure 2.1 Vector map of eGFP-C1 construct.

pEPI-eGFP: A 6.7kb eukaryotic expression construct containing the human immediate early cytomegalovirus (CMV) promoter driving the enhanced green fluorescent protein (eGFP) reporter gene terminated by the SV40 polyadenylation signal, a multiple cloning site (MCS) containing restriction sites, the pUC origin of replication for vector propagation in *E.coli*, the SV40 origin of replication for vector propagation in mammalian cells expressing the SV40 T-antigen, a bacterial promoter to drive the expression of the Tn5 Kanamycin resistance gene within *E.coli*, the early SV40 promoter driving the expression of the Neomycin resistance gene within mammalian cells terminated by the Herpes Simplex Virus thymidine kinase (HSV TK) polyadenylation signal, and the β-IFN S/MAR element. As indicated the S/MAR element contains a high A-T content and 5 polyadenylation signals within it (kindly donated by J. Bode).

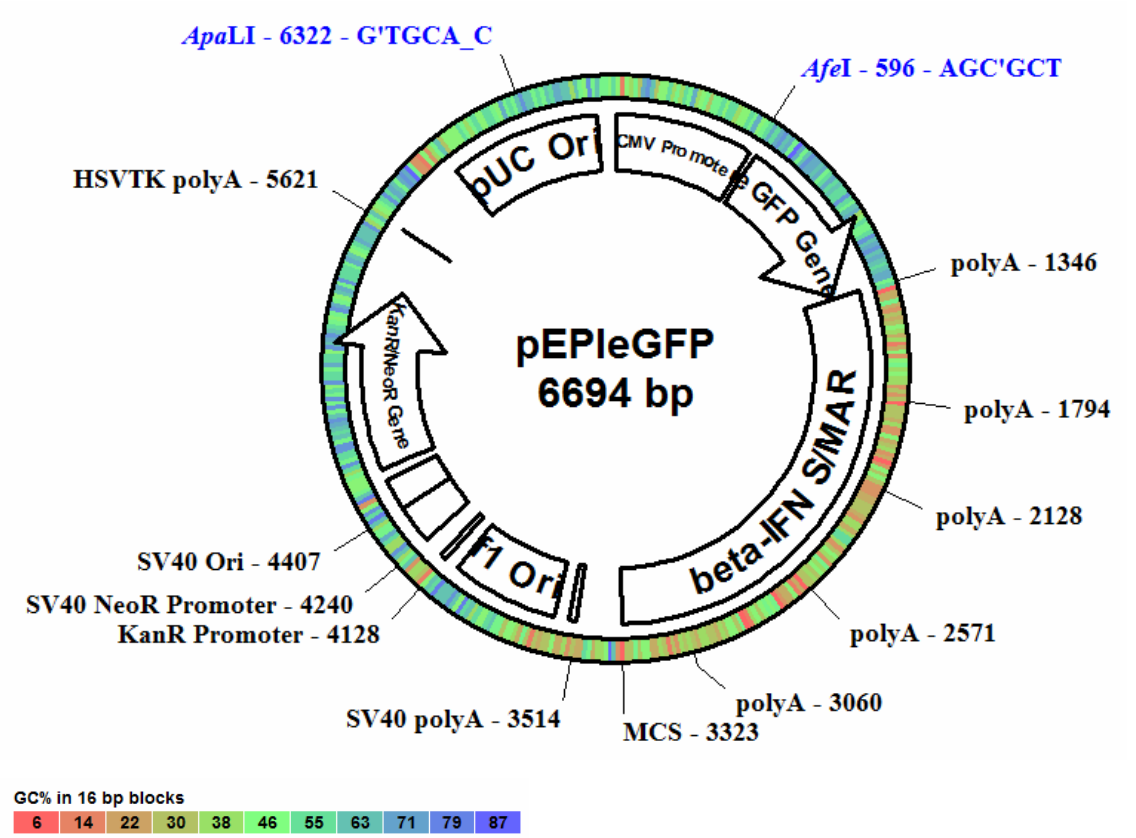


Figure 2.2 Vector map of pEPI-eGFP construct.

pEPI-M18: A 5.4kb eukaryotic expression construct containing the human immediate early cytomegalovirus (CMV) promoter driving the enhanced green fluorescent protein (eGFP) reporter gene terminated by the SV40 polyadenylation signal, a multiple cloning site (MCS) containing restriction sites, the pUC origin of replication for vector propagation in *E.coli*, the SV40 origin of replication for vector propagation in mammalian cells expressing the SV40 T-antigen, a bacterial promoter to drive the expression of the Tn5 Kanamycin resistance gene within *E.coli*, the early SV40 promoter driving the expression of the Neomycin resistance gene within mammalian cells terminated by the Herpes Simplex Virus thymidine kinase (HSV TK) polyadenylation signal, and the 'mini-S/MAR' (733bp) element derived from the β-IFN S/MAR element (kindly donated by J. Bode).

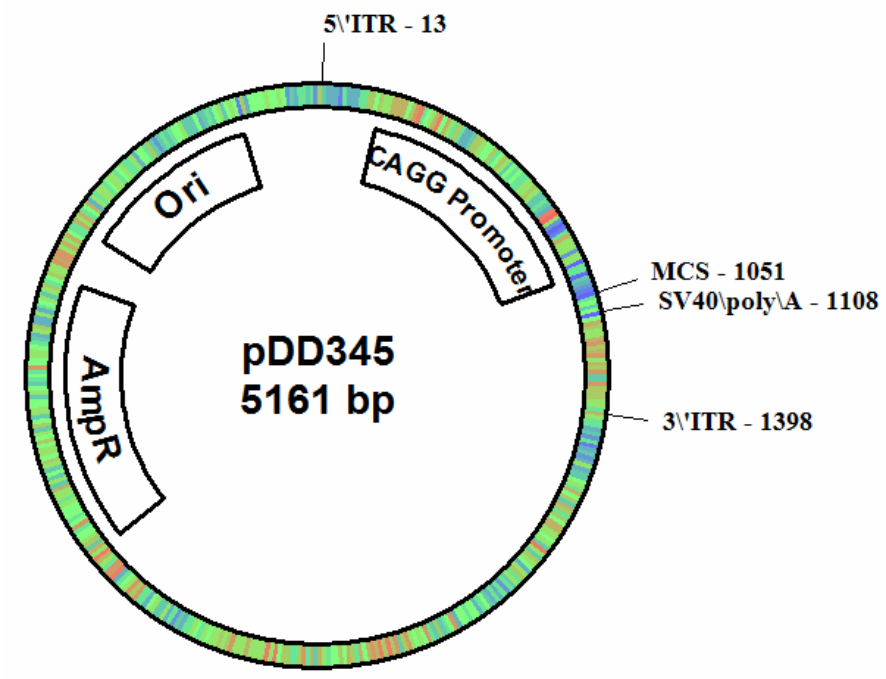


Figure 2.3 Vector map of pDD345 construct.

pDD345: A 5.1kb vector containing the Adeno-Associated Virus (AAV-2) type 2 wild-type cis acting terminal repeats (5' and 3' ITRs) cloned into the pBR322 plasmid backbone. Between these terminal repeats are a copy of the CAGG promoter, which does not drive the expression of any transgene, a multiple cloning site (MCS), and a copy of the SV40 polyadenylation sequence. The remainder of the vector contains an expression cassette expressing the Ampicillin resistance gene, and an origin of replication from pMB1 for vector propagation in *E.coli.*

2.1.3.3. Plasmid Vectors Constructed within RHUL

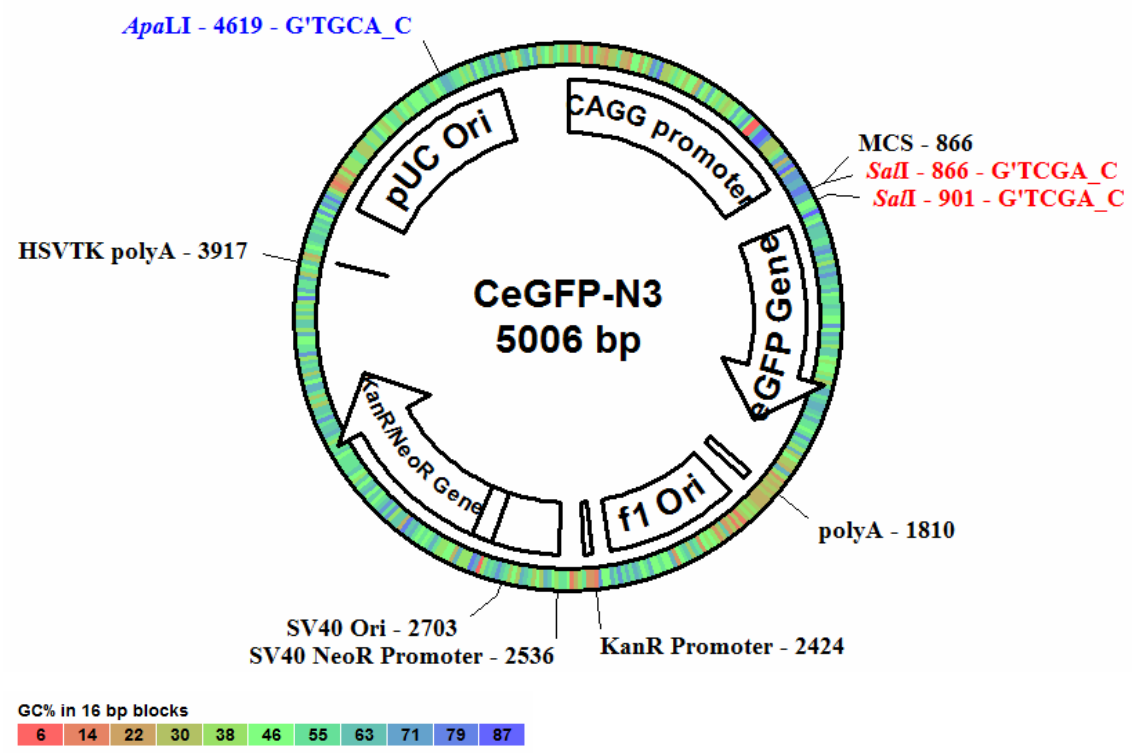


Figure 2.4 Vector map of CeGFP-N3 construct.

CeGFP-N3: A 5kb eukaryotic expression construct containing the human cytomegalovirus early enhancer/chicken β-actin (CAGG) promoter driving the enhanced green fluorescent protein (eGFP) reporter gene terminated by the SV40 polyadenylation signal, a multiple cloning site (MCS) containing restriction sites, the pUC origin of replication for vector propagation in *E.coli*, the SV40 origin of replication for vector propagation in mammalian cells expressing the SV40 T-antigen, a bacterial promoter to drive the expression of the Tn5 Kanamycin resistance gene within *E.coli*, the early SV40 promoter driving the expression of the Neomycin resistance gene within mammalian cells terminated by the Herpes Simplex Virus thymidine kinase (HSV TK) polyadenylation signal.

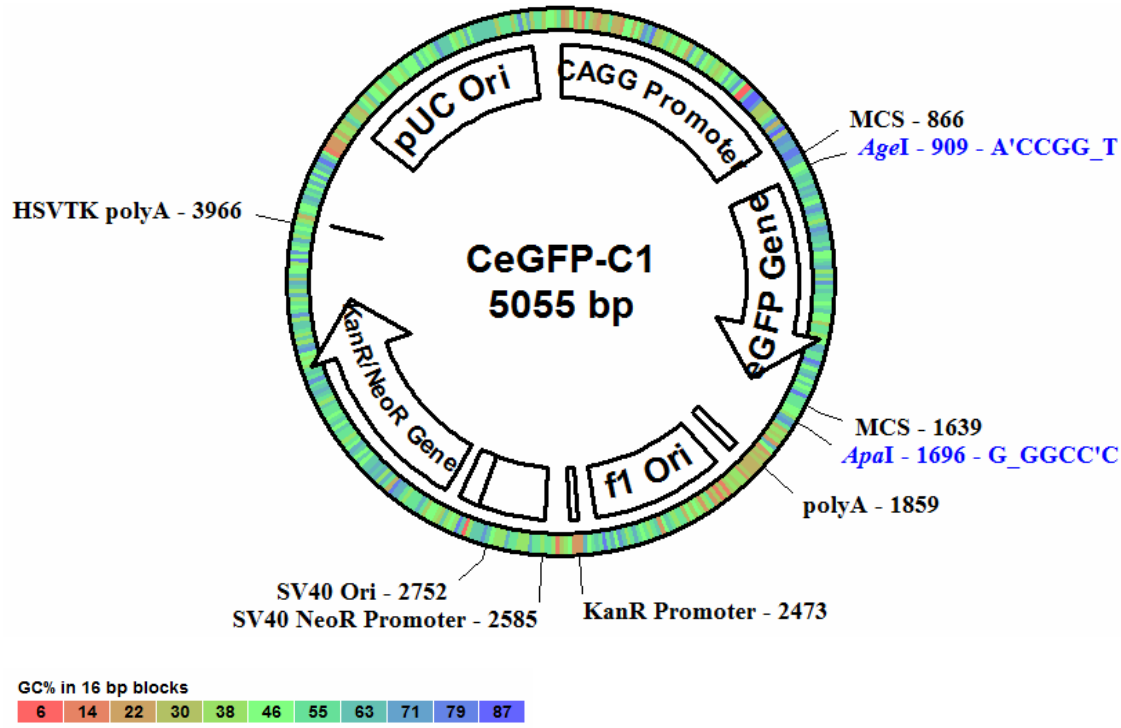


Figure 2.5 Vector map of CeGFP-C1 construct.

CeGFP-C1: A 5kb eukaryotic expression construct containing the human cytomegalovirus early enhancer/chicken β-actin (CAGG) promoter driving the enhanced green fluorescent protein (eGFP) reporter gene terminated by the SV40 polyadenylation signal, a multiple cloning site (MCS) containing restriction sites, the pUC origin of replication for vector propagation in *E.coli*, the SV40 origin of replication for vector propagation in mammalian cells expressing the SV40 T-antigen, a bacterial promoter to drive the expression of the Tn5 Kanamycin resistance gene within *E.coli*, the early SV40 promoter driving the expression of the Neomycin resistance gene within mammalian cells terminated by the Herpes Simplex Virus thymidine kinase (HSV TK) polyadenylation signal.

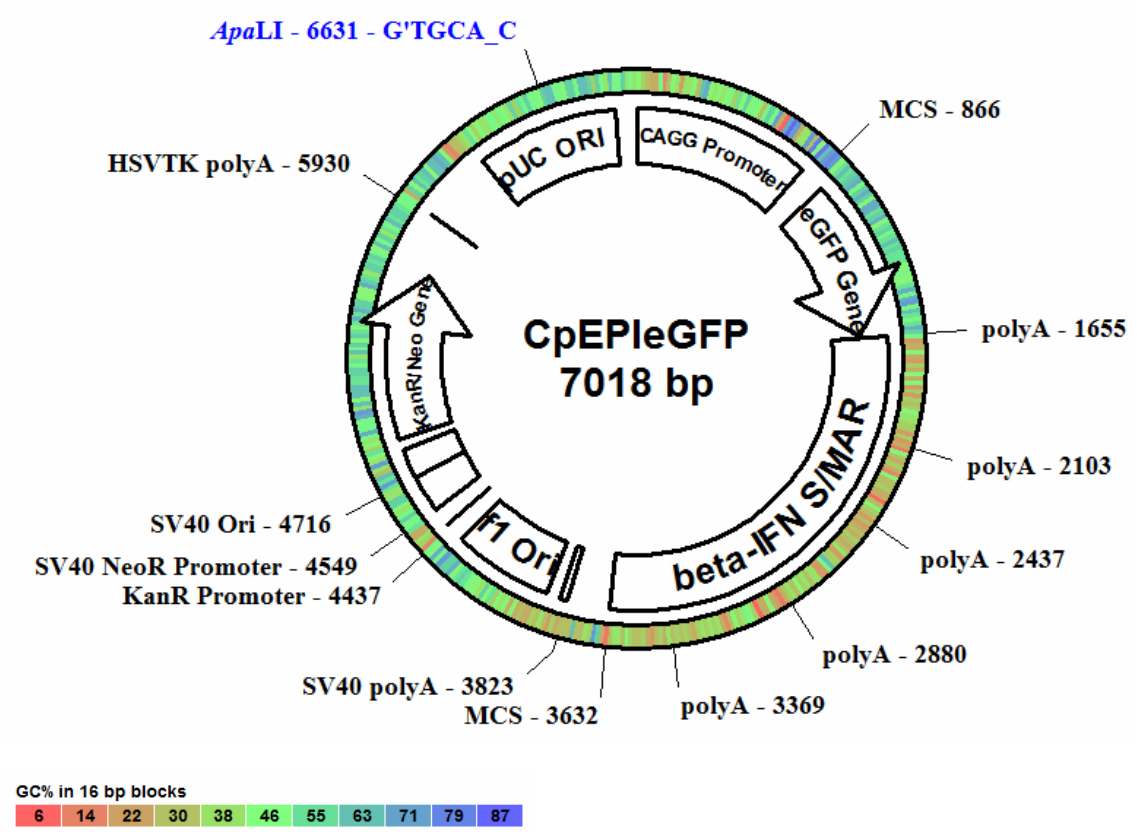


Figure 2.6 Vector map of CpEPI-eGFP construct.

CpEPI-eGFP: A 7kb eukaryotic expression construct containing the human cytomegalovirus early enhancer/chicken β-actin (CAGG) promoter driving the enhanced green fluorescent protein (eGFP) reporter gene terminated by the SV40 polyadenylation signal, a multiple cloning site (MCS) containing restriction sites, the pUC origin of replication for vector propagation in *E.coli*, the SV40 origin of replication for vector propagation in mammalian cells expressing the SV40 T-antigen, a bacterial promoter to drive the expression of the Tn5 Kanamycin resistance gene within *E.coli*, the early SV40 promoter driving the expression of the Neomycin resistance gene within mammalian cells terminated by the Herpes Simplex Virus thymidine kinase (HSV TK) polyadenylation signal, and the β-IFN S/MAR element. As indicated the S/MAR element contains a high A-T content and 5 polyadenylation signals within it.

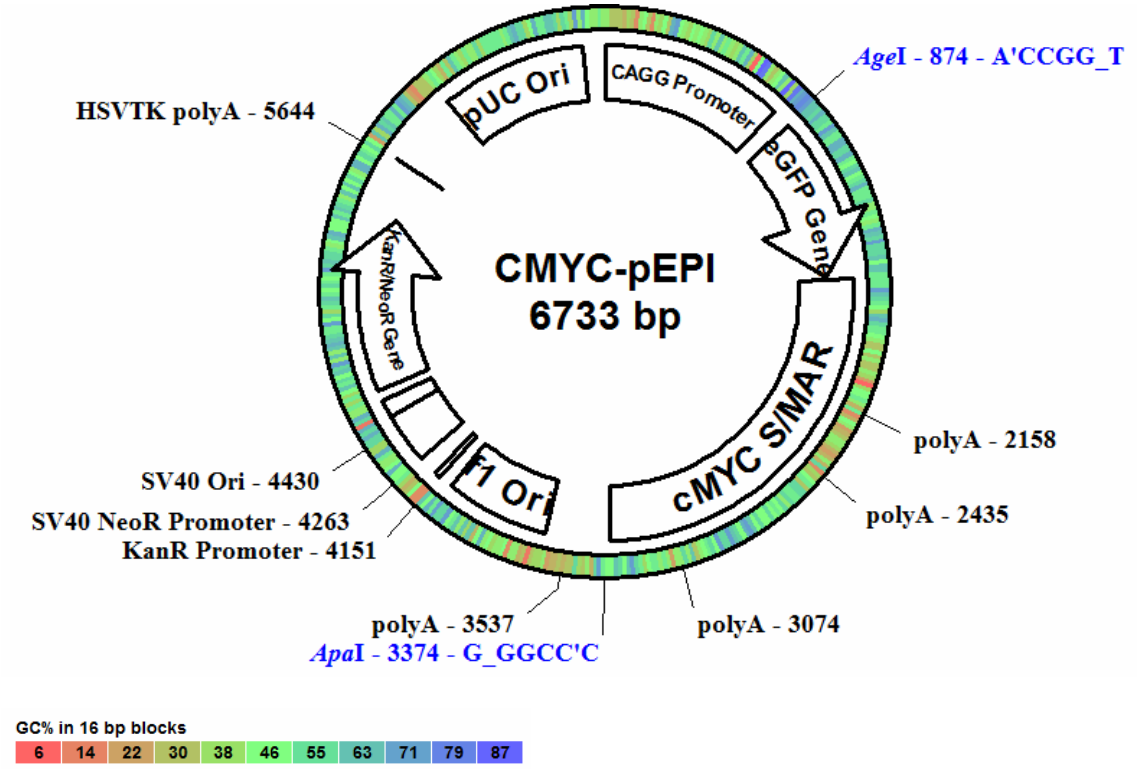


Figure 2.7 Vector map of CMYC-pEPI construct.

CMYC-pEPI: A 6.7kb eukaryotic expression construct containing the human cytomegalovirus early enhancer/chicken β-actin (CAGG) promoter driving the enhanced green fluorescent protein (eGFP) reporter gene terminated by the SV40 polyadenylation signal, a multiple cloning site (MCS) containing restriction sites, the pUC origin of replication for vector propagation in *E.coli*, the SV40 origin of replication for vector propagation in mammalian cells expressing the SV40 T-antigen, a bacterial promoter to drive the expression of the Tn5 Kanamycin resistance gene within *E.coli*, the early SV40 promoter driving the expression of the Neomycin resistance gene within mammalian cells terminated by the Herpes Simplex Virus thymidine kinase (HSV TK) polyadenylation signal, and the cMYC S/MAR element. the region spanning a part of the cMYC S/MAR element has a very low GC content. It is also shown that the cMYC S/MAR element contains 3 polyadenylation signals within its sequence.

2.1.4. Immunostaining

2.1.4.1. Antibodies

Primary Antibodies:

- Mouse monoclonal anti-GFP antibody (Abcam, ab1218)
- Rabbit polyclonal anti-lysenin (Peptide Institute Inc; Osaka, Japan)

Secondary Antibodies:

- Alexa Fluor 488 goat anti-mouse IgM antibody (Invitrogen, A21042)
-488 swine anti-rabbit IgG (SouthernBiotech, 6310-01)

2.1.4.2. General Immunostaining Materials and Reagents

- 8-well chamber slide (Lab-Tek, 70415)
- Bovine Serum Albumin (BSA) (Sigman, B-4287)
- Lysenin (Peptide Institute Inc; Osaka, Japan)
- Normal Goat Serum
- Paraformaldehyde (Sigma, 158127)
- Triton X-100 (Sigma, T8787)
- DAPI Vectashield mounting medium (Vector Laboratories, H-100)

2.1.5. DNA Extraction Reagent

- DNAzol® (Talron Biotech L.T.D., DN127)

2.1.6. Fluorescent In-Situ Hybridisation (FISH) Reagents:

- DAPI II Counterstain (Abbott Molecular, 06J50-001)
- Demecolcine (Sigma, D7358)
- Formamide (Sigma, F7508)
- Human COT-1 DNA (Abbott Molecular, 06J31-001)

- Nick Translation Kit (including: nick translation enzyme, 10x nick translation buffer, dATP, dCTP, dGTP, and dTTP solutions, nuclease-free water) (Abbott Molecular, 07J00-001)
- NP-40 (Abbott Molecular, 07J05-001)
- Red dUTP (Abbott Molecular, 02N34-050)
- Rubber Cement (Fielders)
- Single stranded DNA from salmon testes (Sigma, D7656)

2.1.7. Polymerase Chain Reaction (PCR) Reagents and Primers:

- 2x PCR MasterMix (GeneSys Ltd.)
- The primers located within the Kanamycin resistance gene, optimal annealing temp 60.6°C:

Forward primer strand (called Kan-F), Tm 58.8°C:

5'-TGTCAAGACCGACCTGTCC-3'

Reverse primer strand (called Kan-R), Tm 57.3°C:

5'-AATATCACGGGTAGCCAACG-3'

- The primers used to amplify the cMYC S/MAR element from the BAC clone, optimal annealing temperature 59.5°C:

Forward primer strand (called cMYC-F), Tm 73.9°C, extra bases added at 5' for efficient restriction digestion of product (green), *XhoI* site (red):

5'-GTATACTCGAGCTTCCCAGAACCTGGAAACC-3'

Reverse primer strand (called cMYC-R), Tm 72.2°C extra bases added at 5' for efficient restriction digestion of product (green), *EcoRI* site (red):

5'-GTATAGAATTCGCTCGGCTGAACTGTGTTCT-3'

2.2. Methods

2.2.1. Cell Culture Methods

2.2.1.1. Culture and maintenance of cells

Growth medium: DMEM, 10% FBS, 20mM L-Glutamine, 1x Penicillin Streptomycin

Cells were grown and maintained at 37°C, 8% CO_2. On reaching approximately 70% confluency cells were passaged. Dishes were washed 1x in PBS containing 0.02% EDTA to remove calcium and allow adherent cells to be released. Cells were then treated with trypsin diluted in PBS/0.02%EDTA, pre-warmed to 37°C, and incubated at 37°C for a maximum of 5 minutes. The detachment of the adherent cells was monitored under the microscope. Once detached, cells were resuspended in 10ml growth medium pre-warmed to 37°C and the solution was spun for 5 minutes at 650g at room temperature to pellet the cells. The supernatant was then discarded, the pellet was resuspended in 1ml fresh growth medium, and the cells were counted and plated at the required density.

2.2.1.2. Frozen Cell Storage

Sub-confluent cultures of cells were trypsinised and counted. For every $2x10^6$ cells, 1ml of growth medium was added. Then, an equal amount of a solution composed of 20% DMSO and growth medium was added, for a final concentration of 10% DMSO. 1ml aliquots were dispensed immediately into cryotubes, put in a Cryo 1° freezing container (Nalgene) containing isopropyl alcohol which provides a -1°C/minute cooling rate, and stored in a -80°C freezer. The next day the cryotubes were transferred to a liquid nitrogen storage tank for long-term storage.

2.2.1.3. Thawing Frozen Cell Stocks

Cryotubes containing frozen cells were immediately immersed in a 37°C water bath for rapid thawing to avoid cell death. Once thawed the cells were transferred to a falcon tube and slowly and gently resuspended in 20ml growth medium pre-warmed to 37°C. The cells were spun for 5 minutes at 650g. The supernatant was discarded, the pellet resuspended in growth medium, and then seeded to the required density. Cells were maintained at 37°C in a humidified atmosphere containing 8% CO_2.

2.2.1.4. Induction of Myogenic Differentiation In C2C12 Cultures

Differentiation medium: DMEM, 1% Horse Serum, 40mM L-Glutamine, 1x Penicillin Streptomycin

C2C12 cells were plated at a confluency of approximately 70%. The cells were washed once the following day with pre-warmed PBS, then switched to pre-warmed differentiation medium. The cells were washed and the medium replaced daily for 5-10 consecutive days.

2.2.1.5. Lipofection: Liposome-Mediated DNA Transfection

C2C12 and HeLa cells were plated at a density of $1x10^5$ cells per $10cm^2$ dish area and HepG2 cells at $5 x10^5$ cells per $10cm^2$ to achieve a confluency of approximately 70% the following day. For a $10cm^2$ area dish, a total of 4µg of plasmid was suspended in 250µl of serum-free DMEM in one sterile eppendorf or falcon tube, and 10µl of Lipofectamine 2000 reagent was suspended in 250µl of serum-free DMEM in another sterile eppendorf or falcon tube. The solutions were left to incubate for 5 minutes before gently mixing the two solutions together and leaving them to incubate for 20 minutes. The growth medium on the cells was replaced with 2ml of fresh pre-warmed medium and the plasmid/lipofectamine solution was then added to the medium. Cells were incubated in a 37°C in a humidified atmosphere containing 8% CO_2 for 24 hours. Negative controls for each experiment consisted of untreated cultures, or cultures treated with DNA or lipid only.

The amount of each plasmid transfected was calculated in order that the same number of plasmid molecules of each plasmid was transfected into each culture rather than

the same amount (μg), considering the plasmids tested were of different sizes. Plasmid DNA whose sequences had no resemblance to the plasmids being tested, in order to avoid recombination, was used to keep the total amount (μg) of DNA being transfected in each transfection reaction the same (plasmid pDD345).

2.2.1.6. G-418 Sensitivity Curves

C2C12 and HeLa cells were plated in triplicates at $1x10^5$ cells per $10cm^2$ dish area and HepG2 cells at $5\ x10^5$ cells per $10cm^2$. After 24 hours Geneticin (G-418) was added to the growth media at concentrations ranging between 0.1mg-1mg G-418/ml. G-418/growth media was replaced daily for 7 days. At day 7 the cells were trypsinised and pelleted and resuspended in 1ml PBS. 0.1ml of a 0.4% trypan blue stock solution was added to the cell suspension and gently mixed. An aliquot of cells was examined under a phase contrast microscope at 5x magnification and the number of non stained viable cells was counted. The lowest concentration at which no viable cells remained by day 7 was the concentration used for selection. This was determined as 0.6mg/ml G-418 for the HeLa and C2C12 cell lines and 0.8mg/ml G-418 for the HepG2 cell line.

2.2.1.7. C2C12 Cell Arrest In G0 Phase of Cell Cycle

Low serum methionine-depleted medium: Methionine-depleted DMEM, 2% FBS, 20mM L-Glutamine, 1x Penicillin Streptomycin

Cells were plated at $5x10^4$ per $10cm^2$ dish area in low serum methionine-depleted medium. The cells were washed once in pre-warmed PBS before changing the medium daily for 7 consecutive days.

2.2.1.8. Reserve Cell Isolation

This method was conducted as in the studies by Kitzman *et al* and Yoshida *et al* (Yoshida,N. 1998; Kitzmann,M. 1998). C2C12 cells were differentiated for 10 days in differentiation medium. Myotubes, which are more sensitive to trypsinisation than myoblasts, were trypsinised with 0.15% trypsin and washed with PBS repeatedly until all the myotubes had detached and only mononuclear cells had remained adherent to

the flask, as determined by observation under a phase contrast microscope. The adherent cells were then trypsinised and plated into fresh flasks in growth medium in order to reactivate the quiescent reserve cells.

2.2.2 Bacterial Culture and Molecular Cloning

2.2.2.1. Preparation of selective agar

LB broth with 1% agar was sterilised by autoclaving. The solution was then allowed to cool to 40°C and antibiotics were added under sterile conditions and the solution was mixed well. For Ampicillin and Kanamycin plates, the final antibiotic concentration was 50µg/ml. For Tetracycline the required final concentration was 10µg/ml, and for Chloramphenicol it was 60µg/ml. The agar was then poured onto plates under sterile conditions and the agar was allowed to set. The plates were stored at 4°C.

2.2.2.2. Bacterial streaks

A small sample of a single bacterial colony was collected from a fresh culture plate with a flame sterilised inoculation loop and dispersed in 1 ml of LB broth. The loop was again flame sterilised, dipped in the broth, and used to streak a loop full of bacterial culture across the surface on one edge of a LB agar plate containing the appropriate antibiotic. Further streaks were then made along adjacent regions of the plate. The plate was then incubated upside down at 37°C. Plates were sealed by parafilm and stored at 4°C, where the bacteria remained viable for approximately 2-4 weeks.

2.2.2.3. T4 DNA Polymerase mediated conversion of 5' overhangs to blunt ended termini

T4 DNA Polymerase from Promega was used to fill the 5' protruding termini with dNTPs of DNA fragments digested by restriction enzymes in order to generate blunt ended DNA fragments. 5 units of T4 DNA Polymerase/µg was added to 1X T4 DNA Polymerase Buffer as supplied by Promega and 40µM of dNTPs. Alternatively, if the

restriction digest buffer was compatible with the function of the T4 DNA Polymerase, the enzyme was added directly to the restriction digest reaction after the digestion incubation had terminated, in addition to 40µM of dNTPs. The reaction was incubated at 37°C for 5 minutes and was stopped by heating at 75°C for 10 minutes.

2.2.2.4. Ligation of a DNA insert into a plasmid vector backbone

Insert and backbone DNA fragments to be ligated were run on 0.8-2% agarose-EtBr containing gels, visualised, and excised with a sterile scalpel under low UV radiation. The fragments were extracted from the gels and purified using Qiagen's QiaQuick Gel Extraction Kit according to the manufacturer's protocol. The fragments were typically dissolved in 10-20µl of sterile ddH_2O and kept on ice or frozen at -20°C. The concentration of the purified fragments was determined by the use of the ND-100 NanoDrop spectrophotometer. The molar ratio of vector : insert typically used was 1:3 for ligations involving cohesive ends. Ligation ratios were altered to 1:2, 1:6, or 1:10 when the ration 1:3 proven unsuccessful. The formula used for calculating the molar ratios from the mass ratios was:

molar ratio insert/vector = vector(ng) x insert size(kb) / insert(ng) x vector size(kb)

The vector backbone and insert were ligated using New England Biolabs' Quick Stick Ligase according to the manufacturer's protocol, in a 21µl reaction at room temperature for 5 min. The ligations were purified using Qiagen's QiaQuick Gel Extraction Kit according to the manufacturer's protocol prior to elecroporation.

2.2.2.5. Plasmid DNA transformation of bacteria by electroporation

LB SOC was pre-warmed to room temperature. OneShot Top10 Electrocompetent *E.Coli* Cells (Invitrogen) were thawed on ice. Once thawed 1-5ng of pre-chilled plasmid DNA, not exceeding a volume of 10% of the solution containing the electrocompetent bacterial cells, was added to the 30µl of cells. The tube was gently flicked to mix the solutions. The contents were then transferred to a pre-chilled 0.1cm electroporation

cuvette. The electroporator (BioRad) was set at 50μF and 150Ω, the powerpack at 1500V, 25mA and 25W, and the cells were electroporated. Electroporated cells were swiftly transferred to 1ml LB-SOC media and incubated in a 37°C shaking incubater at 250rpm for 45 minutes. The cells were then plated onto selective agar plates pre-warmed to 37°C overnight for a maximum of 16 hours.

2.2.2.6. Screening of transformants

After plating the cells onto selective media and incubating for 16 hours, a single colony was picked under sterile conditions with the a sterile pipette tip and used to inoculate 5ml of autoclaved LB broth, pre-warmed to 37°C, containing the appropriate antibiotic. This was then placed in a 37°C incubator and shaker set at 250rpm for 8 hours. DNA was extracted from 1.5ml of preculture as per the Qiagen Miniprep Kit protocol. For a maxi-prep, 0.25ml of preculture was used to inoculate 250ml sterile LB broth containing the required antibiotic and was grown in the incubator/shaker overnight for a maximum of 16 hours. The solution was then spun at 4°C at 3000g for 30 minutes to pellet the bacteria. The plasmid was then extracted as per the Qiagen Maxiprep Kit protocol. Plasmid DNA was further characterised by restriction analysis and gel electrophoresis. This procedure was repeated for 10 individual bacterial colonies unless stated for cloning purposes where multiple batches of 24 colonies were typically tested.

2.2.2.7. Analysis of DNA By Restriction Endonuclease Digestion and Electrophoresis

10x TBE (1 Litre): 108g Tris Base, 55g Boric Acid, 40ml EDTA (pH 8.0) dissolved in 1L double distilled H_20.

10x DNA Loading Buffer: 50% glycerol, 100mM Na_3EDTA, 1% SDS, 0.1% bromophenol blue

In order to verify the quality of each plasmid preparation, approximately 1μg of plasmid DNA was digested with one or more restriction enzymes with corresponding sites within the construct. Digests were carried out according to the manufacturer's

(New England Biolabs (NEB), Promega, or Fermentas) recommended conditions and buffers in a 50µl total reaction volume, and incubations were carried out in water baths. Then a sample of digested plasmid DNA (~100-200ng) was mixed with the appropriate volume of 5x loading buffer, and loaded into the well of a typically 0.8% agarose gel. A DNA molecular weight marker and undigested samples of plasmid DNA were loaded into wells alongside the sample, and electrophoresed at 50-100V. 1 x TBE was used as the electrophoresis buffer. Ethidium bromide was incorporated into a gel prior to gel setting at a concentration of 0.5µg/ml, and the DNA fragments were visualised under UV light. Gel images were recorded electronically on a digital imaging system (UVP) or by photographic film with an automatic Polaroid DS34 direct screem instant camera.

2.2.2.8. Amplification and purification of plasmid DNA

The small scale purification of plasmid DNA (Mini-Prep) was carried out by the use of Qiagen's mini-prep plasmid DNA purification kit as per the manufacturer's protocol from a 5-10ml starter culture grown at 37°C for 8 hours.

The larger scale amplification of plasmid DNA (Maxi-Prep) was conducted by the dilution of a starter culture by 1:1000 in 150-200ml sterile LB broth containing the appropriate antibiotic, grown overnight at 37°C in a shaking incubator rotating at 250rpm. The plasmid was purified by the use of Qiagen's maxiprep kit. DNA isolation using this kit was based upon the binding of DNA to the anion-exchange resin columns under low-salt and pH conditions. A medium-salt wash removed protein, RNA, and other low molecular weight impurities. Finally, a high-salt buffer wash led to the elution of the DNA, which was concentrated and desalted and precipitated with isopropanol. The DNA was resuspended in sterile ddH_2O at room temperature and stored at -20°C.

2.2.2.9. Glycerol stocks:

600µl of an overnight bacterial culture was put into a cryotube and 300µl of autoclaved 80% glycerol was added, gently mixed, and immediately stored at -80°C.

2.2.2.10. Plasmid Vectors Construction At RHUL Lab

- CeGFP-C1: This vector was constructed by excising the CMV promoter from peGFP-C1 and ligating the CAGG promoter excised from the CeGFP-N3 vector. This was done by digesting CeGFP-N3 with *SalI*, blunting with T4 DNA Polymerase, then digesting with *ApaLI* to yield the insert fragment of 1288bp, and another 3718bp fragment. peGFP-C1 was digested with *AfeI*, blunted with T4 DNA Polymerase, then digested with *ApaLI* yielding a 968bp fragment and the 3763bp backbone into which the CAGG promoter was inserted. The 1288bp CAGG promoter was ligated into the 3763bp peGFP-C1 backbone as a cohesive-blunt directional ligation.

- CpEPI-eGFP: This vector was constructed by excising the CMV promoter from pEPI-eGFP and ligating the CAGG promoter excised from the CeGFP-N3 vector. This was done by digesting CeGFP-N3 with *SalI*, blunting with T4 DNA Polymerase, then digesting with *ApaLI* to yield the insert fragment of 1288bp, and another 3718bp fragment. pEPIeGFP was digested with *AfeI*, blunted with T4 DNA Polymerase, then digested with *ApaLI* yielding a 968bp fragment and the 5726bp backbone into which the CAGG promoter was inserted. The 1288bp CAGG promoter was ligated into the 5726bp pEPI-eGFP backbone as a cohesive-cohesive directional ligation.

- CMYC-pEPI: This vector was constructed by amplifying the cMYC S/MAR element by PCR from the BAC clone template containing the element (pBACe3.6 RPCI Mouse BAC clone) and inserting it into peGFP-C1. The amplified cMYC S/MAR fragment was flanked with an *XhoI* restriction site at its 5' end and an *EcoRI* site at its 3'end. peGFP-C1 contains *EcoRI* and *XhoI* sites within its multiple cloning site. The amplified product and peGFP-C1 were digested with *EcoRI* and *XhoI,* and the product composed of the cMYC S/MAR element was inserted into the peGFP-C1 backbone as a cohesive-cohesive directional ligation. The cMYC S/MAR element was then excised from the vector by digestion with *AgeI* and *ApaI*, and CeGFP-C1 was also digested with *AgeI* and *ApaI*, and the cMYC S/MAR was ligated into the CeGFP-C1 backbone as a cohesive-cohesive directional ligation.

2.2.3. Fluorescence Immunostaining of cells

4% paraformaldehyde fixing solution (100ml): 4g paraformaldehyde, 10mM NaOH, 10ml of 10x PBS, pH adjusted to 7.4 with 1M HCl, final volume adjusted to 100ml with distilled H_2O

2.2.3.1. eGFP immunostaining for the enhancement of fluorescence signal from eGFP positive cells

Cells were plated in 1cm^2 chamber wells at the required density and left in the incubator overnight to adhere. The following day cells were washed once in PBS pre-warmed to 37°C. The cells were incubated in 4% paraformaldehyde fixing solution pre-warmed to 37°C for 5 minutes in order to fix the cells. The cells were rinsed twice in PBS. Next, they were incubated in PBS/0.1% Triton for 10 minutes to permeabilise the cells' membranes to facilitate antibody entry. Cells were rinsed again twice in PBS. 20% Normal Goat Serum/PBS was added to the cells and incubated for 30 minutes. Mouse monoclonal primary anti-GFP antibody diluted 1 in 400 in 5% Horse Serum/PBS was added and kept at 4°C overnight. The next day the cells were rinsed once and then washed twice for 5 minutes in 5% Horse Serum/PBS. The goat anti-mouse secondary antibody diluted 1 in 200 in 5% Horse Serum/PBS was added and the cells were incubated at room temperature for 1 hour in the dark. Next they were rinsed once and washed 3 times for 10 minutes in PBS. The wells were peeled off the slide, DAPI mounting medium was dropped onto the cells, and the slide was covered with a coverslip. The slides were left at room temperature overnight to set and the fluorescence was detected and images captured the next day at 490nm by the use of a Leica 800 fluorescence microscope at 10x magnification.

2.2.3.2. Immunostaining for the detection of lysenin/eGFP positive cells

Cells were plated in 1cm^2 chamber wells at the required density and left in the incubator overnight to adhere. The following day the medium was removed and the cells were washed once in PBS pre-warmed to 37°C. The cells were incubated in 4%

paraformaldehyde fixing solution pre-warmed to 37°C for 5 minutes in order to fix the cells. The cells were rinsed twice in PBS. They were incubated for 30 minutes at room temperature in 2% BSA/PBS. Next, 0.5µg/ml Lysenin diluted in 2% BSA/PBS was added to the cells and incubated at room temperature for 1 hour. In this step the lysenin bound to sphingomyelin present in large amounts on the cell surface of reserve cells. The cells were rinsed again twice in PBS. For co-staining for eGFP, the cells were incubated in 0.1% Triton/PBS for 10 minutes to permeabilise the cells' membranes to facilitate antibody entry, rinsed twice with PBS, and incubated in 20% Normal Goat Serum/PBS for 30 minutes. Rabbit anti-lysenin primary antibody diluted 1 in 200 in 2% BSA/PBS was added and, if staining for GFP, the mouse monoclonal primary anti-GFP antibody diluted 1 in 400 in 5% Horse Serum/PBS was also added at this stage. This was incubated at 4°C overnight. The next day the cells were rinsed once and then washed twice for 5 minutes in 5% Horse Serum/PBS. The secondary antibody swine anti-rabbit diluted 1 in 200 in 2% BSA/PBS was added and incubated in the dark at room temperature for 1 hour. If also staining for GFP, goat anti-mouse secondary antibody diluted 1 in 200 in 5% Horse Serum/PBS was added as well at this stage. Next the cells were rinsed once and washed 3 times for 10 minutes in PBS. Finally, the wells were peeled off the slide, DAPI mounting medium was dropped onto the cells, and the slide was covered with a coverslip. The slides were left at room temperature overnight to set and the fluorescence was detected and images captured the next day at 490nm by the use of a Leica 800 fluorescence microscope at 10x magnification.

2.2.4. Total DNA extraction

DNAzol® Genomic DNA Isolation reagent was used to extract total DNA from cells according to the manufacturer's protocol. This extraction technique is based upon the use of guanidine – detergent lysing solution which hydrolyzes RNA and selectively precipitates DNA from the lysate with ethanol (Chomczynski *et al*., 1997). Total DNA was resuspended in sterile ddH_2O and stored at -20°C.

2.2.5. HRP-Labelled Southern Blotting

Pre-hybridisation Buffer: 0.5M NaCl, 5% blocking reagent (supplied with kit), made up to 10ml pre-hybridisation buffer ECL Gold (supplied with kit).

20x SSC (1 Litre): 3M NaCl, 0.3M Na_3Citrate, pH adjusted to 7 using 1M HCl, final volume adjusted to 1L with distilled H_2O.

15µg of extracted total genomic DNA was digested with *EcoRI* which cuts once within the plasmids tested. The digested DNA was run on a 0.8% agarose gel in TBE buffer at 100V for 2 hours. The gel was then transferred to a UV gel visualiser in order to confirm digestion and images were taken. The gel was exposed to UV light for 1 minute to cross-link the DNA and then soaked in 0.5M NaOH for 30 minutes to denature the DNA. Next, Whatman paper cut to the size of the gel was soaked in 5x SSC and placed on a flat surface covered in cling film. The gel was placed on the paper, followed by a piece of Hybond N^+ membrane also cut to the size of the gel. Three more Whatman papers soaked in 2x SSC were placed over the membrane, followed by a stack of tissues and a weight of 500g atop the stack. Blotting was done for 2 hours. Next, the membrane was soaked in 5x SSC for 10 minutes. It was then soaked in 1.5x SSC for 10 minutes. The membrane was then placed in pre-hybridisation buffer in the hybridisation container, pre-warmed to 42°C, for 1 hour.

During the pre-hybridisation step the probe was prepared by the use of Amersham HRP Direct Labelling Kit. The template used was the eGFP sequence found in the plasmids tested, generated by PCR amplification. The template fragments were run on a gel to confirm the correct size, then cut out and purified with Qiagen's QiaQuick Gel Extraction Kit. 100ng of template was resuspended in distilled water to a final volume of 10µl. The probe was completely denatured to become single stranded by placing the eppendorf containing the probe in boiling water for 5 minutes then immediately placing in an ice water bath. Following a quick spin in the centrifuge the labelling was then carried out as per protocol provided with the labelling kit, and probe was directly labelled with Horseradish Peroxidase enzyme.

Hybridisation of the probe to the target DNA on the membrane was carried out by adding the labelled probe to the pre-hybridisation solution, mixing well, and leaving to hybridise overnight.

The next day the rinsing solutions that consisted of 2x SSC and 0.5x SSC were pre-warmed to 42°C. The membrane was washed in the 2x SSC first for 20 minutes, then the 0.5x SSC for 20 minutes, both at 42°C in the hybridisation oven.

To visualise the locations where the probe hybridised to target DNA the Amersham ECF Chemifluorescent substrate was added to the membrane in a dark room as per protocol provided with the substrate and left to react for 1 minute. The membrane was placed in a plastic sleeve and an autoradiography film Amersham Hyperfilm was placed on top between 5 minutes and overnight depending upon the sensitivity required. The film was developed by being placed in developer, rinsed in distilled water, then placed in fixer, rinsed in distilled water, then air dried.

2.2.6. Fluorescent In-Situ Hybridisation (FISH)

Fixative Solution: A solution of 3:1 methanol : acetic acid.

Hybridisation Buffer (7ml): 1g Dextran sulphate, 5.5ml formamide, 0.5ml 20x SSC, final volume made up to 7ml with distilled H_2O.

Hypotonic Solution: 0.4% KCl, 0.4% Na_3citrate

Denaturant Solution: 70% formamide, 2x SSC.

Wash A: 50% formamide, 2x SSC.

Wash B: 2x SSC

Wash C: 2xSSC, 0.1% NP40

$1x10^5$ cells were typically plated and grown to 85% confluency in a flask. Demecolcine was added to the medium for a final concentration of 1µg/ml and incubated in a 37°C, 8% CO_2, humidified chamber for 1 hour to block mitosis of the cells at metaphase where the chromosomes are most condensed. Next the cells were trypsinised and pelleted. The pellet was gently resuspended in 2ml hypotonic solution pre-warmed to 37°C by flicking the tube until the pellet was resuspended. This was incubated in a 37°C water bath for 7 minutes in order to release the nuclei. Next the nuclei were pelleted by centrifugation at 650rpm for 5 minutes, and gently resuspended in 2ml freshly made Fixative Solution and was incubated at room temperature for 30 minutes. The nuclei were pelleted again by centrifugation, and resuspended again in 2ml fresh

Fixative Solution and incubated at room temperature for 20 minutes. The nuclei were pelleted once more and resuspended in 250µl fresh fixative. This was stored at -20°C until use.

For chromosome spreads, glass slides were used that were aged by immersing for 24 hours in 3% HCl/ 70% Methanol/ 27% H_2O. The slides were rinsed 3 times in distilled H_2O and are stored at 4°C until use. To prepare chromosome spreads the glass slides were allowed to air dry and the cell suspensions in fixative were kept on ice. 50µl of the nuclei suspension was pipetted and dropped onto the glass slides from a height of approximately 1 meter. The fixative was allowed to evaporate and the slides were left at room temperature overnight to dry.

To prepare the slides for staining they were dehydrated by immersing in a 70%, 80%, and 100% ethanol series for 2 minutes at each concentration. The slides were allowed to air dry, then the chromosomes were denatured by placing in Denaturant Solution pre-warmed to 40°C for 1 minute, pre-warmed to 60°C for 1 minute, and pre-warmed to 76°C for 5 minutes. They were then placed back in the denaturing solution pre-warmed to 60°C for 1 minute, and the solution was pre-warmed to 40°C for 1 minute. These stepwise increased in temperature keep the condensed chromosomes from adopting a 'fuzzy' appearance which can result from sudden large temperature changes. The slides were again dehydrated in an ethanol series as described above and allowed to air dry.

To prepare the probe the Nick Translation Kit from Abbott Molecular was used as per the protocol supplied. The probe template used was the eGFP-C1/CeGFP-C1 vector. An eppindorf was pre-cooled on ice, and the following components were added in: 1µg of template DNA, 2.5µl of 0.2mM Red-dUTP, 5µl of 0.1mM dTTP, 10µl of 0.3mM dNTP mix, 5µl of 10X nick translation buffer, 10µl of nick translation enzyme, and a volume of nuclease-free H_2O to make the final reaction volume up to 50µl. The solution was briefly vortexed and centrifuged, then incubated at 15°C overnight for 16 hours. The reaction was then arrested by heating in a water bath at 70°C for 10 minutes. The solution was then briefly centrifuged and stored on ice or at -20°C until further use.

Once the probe was generated and purified using Qiagen's QiaQuick Gel Extraction Kit 1µl of probe mixed with 1µl of 10mg/ml sonicated salmon sperm DNA from Sigma, 1µl of sonicated Human Cot-1 DNA from Abbott Molecular, and 7µl of Hybridisation Buffer

per slide. The components were mixed well and the DNA was denatured by placing the eppendorf on a heating block at 76°C for 5 minutes.

Hybridisation was carried out by applying 10µl of the probe mixture to each slide and immediately covering with a glass coverslip and avoiding creating bubbles. The edges of the coverslips were sealed with rubber cement and allowed to air dry for 5 minutes at room temperature. The slides were incubated upsidedown at 37°C, 8% CO_2, humidified chamber for 48 hours for hybridisation to take place.

After hybridisation the rubber cement was peeled off and the slides were immersed in Wash A at 43°C for 10 minutes. The slides were gently shaken and the coverslips fell off at this stage. The slides were washed again in Wash A at 43°C for 10 minutes 2 more times. Next they were washed with Wash B at 43°C for 10 minutes. Finally, they were washed in Wash C at 43°C for 5 minutes. The slides were allowed to completely air dry. 10µl DAPI mounting medium was applied to each slide and was covered by a coverslip. The slides were visualised using a CARV II non-laser Confocal Imager fluorescence microscope from Image Solutions at 100x magnification.

2.2.7. Fluorescent Activated Cell Sorting (FACS) and flow cytometry analysis

2.2.7.1. Analysis of Enhanced Green Fluorescent Protein (eGFP) expression by flow cytometry

$1x10^5$ cells were typically analysed by flow cytometry for the expression of eGFP at appropriate intervals after transfection with vectors carrying eGFP expression cassettes. Cells to be analysed were trypsinised, pelleted, and the pellet resuspended in 500µl of 1% PFA/PBS solution. The samples were stored in the dark at 4°C for a maximum of 5 days until analysis. The cells were analysed using a Beckman Coulter EPICS XL flow cytometer with the use of XL System II™ software (at the Dept. of Immunology, ICSM, Chelsea Westminster Hospital, South Kensington by colleagues Timos Papagatsias and Takis Athanasopoulos) or a BD Biosciences FACSCalibur flow cytometer with the use of FACScomp™ software (at Royal Holloway). Cells were gated using non-transfected fixed cells according to forward scatter and side scatter in order to distinguish single cells from doublets or cell clumps (see Figure 2.8). eGFP gating

was also conducted using untransfected cells according to cell count and FITC fluorescence, where any cells from test samples emitting fluorescence greater than the gate set by the negative control were considered to be eGFP positive and counted as such, as was determined with the positive control consisting of cells 24 hours post transfection with the vector CeGFP-C1 containing an eGFP expression cassette (see Figure 2.9). The mean fluorescence intensity (MFI) was determind by the flow cytometer by measuring the fluorescence of each individual eGFP positive cell that is passed through, then calculating the mean intensity of those positive cells.

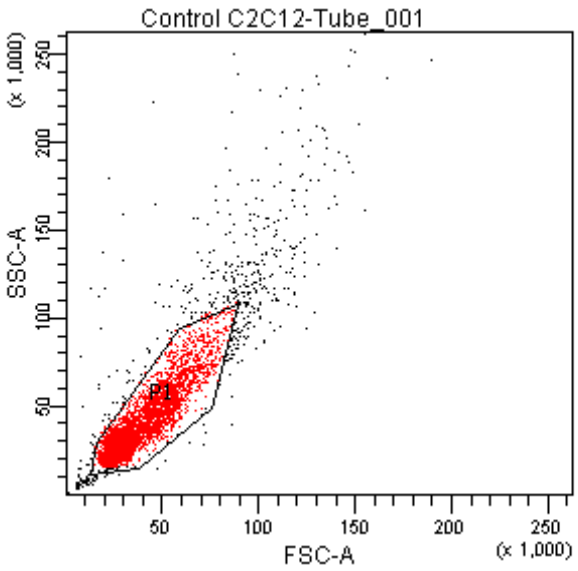


Figure 2.8 Flow cytometry gating of C2C12 cells using untransfected cells according to forward scatter and side scatter.

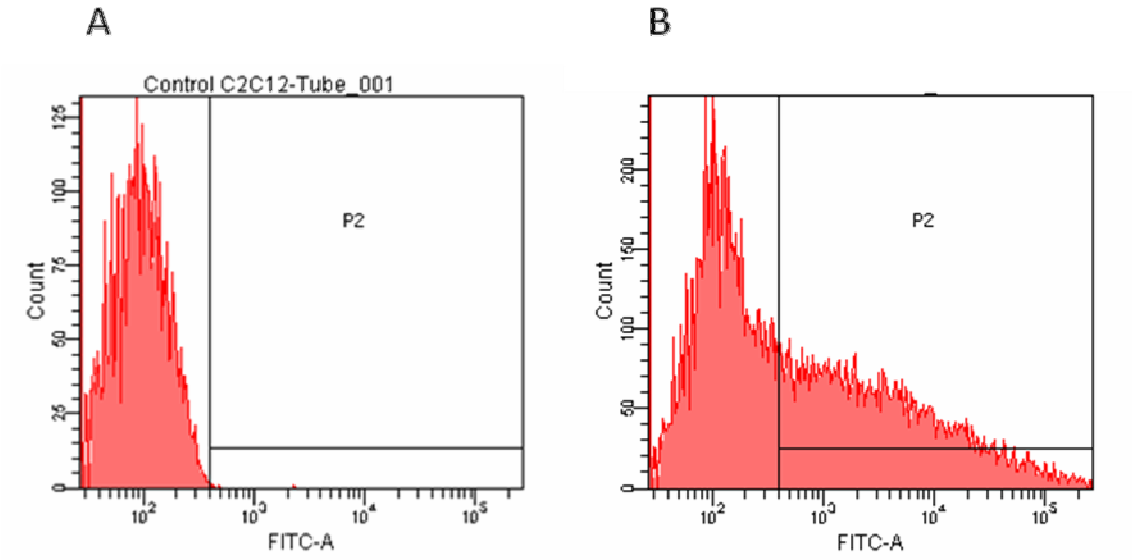


Figure 2.9 Flow cytometry FITC gating of C2C12 cells using untransfected C2C12 cells as a negative control according to cell count and FITC fluorescence (A). (B) shows C2C12 cells 24 hours post transfection with a vector containing an eGFP expression cassette (CeGFP-C1) indicating the presence of eGFP positive cells within the sample population.

2.2.7.2. Cell cycle analysis by flow cytometry

Hypotonic Solution: 0.4% KCl, 0.4% Na_3citrate

DAPI staining solution: 10 mM Tris-HC1 (pH 7.5), 0.1% Triton X-100, and 4 µg/ml DAPI

$1x10^3$-$1x10^4$ cells were typically analysed by flow cytometry to determine the percentage of cells in each phase of the cell cycle, taken at appropriate intervals from C2C12 cell cultures in low serum methionine depleted medium. Cells to be analysed were trypsinised, pelleted, and the pellet resuspended in 400µl hypotonic solution. The solution was pipetted thoroughly and incubated at 37°C for 7 min in order to release the nuclei from the cells. Next the nuclei were pelleted by centrifugation at 650rpm for 5 minutes, and gently resuspended in 800µl DAPI staining solution and the tube was gently flicked to mix. The samples were kept on ice for a maximum of 30 minutes before being analysed by a PAS2 flow cytometer (Partec, Münster, Germany) by UV excitation light generated by a mercury lamp, and data was analysed and graphs generated using Windows™ FloMax® software.

2.2.7.3. Fluorescent Activated Cell Sorting (FACS)

$1x10^5$ cells were sorted for eGFP positive cells. Cells to be sorted were trypsinised, pelleted, and the pellet resuspended in 300µl growth medium. The samples were kept on ice until analysed (at the University College London Wolfson Institute of Biomedical Research by colleague Thomas Adejumo) in a Beckman Coulter MoFlo XDP Cell Sorter with the use of Summit software. The negative control used was untransfected C2C12 cells and the positive control was C2C12 cells 24 hours post transfection with the vector CeGFP-C1 containing an eGFP expression cassette. The cells were gated according to forward scatter and side scatter to distinguish single cells from doublets (see Figure 2.10). eGFP gating was set by comparison of the negative control with the positive control (see Figure 2.11). Samples were sorted on the basis of the determined gating, and eGFP positive cells were resuspended in growth medium and replated.

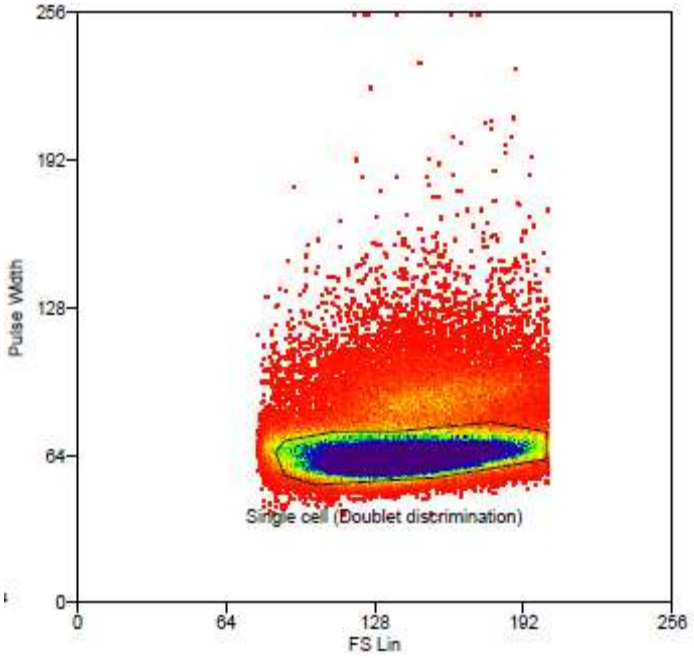


Figure 2.10 FACS gating of C2C12 cells according to forward scatter and side scatter to discriminate doublets from single cells.

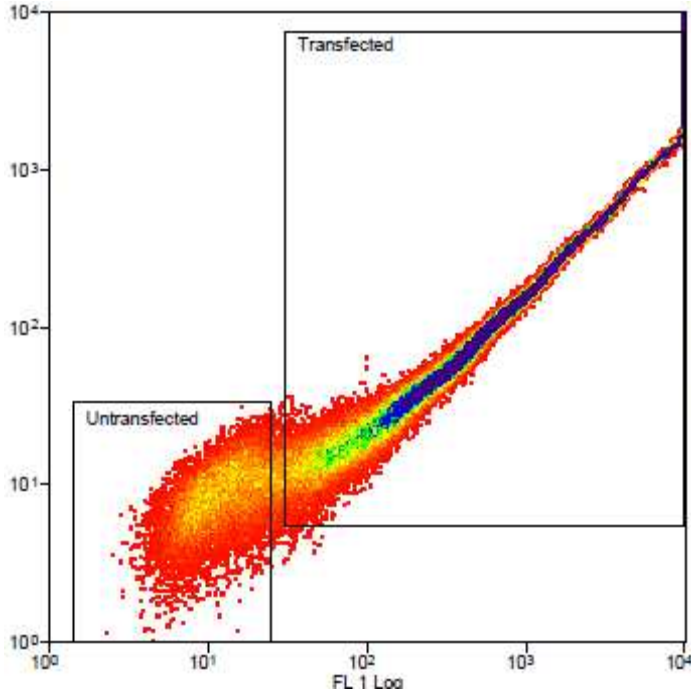


Figure 2.11 FACS gating for eGFP positive C2C12 cells using untransfected cells and cells 24 hours post transfection with a vector containing an eGFP expression cassette (CeGFP-C1).

2.2.8. Polymerase Chain Reaction (PCR)

PCR was carried out in 25µl volumes in sterile tubes in a PTC-200 Peltier Thermal Cycler. Each PCR reaction contained 12.5µl GeneSys MasterMix, which included 1x amplification buffer, 10µM dNTPs, 1.5mM MgCl, and thermostable DNA polymerase, 1µl of 10µM forward primer, 1µl of 10µM reverse primer, 10µl of template DNA at the required concentration, and 0.5µl sterile ddH_2O. The PCR cycle began with 2 minutes at 94°C, then 35 cycles of 30 seconds at 94°C for denaturation, 30 seconds at optimal annealing temperature previously determined by a temperature gradient, and 45 seconds at 72°C for elongation. Following this was 10 minutes at 72°C for final elongation. PCR generated fragments were stored at 4°C and and analysed by electrophoresis. The primers were designed by Primer3 program, and a pair of forward and reverse primers were chosen based upon the similarities of their lengths, GC content, and melting temperatures (T_m). In order to determine the optimal annealing temperature for each set of primers a PCR reaction was set up using 1ng of template DNA and run according to the program described above. A temperature gradient ranging from 65-72°C was conducted for the annealing temperature, and the optimal annealing temperature was chosen after electrophoresis of the PCR products based upon the absence of non-specific sequence amplification and the qualitatively determined brightness of the band, denoting efficient target sequence amplification.

2.2.9. Statistical Analyses:

eGFP positive cell populations were assumed to have normal distributions. NCSS program was used to conduct the statistical analyses. For the comparison of two populations a student's T-test was used. For 3 or more populations a One-way ANOVA test was conducted. The post-hoc Tukey-Kramer test was used when comparing 3 or more populations to determine which groups differed significantly to others where significance was confirmed by the student's T-test or ANOVA. Significance was set at $P < 0.05$.

CHAPTER 3: Evaluation of The Expression and Episomal Establishment Of S/MAR Vectors in The Human Epithelial Cervial Cancer Cell Line HeLa, Murine Myoblast Cell Line C2C12, and Human Hepatocyte Cell Line HepG2

3.1. Introduction:

The aim of the experiments in this chapter was to investigate the expression and episomal status of a plasmid vector containing an S/MAR element following transfection into human epithelial cervical cancer cell line HeLa, the murine skeletal muscle cell line C2C12, and the human hepatocyte cell line HepG2 over time, compared with those of a control vector in order to investigate the effects of the presence of the S/MAR element. The S/MAR vector, pEPI-eGFP, was a plasmid containing a CMV-driven eGFP reporter gene, with the 2.0kb human IFN-β S/MAR element positioned at the 3' end of the reporter gene within the open reading frame, and preceding the 5' end of the SV40 polyadenylation signal. The vector also contained a Neomycin resistance gene driven by the SV40 promoter (vector maps in Chapter 2: Materials & Methods). The control vector, eGFP-C1, was composed of the same elements and sequences as pEPI-eGFP, but lacked the S/MAR element.

All cell lines used were immortal lines in order to be able to investigate the expression and retention of the vectors tested in replicating cells after many cell divisions over time. The C2C12 line, as indicated in Chapter 1 Introduction (section 1.3.3), was considered a suitable skeletal muscle line for the purposes of the investigations in this thesis. The other two cell lines were chosen for other reasons. In considering the differences observed in different cell lines in the expression and retention of these vectors over time in studies reviewed in Chapter 1 Introduction (section 1.10), cell lines with different characteristics to C2C12 were chosen in order to investigate whether any differences may be observed and, if so, what these variations may indicate. Thus the HepG2 line was chosen as it represented a different tissue type, and also had a longer doubling time than the other two cell lines chosen. Hela was chosen as a control line, as the integration status of eGFP-C1 and pEPI-eGFP in HeLa cells over time had

already been investigated by Schaarschmidt *et al* (Schaarschmidt *et al*., 2004). However, the expression of these vectors in this cell line had not yet been evaluated.

Also evaluated was the 'pEPI-M18' vector, devised and kindly provided by J. Bode, in C2C12 cells. This vector was a plasmid composed of the same elements and sequences as pEPI-eGFP, except a 0.7kb segment derived from the human β-IFN S/MAR, termed 'mini S/MAR', replaced the full length S/MAR. This mini S/MAR element was created by a recombination event which had occurred within the full length S/MAR element after the long-term cultivation of a population of CHO-K1 cells transfected with a minicircle containing the full length S/MAR element, leading to a decrease in S/MAR size from 4kb to 0.7kb. A region of 1.2kb had been excised from within the S/MAR element. This occurred in two separate independent events where the deletions were found to be identical. This vector, termed 'pEPI-M18' had indicated improvements within populations transfected by it over those containing pEPI-eGFP, in terms of percentage of positive cells and mean fluorescence intensity (personal communication with J. Bode).This vector was tested in order to determine whether this shorter version of the β-IFN S/MAR would be able to confer the same properties to its vector as the full length element. If successful, the smaller size of the vector would allow it a larger carrying capacity for therapeutic transgenes to be inserted than pEPI-eGFP.

Samples were analysed by PCR at several time points to determine whether copies of the vector sequences remained present within the transfected populations over time. Flow cytometry and immunostaining of the samples were used as assessments of the expression of the eGFP reporter gene over time in each population, thus giving an indication of the efficacy, in terms of both number of positive cells and level of expression, of each vector tested at expressing the transgene over time. Southern blotting, plasmid rescue, and fluorescent in-situ hybridisation was used to determine the episomal/integrant status of the transfected vectors at the final day of the experiments.

Table of plasmid vectors tested in this chapter:

Vector Name	Vector Feature	Promoter-Transgene	Antibiotic Resistance	Vector Length (kb)
eGFP-C1	No S/MAR element	CMV-eGFP	Kan/Neo	4.7
pEPI-eGFP	Full-length 2kb β-IFN S/MAR element	CMV-eGFP	Kan/Neo	6.7
pEPI-M18	Truncated 0.7kb β-IFN S/MAR element	CMV-eGFP	Kan/Neo	5.4

Table 3.1 Vectors used in this chapter, indicating the name, the presence of an S/MAR element, the transgene promoter, the transgene used for assessment studies, the antibiotic resistance gene, and the length (kb) of each construct.

3.2. Results

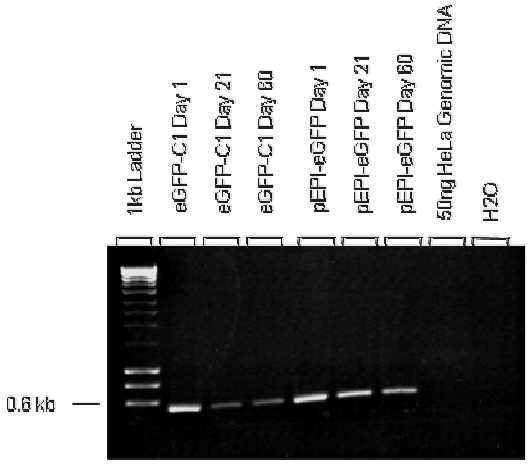
3.2.1 Experimental Design

To carry out the aims of this chapter, cell populations of three cell lines were transfected with each vector, selected for neomycin resistence, and passaged for many generations. This was conducted by plating two 75cm^2 flasks with 7.5x10^5 HeLa cells each, three 75cm^2 flasks with 7.5x10^5 C2C12 cells each, and two 75cm^2 flasks with 37.5x10^5 HepG2 cells each. After 24 hours the cells had reached ~70% confluency. Three flasks of cells, one of each cell line, were transfected with the control vector eGFP-C1, three flasks were transfected with the S/MAR vector pEPI-eGFP, and the final flask, which contained C2C12 cells, was transfected with the vector containing the 'mini-S/MAR', pEPI-M18, by lipofection (see Chapter 2: Materials & Methods). 24 hours after transfection each transfected population of cells was analysed under a fluorescent microscope to confirm the presence of eGFP positive cells, to ensure that the cells had been successfully transfected with plasmid. Cell samples were collected from each population for analysis and the remaining cells were seeded into 25cm^2 flasks at a density of 5x10^4 cells for the HeLa and C2C12 cell populations, and of 1x10^5 cells for the HepG2 populations. The cells were put under selection by the addition of Geneticin (G-418) to the culture media at the appropriate concentration for each cell line (see Chapter 2: Materials & Methods). Fresh media containing G-418 was added to each culture every 48 hours, and the cells were passaged when culture confluency reached ~70%. Selection was maintained for 21 days, and on the final day cell samples were again collected from each population for analysis. The remaining cells were re-seeded again into fresh flasks, and the HeLa and C2C12 populations were passaged for a further 39 days, and the HepG2 populations for a further 99 days, due to the longer doubling time of HepG2 cells. On the final day of the experiment, Day 60 for HeLa and C2C12 and Day 120 for HepG2, cell samples were collected again from each population for analysis.

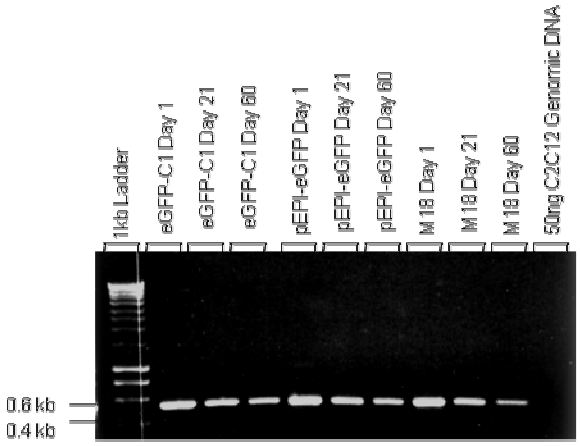
3.2.2 Detection of Vector Sequences in Transfected HeLa, C2C12, and HepG2 Populations By Polymerase Chain Reaction (PCR)

PCR analyses were conducted on samples taken on Days 1, 21, and 60 for HeLa and C2C12 transfected populations, and Days 1, 21, and 120 for the HepG2 transfected populations. 50ng of total DNA extracted from each sample was used per PCR reaction. The primers used were designed by Primer3 and lie within the Kanamycin resistance gene present in eGFP-C1, pEPI-eGFP, and pEPI-M18. The PCR was run for 35 cycles and the expected product size was 503bp. Products were run on 0.8% agarose gels counterstained with ethidium bromide and visualised under UV light see Materials & Methods). PCR on each sample confirmed the presence of eGFP-C1 and pEPI-eGFP vectors within cells in the transfected HeLa and HepG2 populations, and of eGFP-C1, pEPI-eGFP, and pEPI-M18 vectors within the transfected C2C12 populations (see Figure 3.1)

(A) HeLa:



(B) C2C12:



(C) HepG2:

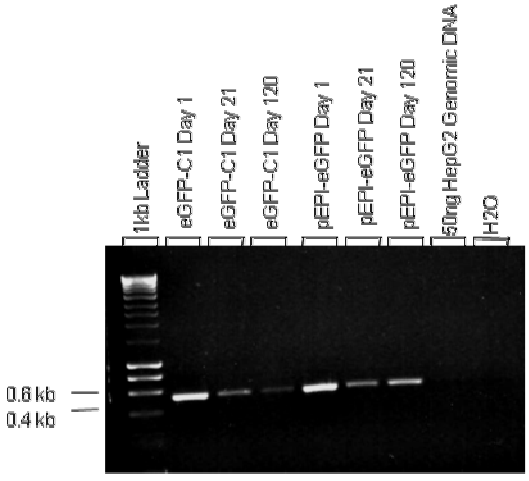


Figure 3.1 PCR analysis of 50ng of total HeLa DNA (A) or HepG2 DNA (C) transfected with eGFP-C1 or pEPI-eGFP, and C2C12 DNA (B) transfected with eGFP-C1, pEPI-eGFP, or pEPI-M18 vector on samples taken at Days 1, 21, and 60 (or Day 120 in the case of HepG2). The negative controls were 50ng untransfected C2C12, HeLa, or HepG2 genomic DNA, and H_20. The primers used were within the Kanamycin resistance gene, and product size expected was 503bp. Copies of the vectors were present within cells of the transfected populations in each cell line at each time point.

3.2.3 Transgene Expression Analysis of Transfected Cell Populations By Flow Cytometry And Immunostaining

Transfected cell populations were analysed by flow cytometry and immunostaining to assess eGFP transgene expression at three different time-points: Day 1, Day 21, and Day 60 for HeLa and C2C12 cells, and Day 1, Day 21, and Day 120 for HepG2 cells. Day 1 was 24 hours post transfection, Day 21 was the final day of selection, Day 60 was the final day of the experiment for the HeLa and C2C12 cells where cells were allowed to proliferate in growth medium for 39 days post selection, without selection pressure, and Day 120 the final day of the experiment for the HepG2 cells where cells were allowed to proliferate in growth medium for 99 day post-selection, without selection pressure.

For flow cytometry analysis, cells were fixed in 1% PFA/PBS and $1x10^5$ cells were analysed per sample. The mean percentage of eGFP positive cells and the mean fluorescence intensity of the positive cells were measured (as described in Chapter 2: Materials & Methods).

For the immunostaining $1x10^4$ cells of each sample were seeded into $1cm^2$ chamber slide wells and then fixed after 24 hours in 4% PFA/PBS prior to staining. The cells were stained in order to amplify the eGFP signal using mouse anti-eGFP primary antibody and goat anti-mouse secondary antibody conjugated with green fluorescent Alexa Fluor 488 dye.

The staining was conducted in triplicates (n = 3), 200 cells were counted per image, and of those 200 the number of positive cells was counted. The percentages ± the standard error of the mean (SEM) were calculated for each transfected population at each of the three chosen time points. The mean percentages of each transfected population at each time point were statistically compared, as were the transfected populations at all time points for each cell line, by One-Way ANOVA. The post-hoc Tukey-Kramer test was used to determine the means that were significantly different to other means in group comparisons (3+ transfected populations).

3.2.3.1 Expression Analysis Of HeLa Cell Populations Transfected With eGFP-C1 or pEPI-eGFP

Figure 3.2 shows the mean percentage of positive cells in the transfected populations determined by cell counts from the immunostaining images (a), and the mean percentages of eGFP positive cells as measured by flow cytometry (b). Data from Day 1 indicate that the transfection efficiencies of eGFP-C1 and pEPI-eGFP in this experiment were comparable to one another (eGFP-C1 mean number of eGFP positive cells= 62.4% (flow cytometry) and 50.3 ± 3.4% (immunostaining); pEPI-eGFP number of eGFP positive cells= 54.5% (flow cytometry) and 48.5 ± 4.8% (immunostaining)). This was confirmed by an ANOVA analysis on the cell counts data where no significant difference was found between the percentage of positive cells in the eGFP-C1 and pEPI-eGFP transfected populations at Day 1 ($P = 0.8$).

On the final day of selection at Day 21 all cells which survived selection theoretically contained a copy of the vector sequence. However, Figure 3.2 (a,b) show a decline in the percentage of eGFP positive cells from Day 1 to Day 21, which suggests that gene silencing was a possible reason for the lack of detectable eGFP expression, although this would require further examination, possibly by measuring gene transcripts by a method such as Qunatitative Reverse Transcriptase-PCR (Q-RT-PCR). The decline in eGFP positive cells was indicated by the mean percentages of positive eGFP-C1 and pEPI-eGFP transfected cells (means of 9.4% from flow cytometry and 2.3 ± 0.7% from the immunostaining cell counts; means of 18% from the flow cytometry and 4 ± 1.5% from the immunostaining cell counts, respectively). The ANOVA analysis showed that there was no significant difference in the percentage of positive cells between the two populations at Day 21 ($P = 0.4$). A further decrease in the percentage of positive cells in both populations was evident at Day 60. The data indicate that a very small percentage of cells continued to express the transgene at Day 60 (for the eGFP-C1 population means of 0.3% from flow cytometry and 1.25 ± 0.4% from the immunostaining cell counts; and for the pEPI-eGFP population means of 2.3% from the flow cytometry and 0% from the immunostaining cell counts), and that the mean percentage of positive cells in the eGFP-C1 transfected population was significantly higher than that in the pEPI-eGFP transfected population ($P = 0.04$). This is illustrated

by the representative immunostaining images in Figure 3.3 which show the decline observed in eGFP positive cells over time in both transfected populations.

Figure 3.4 shows the mean fluorescence intensity of the eGFP-C1 and pEPI-eGFP transfected populations, which is an indication of the mean level of transgene expression resulting from the transfected vector in each population. In taking the flow cytometry data shown for Day 1 with the data shown by immunostaining images, which were all taken at the same exposure, there was evidence to suggest that there was a higher proportion of eGFP-C1 transfected cells that produced more transgene product per cell than was produced by pEPI-eGFP transfected cells. In the immunostaining images this can be subjectively noted by the brightness and colour intensity of the positive cells (Figure 3.3).

Figure 3.4 shows that the mean fluorescence intensity of the eGFP positive cells declined in the eGFP-C1 population between Days 1 and 21, but remained similar between Days 21 to 60; and in the pEPI-eGFP population the mean fluorescence intensity declined from Day 1 to Day 60, which would indicate a decrease in the amount of transgene expressed per cell over time. The representative immunostaining images were also in support of this as the brightness and intensity of the positive cells at Day 21 and Day 60 in both populations were less than those observed in the Day 1 images, as was also subjectively observed (Figure 3.3).

Total eGFP fluorescence is an indication of the total amount of eGFP expressed in a given population of transfected cells. This is calculated by multiplying the mean percentage of eGFP positive cells by the mean fluorescence intensity. Figure 3.5 shows that an evident decline had occurred in the total amount of transgene expressed in both the eGFP-C1 and pEPI-eGFP transfected populations from Day 1 to Day 21, and a further decrease between Days 21 and 60.

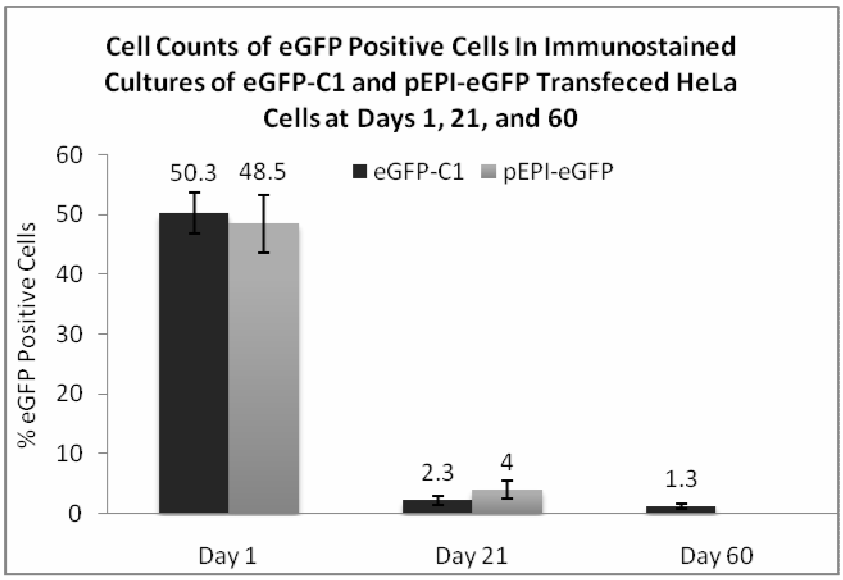


Figure 3.2(a) Mean percentages ± SEM (error bars) of eGFP positive cells derived from cell counts of immunostaining images of HeLa cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60). The values above each bar indicate the percentage of positive cells. The chart shows that the transfection efficiencies of the two vectors in this experiment were relatively high and comparable to one another before significantly declining from Day 1 to Day 60, but there were significantly more positive cells in the eGFP-C1 transfected population at Day 60 than in the pEPI-eGFP population. *n* = 3.

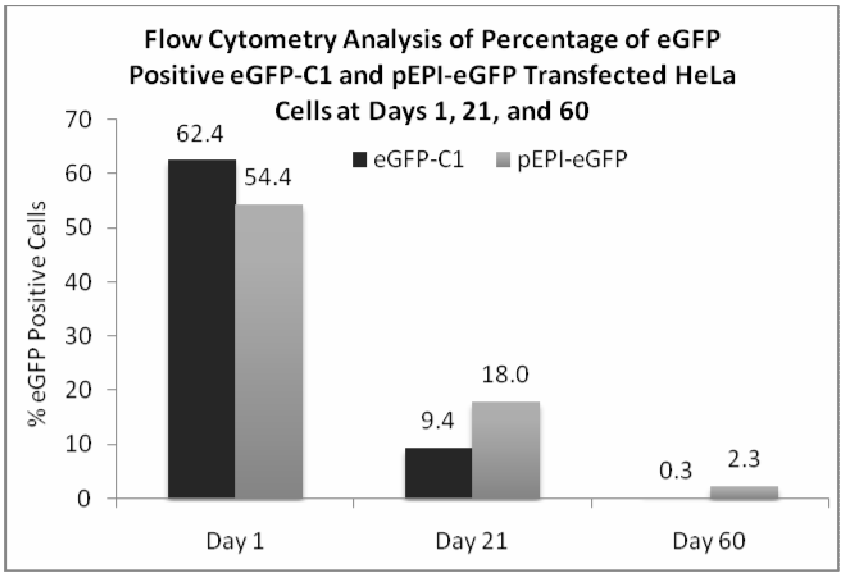


Figure 3.2(b) Flow cytometry analysis of the percentage of eGFP positive HeLa cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. A decline in the percentage of eGFP positive cells with time was evident in both the eGFP-C1 and pEPI-eGFP transfected populations.

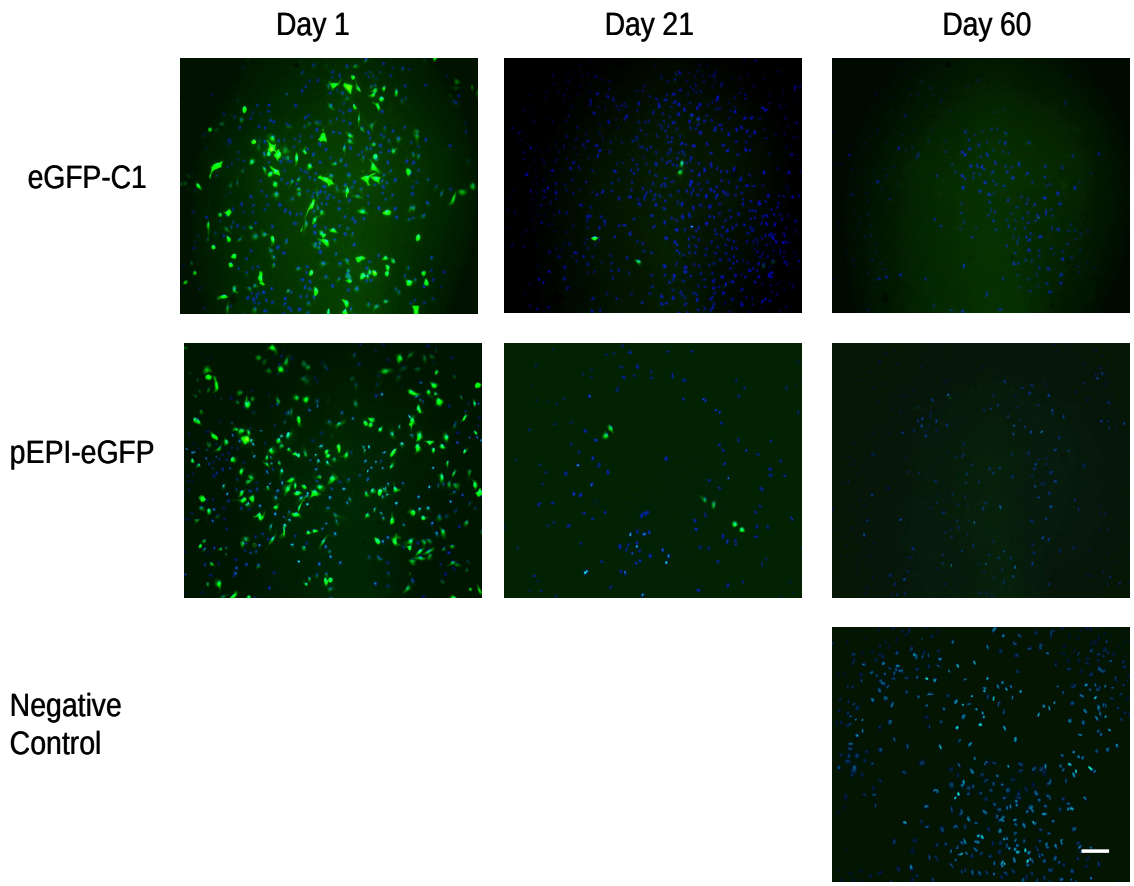


Figure 3.3 eGFP immunostaining of fixed HeLa cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1) (images A and B, respectively), the final day of selection after a 21 day selection period (Day 21) (images C and D, respectively), and the 39th day of proliferation without any selective pressure post-selection (Day 60) (images E and F, respectively). The negative control was untransfected HeLa cells stained for eGFP (image G). Green indicates eGFP positive cells and blue is the DAPI staining of the nuclei. These images are visual representations of the eGFP positive cells of the transfected populations and support the flow cytometry and cell counting data. A qualitative observation of the positive cells of the eGFP-C1 transfected cells at Day 1 is they were brighter and fluoresced more intensely than those of the pEPI-eGFP transfected cells, indicating higher transgene expression in eGFP-C1 transfected cells. The Day 21 images show a decrease in brightness and fluorescence of positive cells of both transfected populations, and a decrease in the amount of positive cells. By Day 60 no eGFP positive cells could be detected in the pEPI-eGFP transfected population. Bar, 90µm.

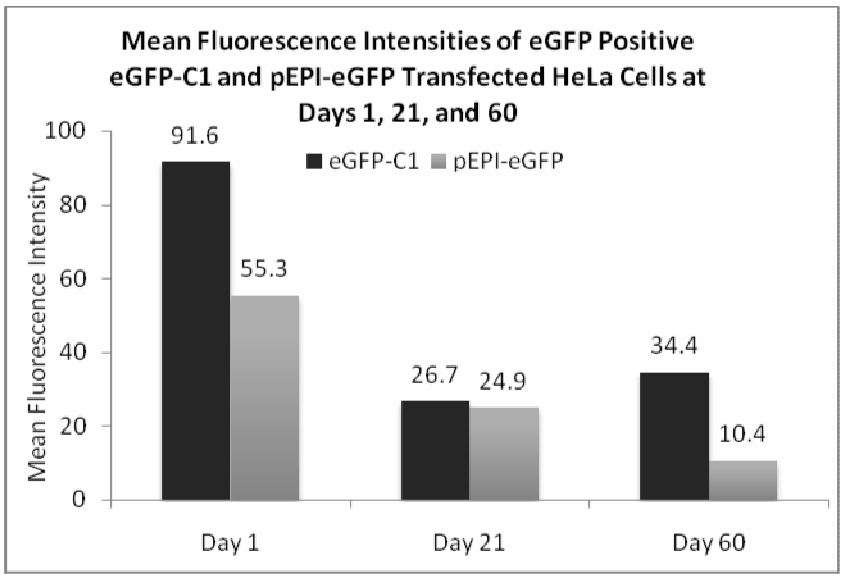


Figure 3.4 Flow cytometry analysis of the mean fluorescence intesity of eGFP positive HeLa cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. The values above the bars are the exact MFI values. These data suggest that on average the eGFP-C1 transfected cells produced more transgene product per cell than pEPI-eGFP transfected cells at Day 1. It also suggests that the mean fluorescence intensity declined in the eGFP-C1 population between Days 1 and 21, but remained similar between Days 21 to 60. In the pEPI-eGFP population the mean fluorescence intensity declined from Day 1 to Day 60, which would indicate a decrease in the amount of transgene expressed per cell over time.

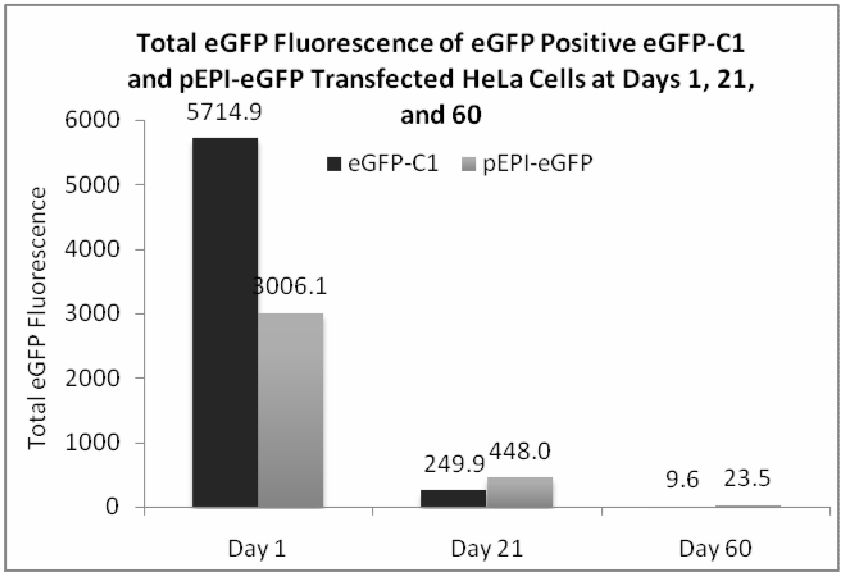


Figure 3.5 Flow cytometry analysis of the total eGFP fluorescence of HeLa cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. The data indicates a decline in the total amount of eGFP transgene expression in both transfected populations over time.

3.2.3.2 Expression Analysis of C2C12 Cell Populations Transfected With eGFP-C1, pEPI-eGFP, or pEPI-M18

Figure 3.6 shows the mean percentages ± the standard error of the mean (SEM) (*n* = 3) of positive cells in the transfected populations as found by cell counts from the immunostaining images (a), and the mean percentages of eGFP positive cells as measured by flow cytometry (b). Figure 3.6 (a) indicates that the transfection efficiency of the eGFP-C1 vector was the highest of the three vectors, followed by the pEPI-M18 vector, and finally the pEPI-eGFP vector which had the lowest efficiency in this experiment, according to the percentage of positive cells counted from the immunostaining images (means of 25.5± 1.6%, 20 ± 1.1%, and 10.8 ± 0.7% positive cells, respectively; P = 0.0033). Figure 3.6 (b) flow cytometry data supports the conclusion that pEPI-eGFP had a lower transfection efficiency than both the other plasmids (means of 37% positive cells in the eGFP-C1 transfected population, 42.1% for pEPI-M18, and 25.4% for pEPI-eGFP).

A decline was observed in the percentage of eGFP positive cells in each transfected population over time. On the final day of selection at Day 21 a decline in mean percentages of positive cells was observed from Day 1, which would indicate that there was a decrease in the amount of transgene expression per cell over time, or the loss of the vector from the cells that were transfected. The data also showed that on Day 21 the eGFP-C1 transfected population had a significantly higher mean percentage of positive cells than the pEPI-M18 and pEPI-eGFP transfected populations (means of 13.3 ± 2.1%, 4 ± 1.2%, and 0% respectively as indicated by cell counting; P = 0.0088). The mean percentages decreased further in all three populations by Day 60, where the only population that still contained positive cells was the eGFP-C1 transfected population (mean of 2 ± 0.8% as indicated by cell counting). However, the mean percentage of positive cells within the eGFP-C1 population was significantly greater than that of the pEPI-M18 and pEPI-eGFP transfected populations at Day 60 (P = 0.038).

The decrease in mean percentage of positive cells was also shown by the flow cytometry data (at Day 21 means of 28.7% in the eGFP-C1 population, 4.9% in the pEPI-M18 population, 5.1% in the pEPI-eGFP population, and at Day 60 means of 0.9%, 0.3%, and 1.7% respectively). Consistent with the charts were the immunostaining

images, illustrative of each transfected population at each time point (Figure 3.7), which were all taken at the same exposure and also indicated the decline observed in eGFP positive cells over time in the three populations.

Figure 3.8 shows the mean fluorescence intensities of the eGFP positive cells of the eGFP-C1, pEPI-eGFP, and pEPI-M18 transfected populations as measured by flow cytometry. The data suggests that on Day 1 the eGFP-C1 and pEPI-M18 transfected populations had similar mean expressions, and that on average, cells in these two populations produced more transgene product per cell than those in the pEPI-eGFP transfected cells. The immunostaining images, suggest that there was a higher proportion of pEPI-M18 transfected cells that produced more transgene product per cell than eGFP-C1 or pEPI-eGFP transfected cells at Day 1, as was subjectively judged by the brightness and colour intensity of the positive cells (see Figure 3.7)

Figure 3.8 also indicates that the mean fluorescence intensity declined with time in all three transfected populations, which would indicate a decrease in the amount of transgene expressed per cell over time. The immunostaining images could only support this in the case of the pEPI-M18 transfected population where the intensity of the positive cells was, subjectively, lower at Day 21 than at Day 1. No difference in intensity was apparent in the case of the eGFP-C1 transfected population from Day 1 to Day 21, however both the flow cytometry and the immunostaining images suggest that the expression in the positive cells of the eGFP-C1 population at Day 21 was greater than that in the positive cells of the pEPI-M18 transfected population. No positive cells were detected by Day 21 in the pEPI-eGFP transfected population by immunostaining. However, the fact that the flow cytometry data showed that positive cells were still present within each population at Day 60, yet the immunostaining showed that by Day 60 positive cells could only be detected in the eGFP-C1 transfected population, indicates that the positive cells of the eGFP-C1 population did have the highest mean fluorescence intensity of the three transfected populations at Day 60.

The total eGFP fluorescence is shown in Figure 3.9 and, not surprisingly, shows an evident decline in the total amount of transgene expressed in all three transfected populations over time.

C2C12 cells have the ability to differentiate and fuse into myotubes when grown under low serum medium. In order to investigate the effects differentiation may have upon the expression of the transgene, $5x10^4$ cells were isolated from each transfected

population at the three time points (Day1, 21, and 60) and seeded into 1cm^2 chamber slide wells. The cells were differentiated by changing the medium to differentiation medium for 5 days (as described in Chapter 2: Materials & Methods). The cells were then fixed in 1% PFA/PBS and immunostained for eGFP. The images obtained illustrated that faintly eGFP positive myotubes were detected in the pEPI-eGFP transfected population at Day 21 (Figure 3.10) where no positive myoblasts had been observed prior to differentiation (Figure 3.7). Yet, the decline in expression was still evident for all three populations where fewer positive myotubes were present in the eGFP-C1 transfected population and no positive myotubes were detected in the pEPI-eGFP or pEPI-M18 populations at Day 60, as was the case with the undifferentiated myoblasts.

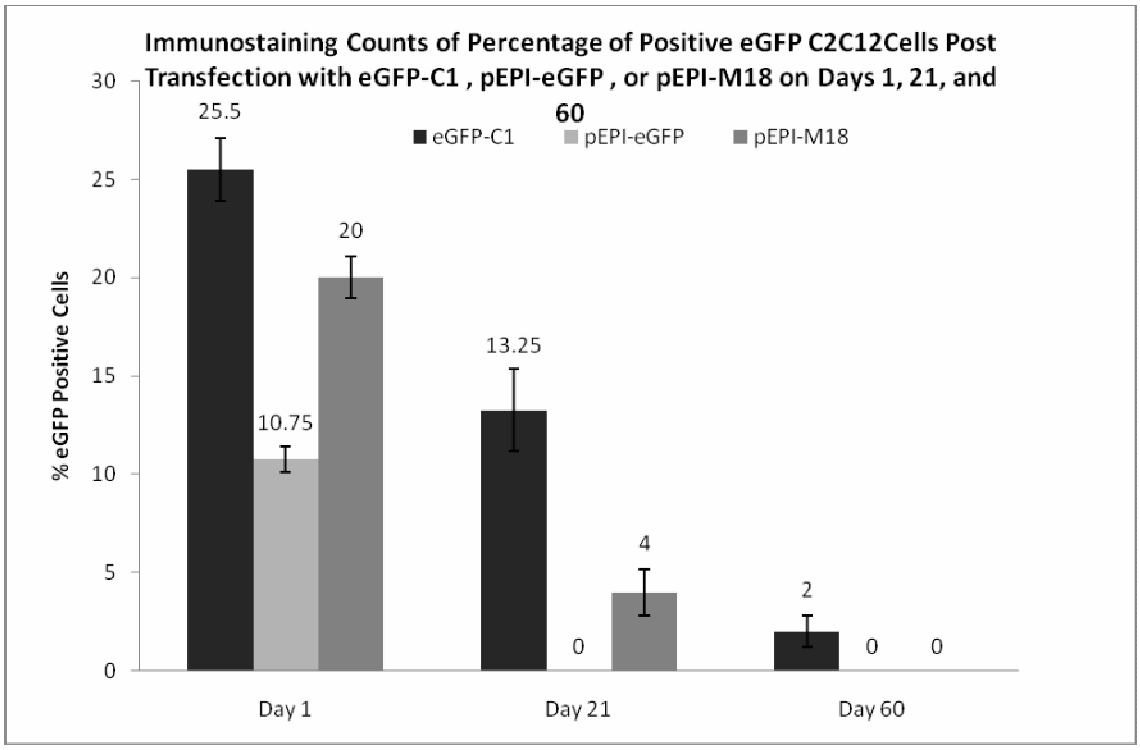


Figure 3.6(a) Mean percentages ± SEM (error bars) of eGFP positive cells derived from cell counts of immunostaining images of C2C12 cells transfected with eGFP-C1, pEPI-eGFP, or pEPI-M18 vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60). The values above the bars indicate the mean percentage of eGFP positive cells. The chart shows that the transfection efficiency of the eGFP-C1 vector was the highest of the three vectors, followed by the pEPI-M18 vector, and finally the pEPI-eGFP vector which had the lowest efficiency in this experiment. A significant decline in the percentage of positive cells was evident by Day 21 for all three transfected populations. *n* = 3.

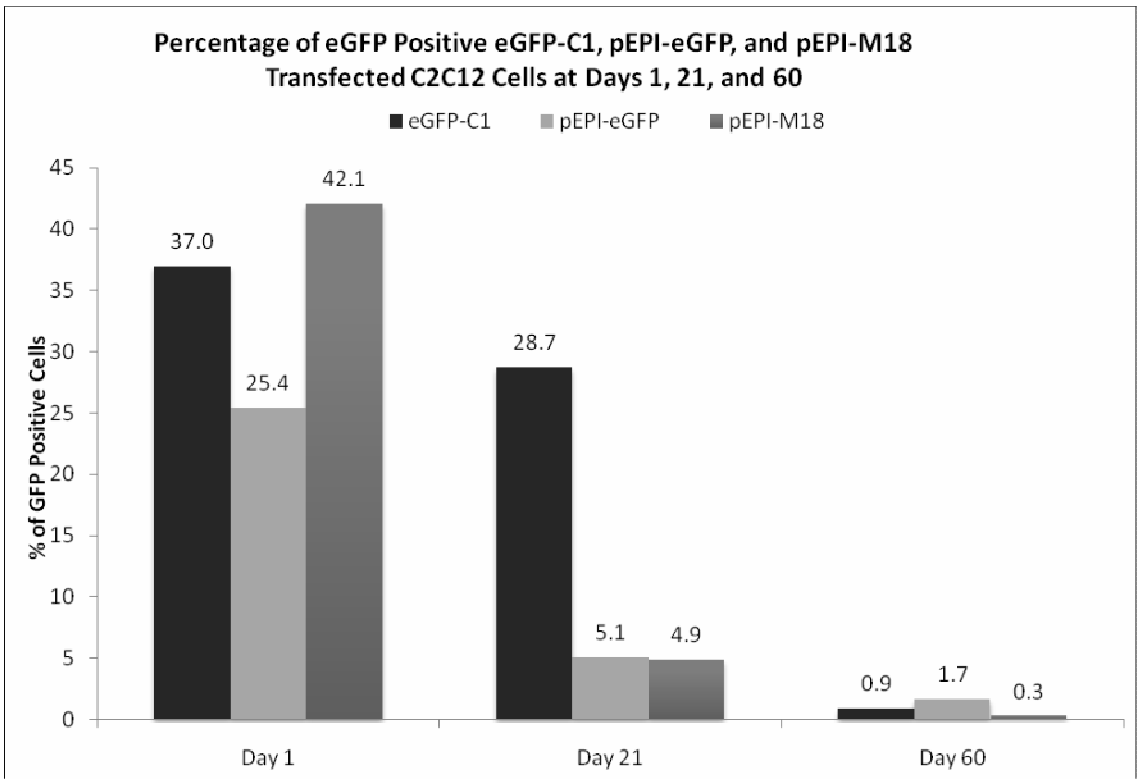


Figure 3.6(b) Flow cytometry analysis of the percentage of eGFP positive C2C12 cells transfected with eGFP-C1, pEPI-eGFP, or pEPI-M18 vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. The values above the bars indicate the mean percentage of eGFP positive cells. A decline in the percentage of eGFP positive cells with time was evident in all three transfected populations.

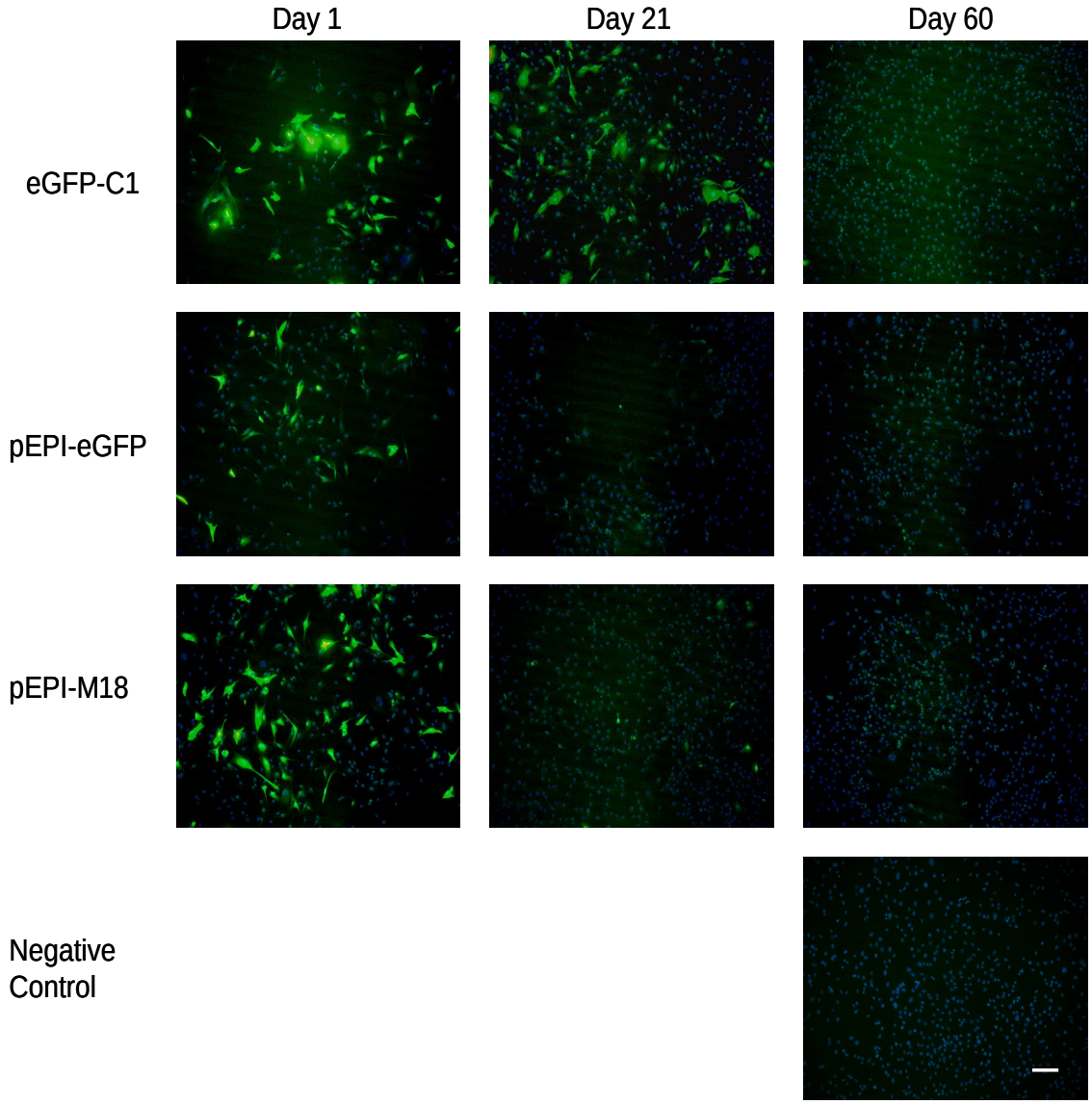


Figure 3.7 eGFP immunostaining of C2C12 cells transfected with eGFP-C1, pEPI-eGFP, or pEPI-M18 vector at 24 hours post transfection (Day 1) (images A, B, and C, respectively), the final day of selection after a 21 day selection period (Day 21) (images D, E, and F, respectively), and the 39th day of proliferation without any selective pressure post-selection (Day 60) (images G, H, and I, respectively). The negative control was untransfected C2C12 cells stained for eGFP (image J). Green indicates eGFP positive cells and blue is the DAPI staining of the nuclei. A qualitative observation of the positive cells of the pEPI-M18 transfected cells at Day 1 is they were brighter and fluoresced more intensely than those of the eGFP-C1 and pEPI-eGFP transfected cells indicating higher transgene expression in those cells. The staining also shows a decrease in the number of positive cells in all three transfected populations from Day 1 to Day 60. Bar, 90µm.

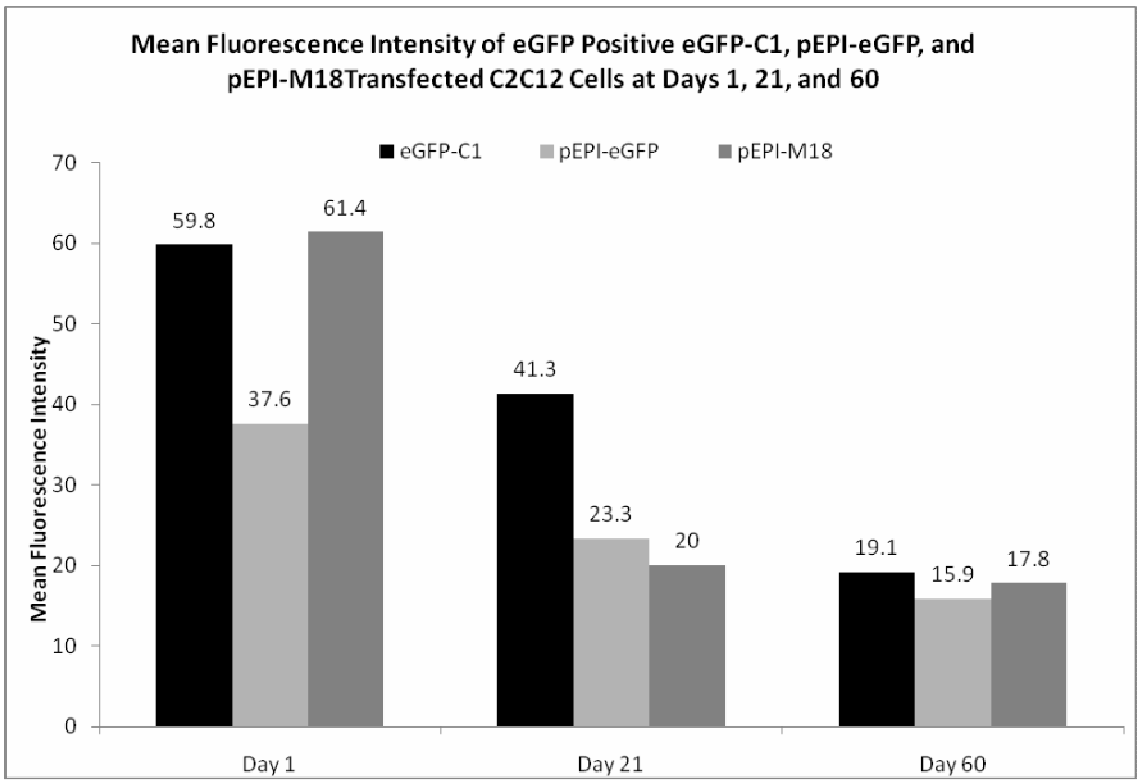


Figure 3.8 Flow cytometry analysis of the mean fluorescence intensities of eGFP positive C2C12 cells transfected with eGFP-C1, pEPI-eGFP, or pEPI-M18 vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. The values above the bars are the exact values of mean fluorescent intensity. These data suggest that eGFP-C1 and pEPI-M18 transfected cells produced similar amounts of eGFP per cell, and that there was a higher proportion of cells in these two transfected populations that produced more transgene product per cell than pEPI-eGFP transfected cells at Day 1. It also suggests that the mean fluorescence intensity declined from Day 1 to Day 60 in all three transfected populations, which would indicate a decrease in the amount of transgene expressed per cell over time.

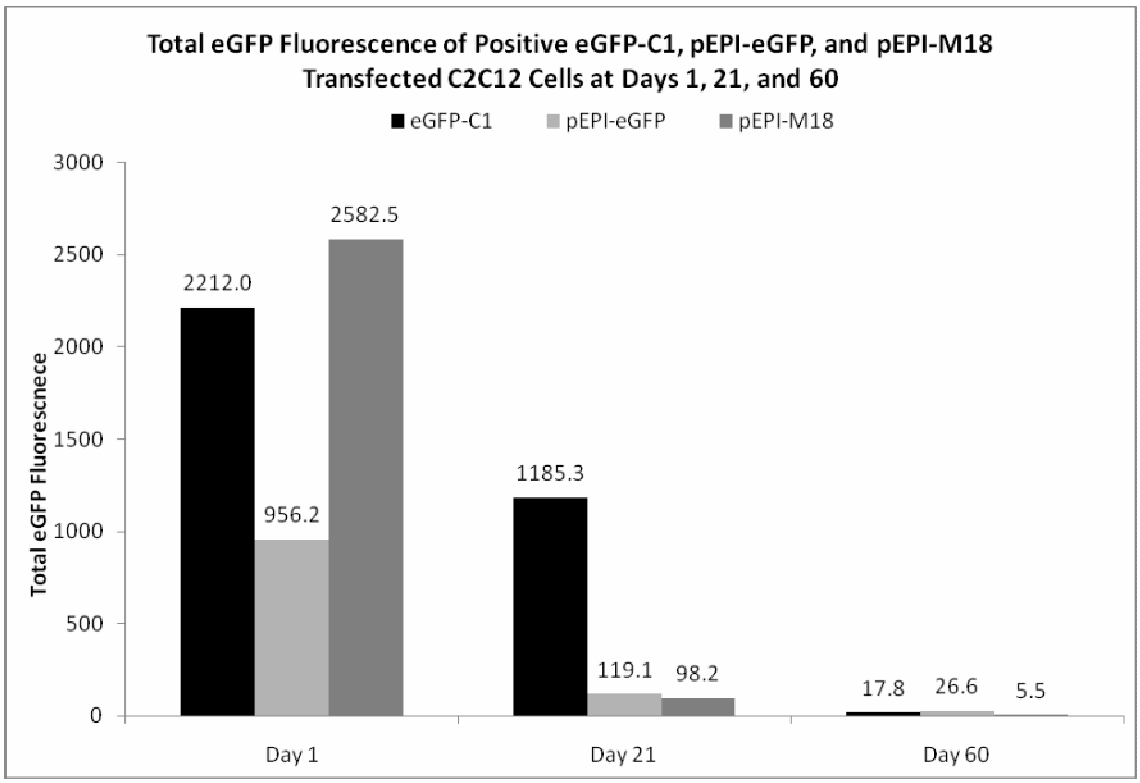


Figure 3.9 Flow cytometry analysis of the total eGFP fluorescence of eGFP positive C2C12 cells transfected with eGFP-C1, pEPI-eGFP, or pEPI-M18 vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. The values above the bars indicate the total eGFP fluorescence.The data indicates a large decline in the total amount of eGFP transgene expression in all three transfected populations over time.

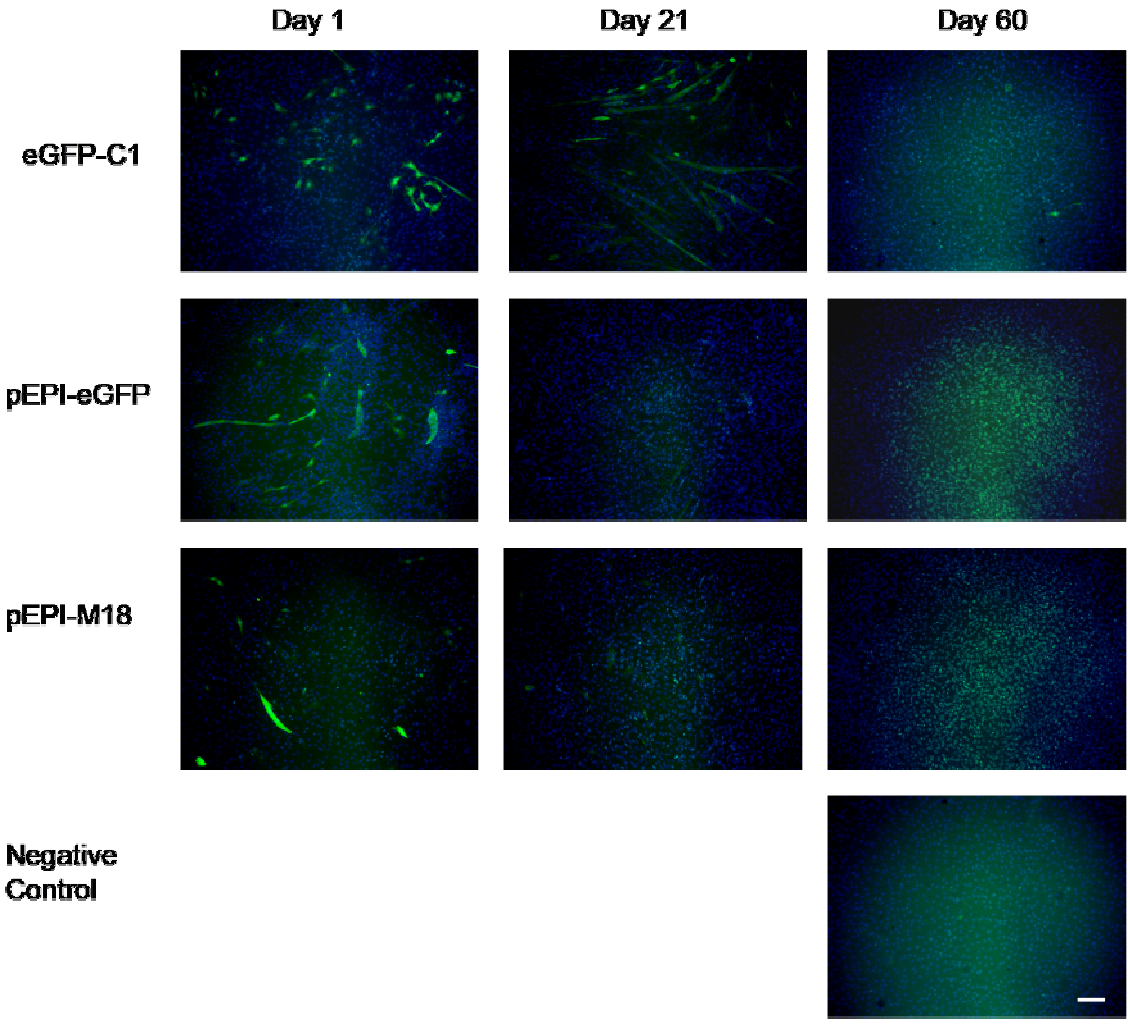


Figure 3.10 eGFP immunostaining of C2C12 cultures transfected as myoblasts with eGFP-C1, pEPI-eGFP, or pEPI-M18 vector and differentiated into myotubes under differentiation medium for 5 days prior to fixing and staining. The transfected populations were differentiated at three time points. Day 1 indicates the eGFP-C1, pEPI-eGFP, and pEPI-M18 transfected populations that were put into differentiation medium 24 hours post transfection (images A, B, and C, respectively), Day 21 indicates the populations that were put under differentiation medium on the final day of selection after a 21 day selection period (images D, E, and F, respectively), and Day 60 indicates the populations that were put into differentiation medium on the 39th day of proliferation without any selective pressure post-selection (images G, H, and I, respectively). The negative control was untransfected C2C12 myotubes (image J). Green indicates eGFP positive cells and blue is the DAPI staining of the nuclei. The images show the

presence of eGFP positive myotubes in each transfected population at Day 1 and at Day 21. At Day 60 eGFP positive myotubes were detected in the eGFP-C1 transfected population but not in the pEPI-eGFP or the pEPI-M18 transfected cells, where expression had been silenced or was expressed at levels too low to be detected by this method. Bar, 90µm.

3.2.3.3 Expression Analysis of HepG2 Cell Populations Transfected With eGFP-C1 or pEPI-eGFP

Figure 3.11 shows the mean percentages ± the standard error of the mean (SEM) (*n* = 3) as found by cell counts of the immunostaining images (a) and the mean percentages of eGFP positive cells within the populations as indicated by flow cytometry (b). The Day 1 data shown in Figure 3.11 (a) indicate that the transfection efficiency of the eGFP-C1 vector was similar to that of the pEPI-eGFP vector in this experiment (means of 32 ± 3.2% and 23.8 ± 0.5% positive cells, respectively; P = 0.07). However, Figure 3.11 (b) flow cytometry data suggests that the transfection efficiency of eGFP-C1 was higher (mean of 14.8% positive cells in the eGFP-C1 transfected population and 2.8% for pEPI-eGFP).

An increase was observed in the mean percentage of eGFP positive pEPI-eGFP transfected HepG2 cells over time. The cell counting data showed an from Day 1 to Day 21 (mean of 47.3 ± 2% at Day 21) and a further increase between Days 21 and 120 (mean of 64.3 ± 3.9% at Day 120). This finding was supported by the flow cytometry data (mean of 44.9% at Day 21 and 60.6% at Day 120) and the immunostaining images where the increase in the proportion of positive cells can be noted (see Figure 3.12). This trend was not observed in the eGFP-C1 transfected cells where the cell counting data showed that the mean percentage of positive cells in this population decreased from Day 1 (mean of 2 ± 0.6% at Day 21, and of 0% at Day 120), and that these values were significantly lower than those of the pEPI-eGFP transfected population at Days 21 and 120 (P = 0.0001). The flow cytometry data did not suggest a decline between Days 1 and 21 of the positive cells in the eGFP-C1 transfected population (mean of 14.8% positive cells at Day 1 and of 17.7% at Day 21) but did between Days 21 and 120 (mean of 2.3% at Day 120). The immunostaining images were also illustrative of this observed decline in the eGFP-C1 population see (Figure 3.12).

In contrast to the case with the HeLa and C2C12 experiments, it was noted here that the percentages of positive cells quantified by cell counting of immunostained cell images were greater than those percentages obtained by flow cytometry analysis.

Figure 3.13 shows the mean fluorescence intensity of the eGFP positive cells of the eGFP-C1 and pEPI-eGFP transfected populations as measured by flow cytometry. The data suggests that the eGFP-C1 transfected cells expressed more transgene per cell, on

average, than the pEPI-eGFP transfected cells throughout the entire period of the experiment. It also suggests that the mean fluorescence intensity continually increased in both populations from Day 1 until Day 120. The immunostaining images, which were all taken at the same exposure, suggest that the eGFP-C1 transfected cells did express a larger amount of transgene than the pEPI-eGFP transfected cells at Day 1. The images also show, subjectively, an increase in intensity in the pEPI-eGFP cells when comparing Day 1 to Day 120.

The total eGFP fluorescence is shown in Figure 3.14 and indicates a considerable increase in the total amount of transgene expressed in the pEPI-eGFP transfected population over time, which was not noted in the eGFP-C1 transfected population, where a decline was seen between Days 21 and 120.

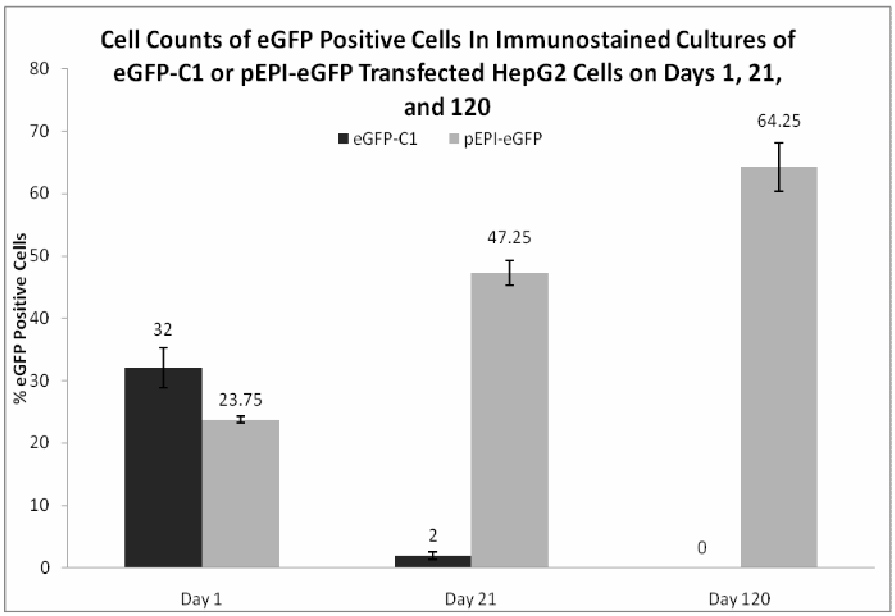


Figure 3.11(a) Mean percentages ± SEM (error bars) of eGFP positive cells derived from cell counts of immunostaining images of HepG2 cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 99th day of proliferation without any selective pressure post-selection (Day 120). The chart shows that the transfection efficiency of eGFP-C1 was similar to that of pEPI-eGFP. An increase was observed in the pEPI-eGFP transfected cells from Day 1 to Day 21 and a further increase between Days 21 and 120, whereas a decrease was shown in the eGFP-C1 transfected and the mean percentages were significantly lower than those of the pEPI-eGFP transfected populations at Days 21 and 120. *n* = 3.

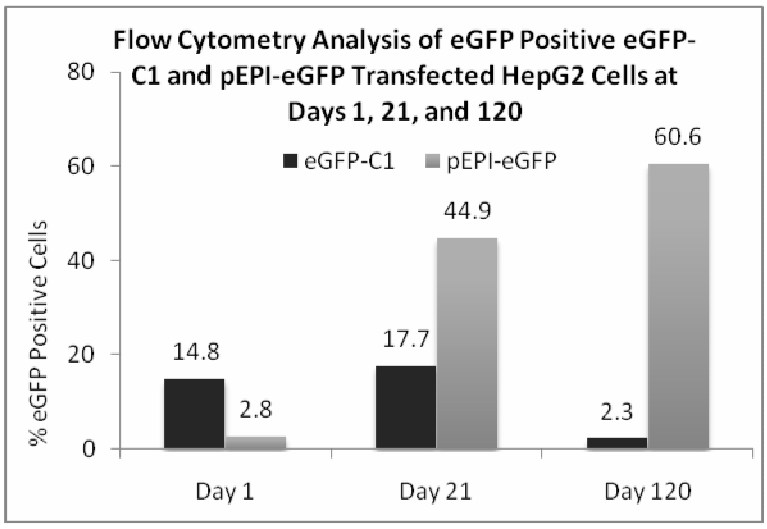


Figure 3.11(b) Flow cytometry analysis of the percentage of eGFP positive HepG2 cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 99th day of proliferation without any selective pressure post-selection (Day 120), where 1x10^{5} cells were analysed per sample. The data suggests that the transfection efficiency of the eGFP-C1 vector was higher than that of the pEPI-eGFP vector in this experiment. An increase was observed in the mean percentage of eGFP positive pEPI-eGFP transfected HepG2 cells over time, whereas a decrease had occurred in the eGFP-C1 transfected cells between Days 1 and 120.

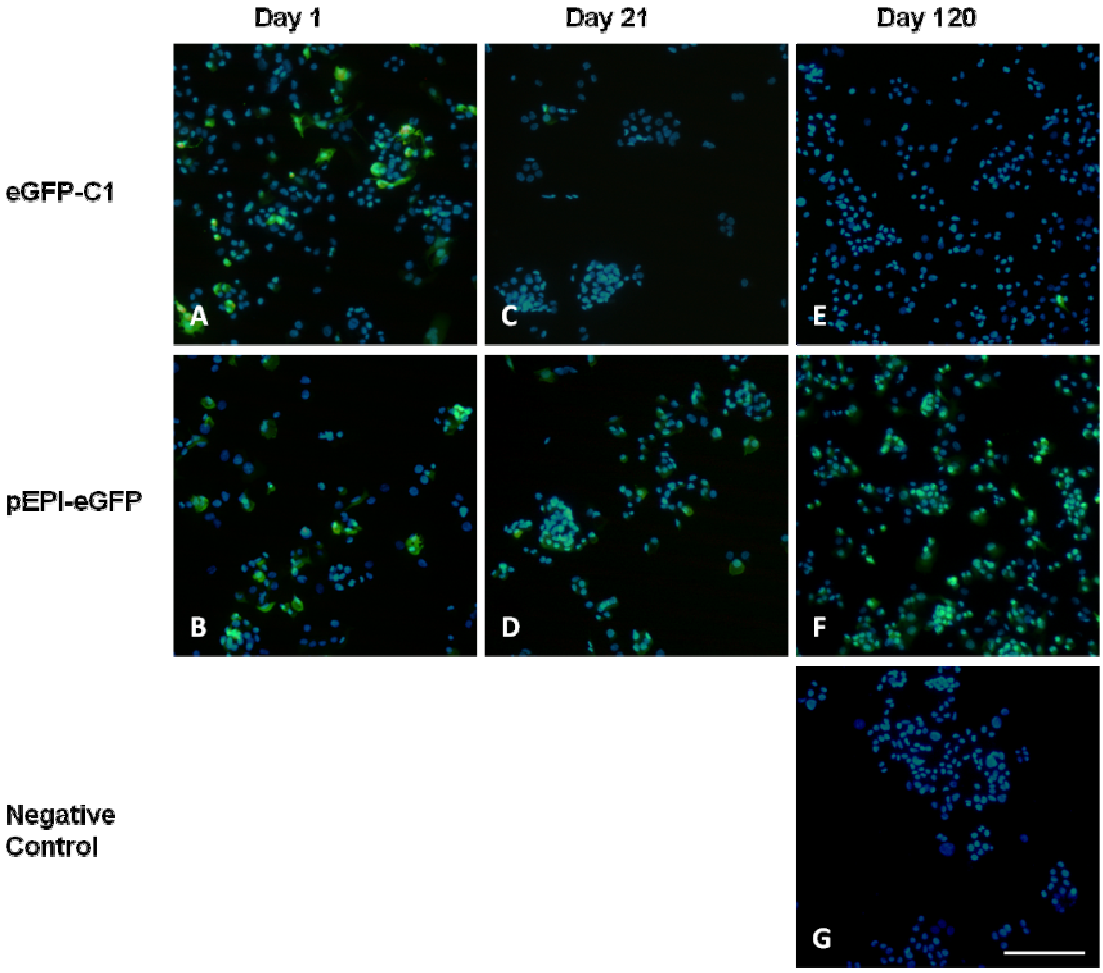


Figure 3.12 eGFP immunostaining of fixed HepG2 cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1) (images A and B, respectively), the final day of selection after a 21 day selection period (Day 21) (images C and D, respectively), and the 99th day of proliferation without any selective pressure post-selection (Day 120) (images E and F, respectively). The negative control was untransfected HepG2 cells stained for eGFP (image G). Green indicates eGFP positive cells and blue is the DAPI staining of the nuclei. These images are visual representations of the eGFP positive cells of the transfected populations and support the flow cytometry and cell counting data. A subjective observation is that the eGFP-C1 transfected positive cells fluoresce more intensely than the pEPI-eGFP transfected positive cells at Day 1. The eGFP-C1 images (top row) indicate a decline in the amount of eGFP positive cells over time. The pEPI-eGFP images (second row) show that the amount of positive cells increased from Day 1 to 21, and again from Day 21 to 120. Bar, 90µm.

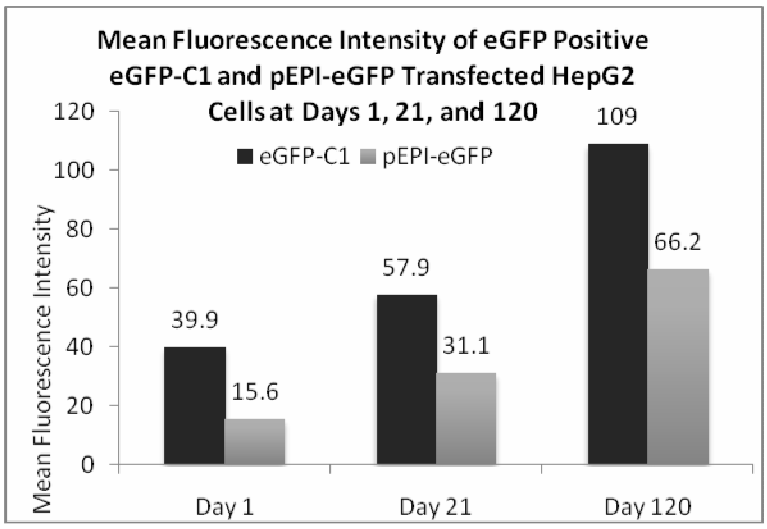


Figure 3.13 Flow cytometry analysis of the percentage of eGFP positive HepG2 cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 99th day of proliferation without any selective pressure post-selection (Day 120), where 1x10^{5} cells were analysed per sample. The data suggests that the eGFP-C1 transfected cells expressed more transgene per cell, on average, than the pEPI-eGFP transfected cells throughout the entire period of the experiment. It also suggests that the mean fluorescence intensity continually increased in both populations from Day 1 until Day 120.

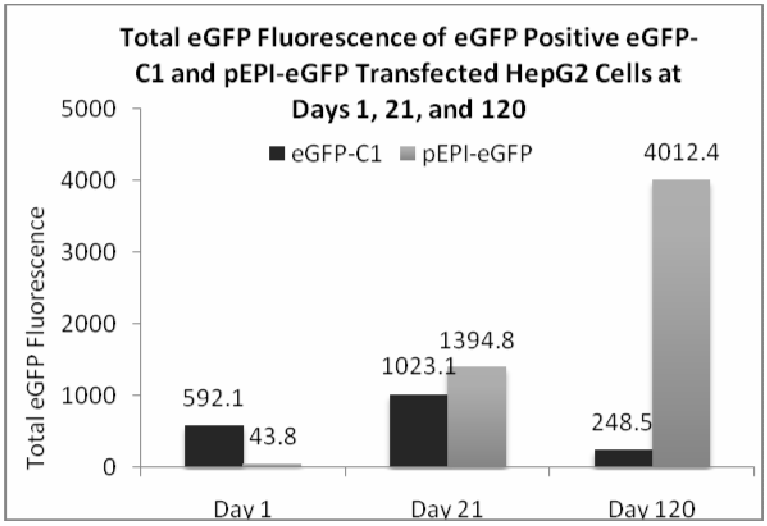


Figure 3.14 Flow cytometry analysis of the percentage of eGFP positive HepG2 cells transfected with eGFP-C1 or pEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 99th day of proliferation without any selective pressure post-selection (Day 120), where 1x10^{5} cells were analysed per sample. The data shows a considerable increase in the total amount of transgene expressed in the pEPI-eGFP transfected population over time, which was not noted in the eGFP-C1 transfected population.

3.2.4 Investigation OF Episomal/Integrant Status Of Transfected Vectors In C2C12 Cells By Southern Blotting

Southern blotting is a method used to detect specific DNA sequences and can confirm whether a vector has remained in an episomal status or has integrated into the host genome. The kit used was the Amersham ECL Direct Nucleic Acid Labelling and Detection Kit. This involved the direct labelling of the DNA probe generated by PCR of the eGFP gene (probe size 503bp) with Horseradish Peroxidase (HRP) enzyme. The labelled probe was then hybridised to complementary sequences on the blot and detected by the addition of HRP's substrate, luminol, upon which the breakdown leads to the release of light. This light was captured on autoradiography film (Materials & Methods).

Southern blotting was conducted on C2C12 cells transfected with eGFP-C1 or pEPI-eGFP plasmid. Samples were taken at Days 1, 7, 21, 30, 40, and 60 and tested. 15µg of total extracted DNA from each sample was digested overnight with *EcoRI* and loaded per lane. A band at 6.7kb, which is the size of pEPI-eGFP, can be seen in each lane from Day 1 to Day 40, and a 4.7kb band for eGFP-C1 could be seen from Day 1 to Day 30 (Figure 3.15). The blot indicates that a greater number of copies of pEPi-eGFP persisted in the C2C12 cells for more cell divisions than the eGFP-C1 plasmid as an episome. No evidence of integration was detected for either population of cells at any time point. However, the method was not sensitive enough to detect any plasmid sequence at Day 60, and it is clear that copies were present within each population as was shown in the PCR results. Therefore this method may not be able to detect integration if it had occurred at a low frequency in a small percentage of the cells. For this reason, a more sensitive method was required.

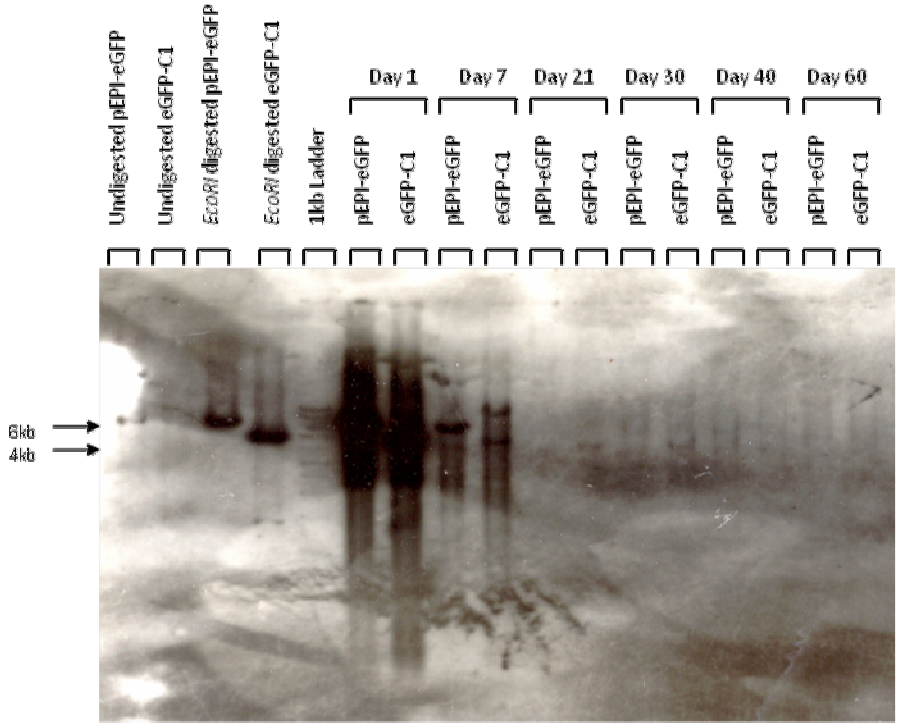
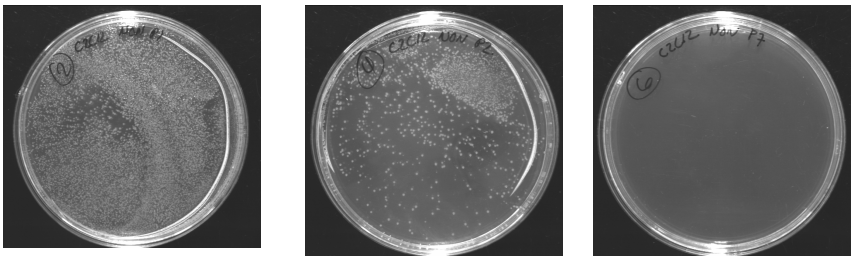


Figure 3.15 Southern blot conducted on C2C12 cells transfected with eGFP-C1 or pEPI-eGFP plasmid. Samples were taken at Days 1, 7, 21, 30, 40, and 60 post-transfection. 15µg of total extracted DNA from each sample was digested overnight with *EcoRI* and loaded per lane. A 6.7kb sized band can be seen in each pEPI-eGFP lane from Day 1 to Day 40, and a 4.7kb sized band in each eGFP-C1l lane from Day 1 to Day 30. The blot indicates that pEPI-eGFP persists in C2C12 cells for more cell divisions than eGFP-C1 as an episome. No evidence of integration can be detected for either population of cells at any time point. However, this method is not sensitive enough to detect the presence of the vector copies, as integrants or as episomes, at Day 60.

3.2.5 Investigation oF Episomal/Integrant Status of Transfected Vectors In C2C12 Cells By Plasmid Rescue

Plasmid rescue is conducted by the extraction of the total gnomic DNA of a sample and its transformation it into bacteria, which are then plated onto selective agar (Materials & Methods). The appearance of colonies indicates the presence of plasmid within the sample that was transformed and denotes the episomal status of a vector. It does not, however, negate nor confirm integration if it occurs. 1µg of total genomic DNA extracted from eGFP-C1- and pEPI-eGFP-transfected samples taken at Days 1, 7, and 21 and electroporated into XL-1 Blue *E.coli*. The eGFP-C1 sample at Day 1 yielded 8000 colonies and the pEPI-eGFP sample at Day 1 yielded 3200 colonies. The eGFP-C1 sample collected on Day 7 had 2424 colonies and the pEPI-eGFP sample had 26. At Day 21 no colonies could be detected from either the eGFP-C1 or pEPI-eGFP sample (Figure 3.16). This indicates either integration of the plasmids into C2C12 host genome, or the reduction of the plasmid copy number per cells to a level undetectable by this method. Considering no colonies were detected beyond Day 21, this method was found not to be suitable for the purposes of this experiment as it was not sensitive enough.

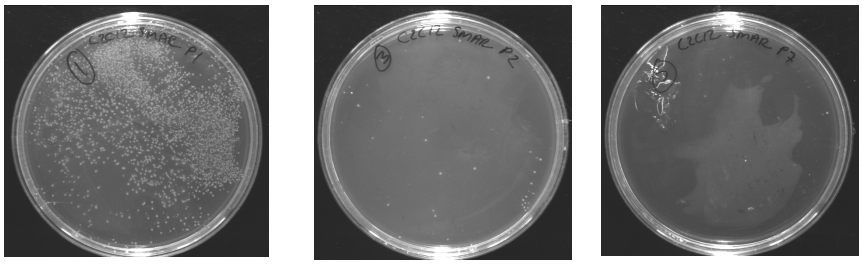
(A) eGFP-C1



Day 1 **Day 7** **Day 21**

8000 **2424** **0**

(B) pEPI-eGFP



Day1 **Day 7** **Day 21**

3200 **26** **0**

Figure 3.16 Kanamycin agar plates showing XL-1 Blue *E.coli* colonies that had been transformed by electroporation with 1µg of genomic DNA extracted from C2C12 cells transfected with eGFP-C1 plasmid (A) or pEPI-eGFP plasmid (B) from samples collected at Day 1, Day 7, and Day 21 post-transfection. Both plasmids appear to be lost at a very high rate and cannot be detected by 21 days post-transfection despite Day 21 being the last day of selection.

3.2.6 Hypothetical Calculation of Plasmid Copy Number to Determine Experimental Sensitivity Required to Detect Episomal Copies Within the C2C12 Populations At Day 60

It is clear that the sensitivity of Southern blotting and plasmid rescue used in this chapter have proven insufficient for the purpose of detecting the vector sequence, episomal or integrated in status, in the multiclonal populations transfected and tested. No vector sequences could be detected by the final day of the experiment although the data from the flow cytometry analysis indicated that some cells in the samples continued to express eGFP, in addition to the PCR data which clearly showed that copies of the vector sequences initially transfected were present within some cells of the populations.

An analysis was performed here which calculated the hypothetical number of plasmid copies that may be found in the transfected populations by the end of the experiment. The number of copies can then be converted to an amount in grams, which would indicate the experimental sensitivity required to detect episomes. The calculations for the analysis require a number of assumptions. The first is that the number of cells containing the episomal vector was the percentage observed by flow cytometry, so if 1.7% of cells proved eGFP positive in the pEPI-eGFP-transfected C2C12 cells, then it was assumed that that was the proportion of cells still carrying the vector. The second assumption was that each of the positive cells had an average of 8 plasmid copies per cell. This was based on previous studies which had reported average copy numbers to be 2.5-13 per cell (Schaarschmidt et al in 2004, Papapetrou et al 2006), therefore 8 was chosen as the average. Further, the size of the mouse genome was estimated to be $2.7x10^9$ bp, respectively (www.genome.gov National Human Genome Research Institute).

The calculations were conducted is as follows:

C2C12 cells are of murine origin, therefore the size of the genome per cell is $2.7x10^9$ bp. If 15µg were used per sample for Southern Blotting, this would be the equivalent to using the collective DNA of $4.49x10^6$ cells. This was calculated according to the following formula:

$$\#\text{ of DNA molecules} = \frac{(\text{total amount of DNA (ng)}) \times (6.022x10^{23}\ \text{molecules/mole})}{(\text{length of DNA molecule (bp)}) \times (1x10^9) \times 650)}$$

The number $6.022x10^{23}$ is Avogadro's constant. This is the number of molecules present in one mole. The number 650 represents the average weight of 1 base pair in double stranded DNA. In order to find the molecular weight of a double stranded DNA molecule, the weight of one base pair is multiplied by the number of base pairs of the molecule. In order to calculate the number of moles per gram from the molecular weight of a molecule, the inverse of this value is taken and multiplied by $1x10^9$ in order to obtain the answer in nanograms. It was calculated that 0.005ng, or 5.05pg, of pEPI-eGFP would be present per 15µg of total DNA, therefore the sensitivity of the Southern blot would have been required to be in the scale of picograms, and considering plasmid rescue required only 1µg of total DNA, its sensitivity would need to be in the femtograms.

The values were calculated for the S/MAR populations only since these were the populations that contained the highest percentages of eGFP positive cells by the final day of the experiment and would therefore theoretically contain the largest amount of vector sequence copies. As can be seen, the amounts are very low and would not be possible to detect using the Southern blotting or plasmid rescue methods employed. It is clear that with multiclonal populations there is a greater difficulty in detecting the vector than if a single clone was picked and grown, where the same integration events would be present in each cell.

3.2.7 Investigation OF Episomal/Integrant Status Of Transfected Vectors In HeLa, C2C12, and HepG2 Final Day Samples By Fluorescent In-Situ Hybridisation (FISH) Analysis

FISH was used to determine the episomal and/or integrant status of the transfected vectors, as conducted in several other S/MAR vector studies (Stehle *et al*., 2007; Baiker *et al*., 2000). Samples of $1x10^5$ cells were taken from each transfected population at the final day of the experiment (Day 60 for HeLa and C2C12 and Day 120 for HepG2 transfected populations) and were arrested at metaphase to yield condensed chromosomal structures. The cells were fixed and dropped onto aged slides to burst the nuclei and expose the DNA, which was then probed for the transfected vector sequences. The probe was generated by random priming using eGFP-C1 as a template, where dTTPs were in direct competition with labelled dUTP, thus resulting in labelled probes in a range of sizes (as described in Chapter 2: Materials & Methods). Following hybridisation and washing, the bound probe was visualised directly by fluorescence microscopy at 100X magnification. Probe detected around the chromosomes or associated with them indicate episomal status of the vector. Probe detected at the same position on both arms of a chromosome suggest integration of the vector into the host genome, as sister chromatids of a chromosome are replicates of each other and the probe would therefore bind at the same position on both chromatids. However, the possibility that the plasmid sequence preferentially binds to certain sequences present on both arms of each chromatid also cannot be excluded, therefore any evidence towards integration that is detected must remain potential. Also, the stronger the signal, the more probe is bound, which indicates a larger amount of plasmid sequence present.

An average of 50 spreads were analysed, and of those, 5-10 positive spreads were found and analysed for each sample. Of the positive spreads, the number of integrated spreads, if present, was counted. The negative controls in this experiment were untransfected HeLa, C2C12, and HepG2 cells (Figures 3.17 (C), 3.19 (C), and 3.21 (C), respectively). The positive controls used for this experiment were HeLa (Figure 3.17 (A) and (B)), C2C12 (Figure 3.19 (A) and (B)), and HepG2 (Figure 3.21 (A) and (B)) cells 24 hours post transfection with eGFP-C1 or pEPI-eGFP plasmid. From the images taken under the fluorescence microscope it can be seen that a large amount of plasmid was present within the positive control cells, as the fluorescence signal emitted was very

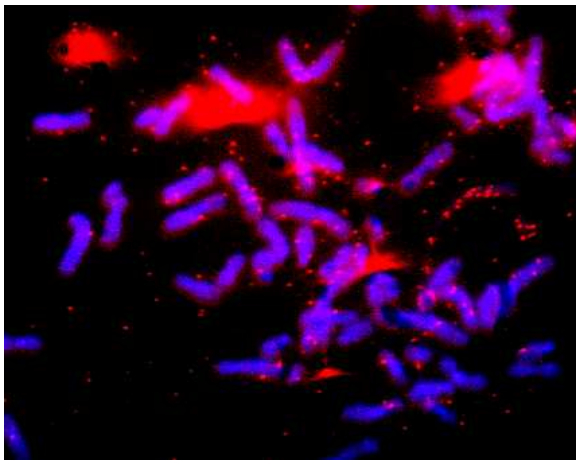
strong. It is also clear that much of the plasmid observed was clustered together. A subjective observation is that eGFP-C1 transfected HeLa, C2C12, and HepG2 cells more effectively than pEPI-eGFP, as the amount of signal spots emmiting fluorescence in the eGFP-C1 images was greater than those seen in the pEPI-eGFP images in all three cell lines.

FISH analysis on the Day 60 sample of the HeLa cells transfected with eGFP-C1 or pEPI-eGFP suggests that pEPI-eGFP remained present as an episome, with no evidence of integration (0/6 spreads) (Fig. 3.18 (A)). eGFP-C1 appears to have integrated into the genome of transfected HeLa cells (1/6 spreads), however episomal copies of the vector were also present (Figure 3.18 (B)).

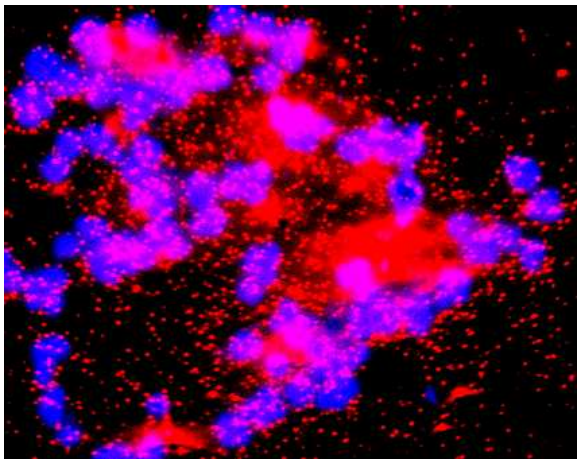
FISH analysis on Day 60 of the C2C12 cells transfected with eGFP-C1, pEPI-eGFP, or pEPI-M18 indicates that pEPI-eGFP remained present as an episome with no evidence of integration (0/6 spreads) (see Figure 3.20 (A)), whereas eGFP-C1 (see Figure 3.20 (B)) and pEPI-M18 (see Figure 3.20 (C)) vectors had been found on duplicate arms of the same chromosome, which suggested integration into the host genome (5/7 in the eGFP-C1 transfected population and 3/8 in the pEPI-M18 transfected population), in addition to episomal copies of both vectors being present.

FISH analysis on Day 120 of the HepG2 cells transfected with the pEPI-eGFP or eGFP-C1 showed no evidence of integration of either vector into the host genome (0/6 and 0/5 spreads, respectively), and episomal copies of the vectors were present within positive cells (see Figure 3.22 (A) and (B), respectively). A subjective observation was that the pEPI-eGFP transfected positive cells contained a greater amount of episomal copies of the vector pEPI-eGFP at Day 120 in comparison to the amount of eGFP-C1 episomal copies in the eGFP-C1 transfected positive cells.

(A) HeLa cells 24 hr after transfection with pEPI-eGFP



(B)) HeLa cells 24 hr after transfection with eGFP-C1



(C)HeLa cells untransfected

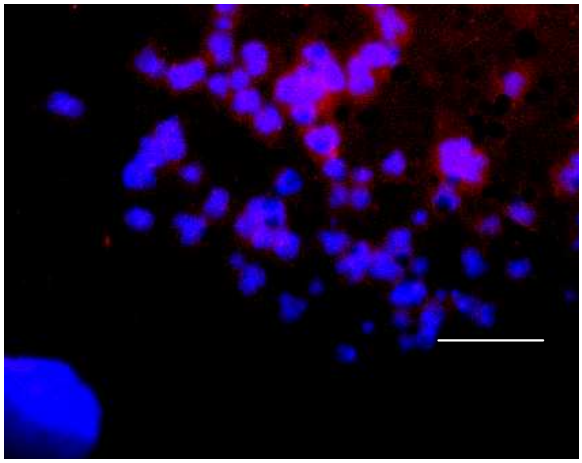
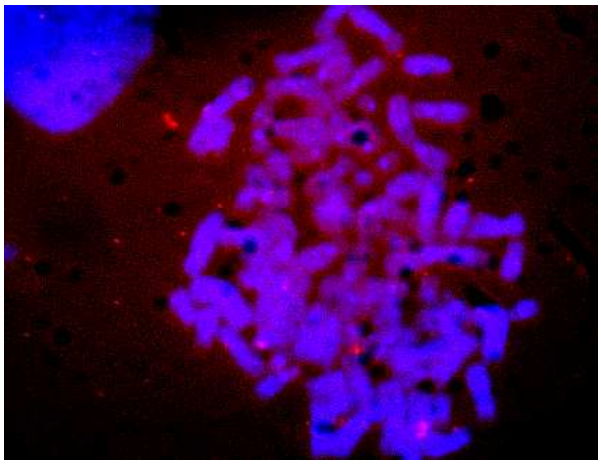


Figure 3.17 Fluorescent In-Situ Hybridisation positive controls of HeLa cells 24 hours post transfection with pEPI-eGFP (A) or eGFP-C1 (B) vector. Vector copies can be seen as the bright red fluorescing spots, and genomic DNA is blue as it was counterstained with Dapi. The vector was seen as mostly clustered together and associated with the chromosomes. There was, subjectively, a greater amount of eGFP-C1 vector present within the eGFP-C1 transfected cells than there was pEPI-eGFP vector present within the pEPI-eGFP transfected cells. Image (C) was the negative control of untransfected HeLa cells. Bar, 5µm.

Transfected HeLa Cells Day 60

(A) pEPI-eGFP



(B) eGFP-C1

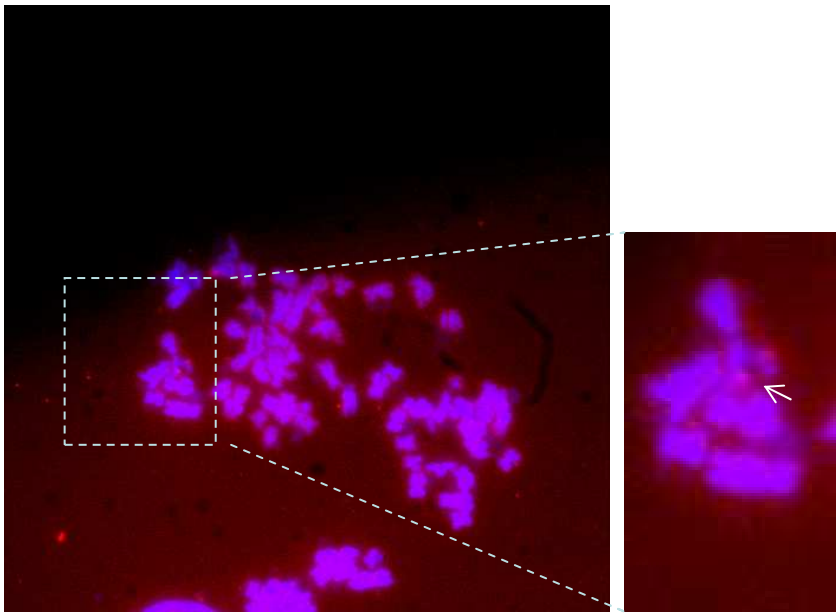
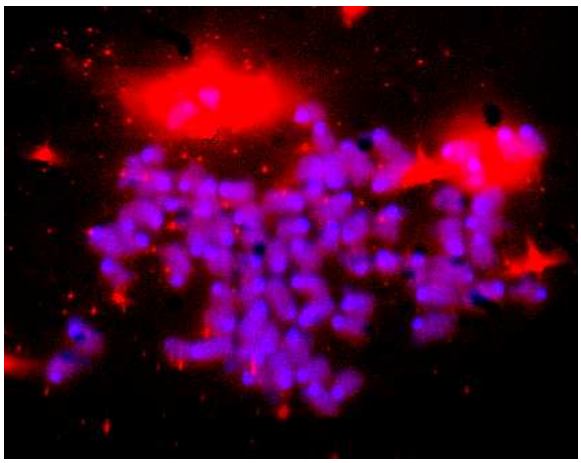
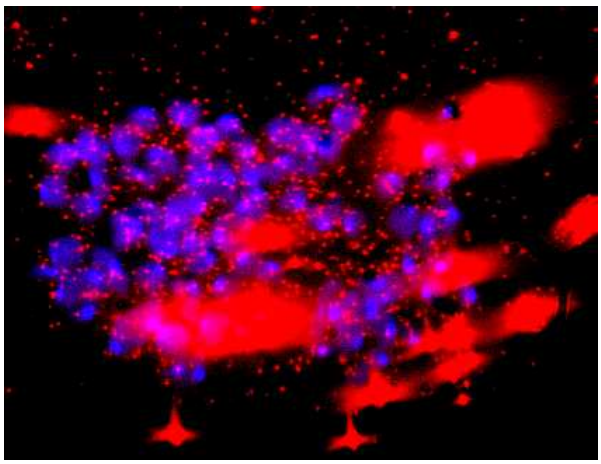


Figure 3.18 Fluorescent In-Situ Hybridisation of HeLa cells transfected with pEPI-eGFP (A) or eGFP-C1 vector (B), Day 60 samples tested. Genomic DNA is counterstained with Dapi (blue). Image (A) shows that pEPI-eGFP is present as an episome with no evidence of integration. Image (B) suggests integration of eGFP-C1 into the HeLa host genome (1 out of 6 spreads), in addition to being retained as an episome.

(A) C2C12 cells 24 hr after transfection with pEPI-eGFP



(B) C2C12 cells 24 hr after transfection with eGFP-C1



(C)C2C12 cells untransfected

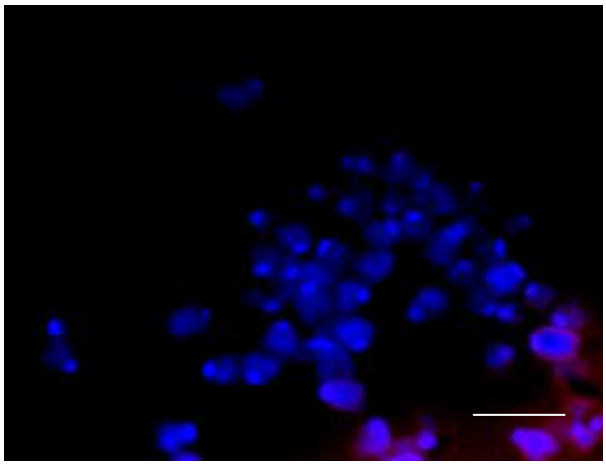
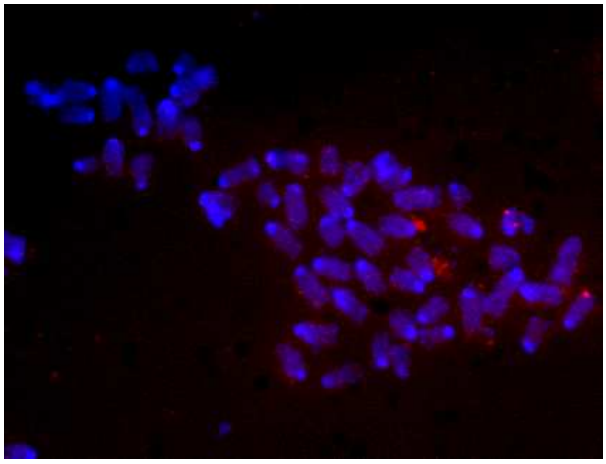


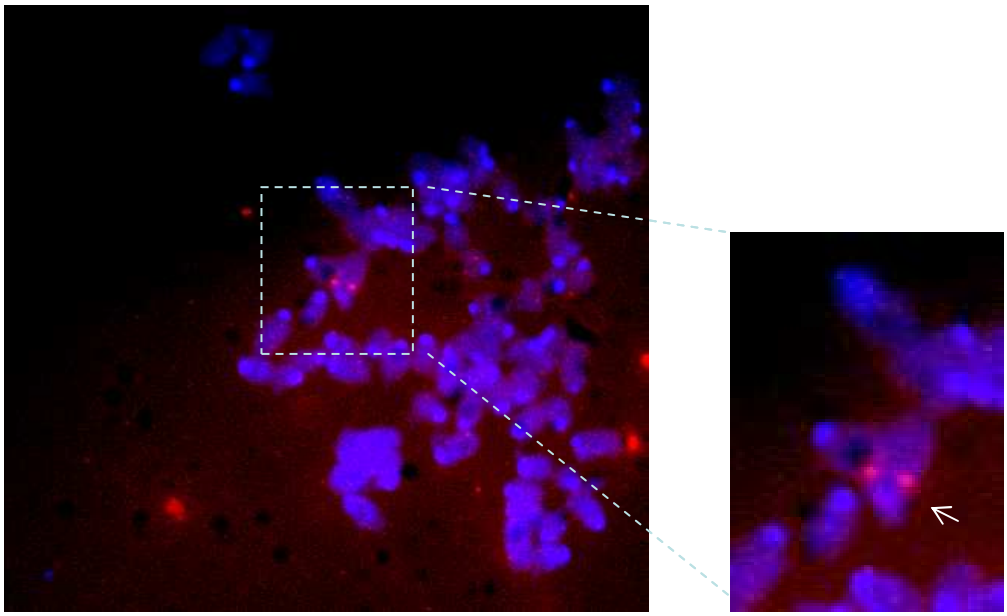
Figure 3.19 Fluorescent In-Situ Hybridisation positive controls of C2C12 cells 24 hours post transfection with pEPI-eGFP (A) or eGFP-C1 (B) vector. Vector copies can be seen as the bright red fluorescing spots, and genomic DNA is blue as it was counterstained with DAPI. The vector was seen as mostly clustered together and associated with the chromosomes. There appeared to be a greater amount of eGFP-C1 vector present within the eGFP-C1 transfected cells than there was pEPI-eGFP vector present within the pEPI-eGFP transfected cells. Image (C) was the negative control of untransfected C2C12 cells. Bar, 5µm.

Transfected C2C12 cells Day 60:

(A) pEPI-eGFP



(B)eGFP-C1



(C)pEPI-M18

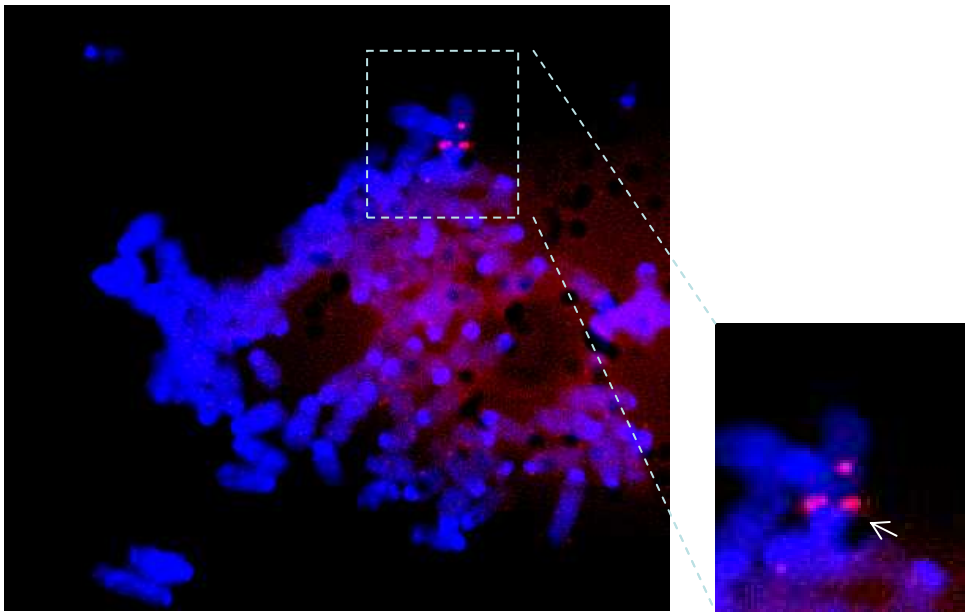
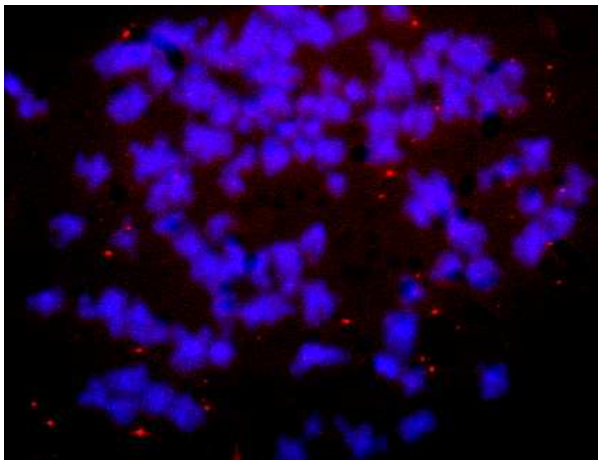
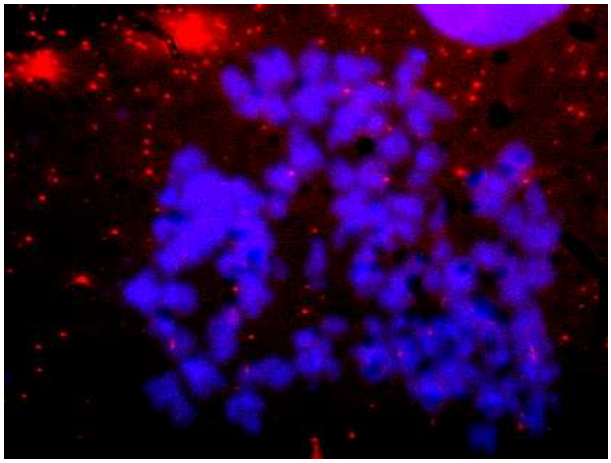


Figure 3.20 Fluorescent In-Situ Hybridisation of C2C12 cells transfected with pEPI-eGFP (A), eGFP-C1 (B), or pEPI-M18 vector (C), Day 60 samples tested. Red fluorescent spots indicate where the probe has hybridised to plasmid sequence. Genomic DNA is counterstained with DAPI (blue). Image (A) shows pEPI-eGFP present as an episome with no evidence of integration (0/6 spreads). Images (B) and (C) suggest eGFP-C1 and pEPI-M18 integration onto the C2C12 host genome (5/7 and 3/8 integrant spreads, respectively) in addition to some copies being present as episomes.

(A) HepG2 cells 24 hr after transfection with pEPI-eGFP



(B) HepG2 cells 24 hr after transfection with eGFP-C1



(C)HepG2 cells untransfected

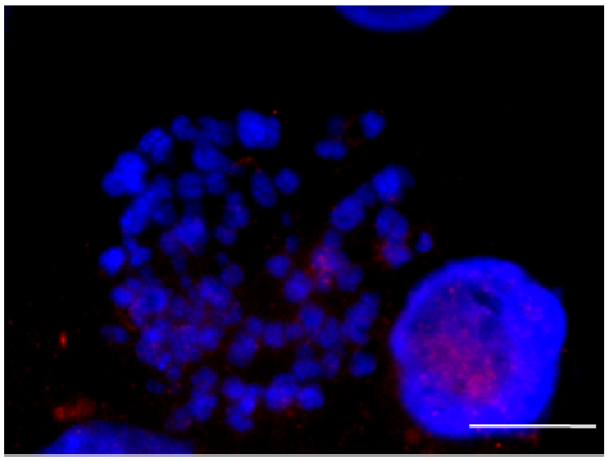
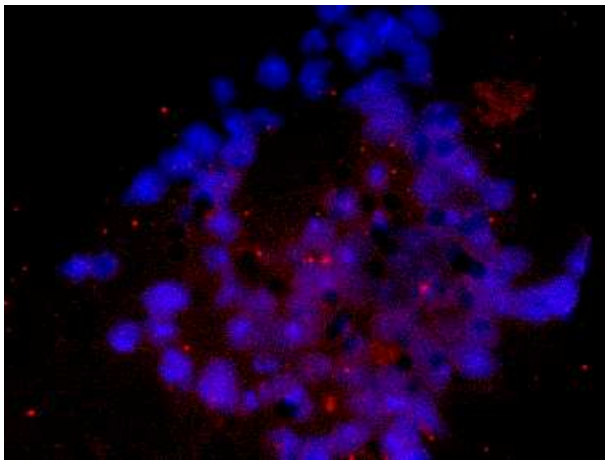


Figure 3.21 Fluorescent In-Situ Hybridisation positive controls of HepG2 cells 24 hours post transfection with pEPI-eGFP (A) or eGFP-C1 (B) vector. Vector copies can be seen as the bright red fluorescing spots, and genomic DNA is blue as it was counterstained with DAPI. The vector was seen as mostly clustered together. There was, subjectively, a greater amount of eGFP-C1 vector present within the eGFP-C1 transfected cells than there was pEPI-eGFP vector present within the pEPI-eGFP transfected cells. Image (C) was the negative control of untransfected HepG2 cells. Bar, 5µm.

Transfected HepG2 cells Day 60

(A) pEPI-eGFP



(B)eGFP-C1

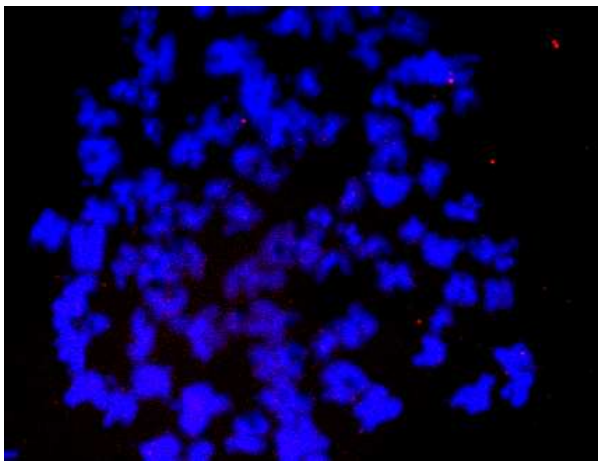


Figure 3.22 Fluorescent In-Situ Hybridisation of HepG2 cells transfected with pEPI-eGFP (A) and eGFP-C1 vector (B), Day 60 samples tested. Red fluorescent spots indicate where the probe has hybridised to plasmid sequence. Genomic DNA is counterstained with DAPI (blue). Images (A) and (B) show that pEPI-eGFP and eGFP-C1 were present as episomes with no evidence of integration into the HepG2 host genomes (0/5 and 0/6 spreads, respectively). Image (A) suggests that there is a larger number of episomes in the pEPI-eGFP transfected cells than the eGFP-C1 transfected cells.

3.2.8 Summary of Results: HeLa Cell Line

		PCR	Cell Counts (% positive cells)	Flow Cytometry			Immunostaining	FISH	
				% positive cells	MFI	Total		# of Slides Indicating Integration	Episomal
Vector	Day								
eGFP-C1	1	√	50.3 ± 3.4	62.4	91.6	5714.9	++	1/6	√
	21	√	2.3 ± 0.7	9.4	26.7	249.9	+		
	60	√	1.25 ± 0.4	0.3	34.4	9.6	+		
pEPI-eGFP	1	√	48.5 ± 4.8	54.5	55.3	3006.1	++	0/6	√
	21	√	4 ± 1.5	18	24.9	448.0	+		
	60	√	0	2.3	10.4	23.5	-		

Table 3.2 Table summarising the results obtained in this chapter for transfected HeLa cells, indicating each vector that was transfected and the time point at which samples were extracted and analysed. PCR results are presented where a '√' denotes that the Kanamycin sequence was amplified and a band was present; Cell Counts indicates the percentage of positive cells as counted from immunostained images of transfected populations; Flow Cytometry results presented include the percentage of positive cells per transfected population, the mean fluorescence intensity (MFI) of the positive cells, and the total amount of expression which was the product of expression and number of positive cells; the Immunostaining data is presented by '-' indicating the absence of positive cells, '+' indicating the presence of less than 10 positive cells, and '++' indicating the presence of more that 10 positive cells; the FISH data is presented as the number of slides which suggested an integration event having occurred out of the total number of slides analysed on Day 60 only, a '√' in the episomal column indicated the presence of episomal copies of the vector observed on the slides, and a '-' indicates none was observed.

3.2.9 Summary of Results: C2C12 Cell Line

Vector	Day	PCR	Cell Counts	Flow Cytometry			Immunostaining		Southern Blotting	Plasmid Rescue	FISH	
				% positive cells	MFI	Total	Myoblasts	Myotubes			Integrant Samples	Episome Present
eGFP-C1	1	√	25.5± 1.6	37	59.8	2212	++	++	√	√	5/7	√
	21	√	13.3 ± 2.1	28.7	41.3	1185.3	+	++	√	-		
	60	√	2 ± 0.8	0.9	19.1	17.8	+	+	-	-		
pEPI-eGFP	1	√	10.8 ± 0.7	25.4	37.6	956.2	+	+	√	√	0/6	√
	21	√	0	5.1	23.3	119.1	-	+	√	-		
	60	√	0	1.7	15.9	26.6	-	-	-	-		
pEPI-M18	1	√	20 ± 1.1	42.1	61.4	2582.5	++	+	N/A	N/A	3/8	√
	21	√	4 ± 1.2	4.9	20	98.2	+	+	N/A	N/A		
	60	√	0	0.3	17.8	5.5	-	-	N/A	N/A		

Table 3.3 Table summarising the results obtained in this chapter for transfected C2C12 cells, indicating each vector transfected and the time point at which samples were analysed. PCR results are presented where a 'v' denotes that the Kanamycin sequence was amplified and a band was present; Cell Counts indicates the percentage of positive cells as counted from immunostained images of transfected populations; Flow Cytometry results presented include the percentage of positive cells per transfected population, the mean fluorescence intensity (MFI) of the positive cells, and the total amount of expression (the product of expression and number of positive cells); Immunostaining data is shown by '-' indicating no positive cells, '+' indicating the presence of less than 10 positive cells, and '++' indicating more that 10 positive cells; Southern Blotting data is presented as a 'v' to indicate the presence of a band,'-' to indicate no band, and N/A if the experiment was not conducted with the samples; Plasmid rescue data is shown with a 'v' to indicate the presence of colonies after transformation, and '-' to indicate no colonies; FISH data is presented as the number of slides suggesting integration out of the total number of slides analysed on Day 60 only, a 'v' in the episomal column indicated the presence of episomal copies of the vector observed on the slides, and a '-' indicates none was observed.

3.2.10 Summary of Results: HepG2 Cell Line

		PCR	Cell Counts (% positive cells)	Flow Cytometry			Immunostaining	FISH	
				% positive cells	MFI	Total		Integration	Episomal
Vector	Day								
eGFP-C1	1	√	32 ± 3.2	14.8	39.9	592.1	++	0/6	√
	21	√	2 ± 0.6	17.7	57.9	1023.1	+		
	120	√	0	2.3	109	248.5	-		
pEPI-eGFP	1	√	23.8 ± 0.5	2.8	15.6	43.8	++	0/5	√
	21	√	47.3 ± 2	44.9	31.1	1394.8	++		
	120	√	64.3 ± 3.9	60.6	66.2	4012.4	++		

Table 3.4 Table summarising the results obtained in this chapter for transfected HepG2 cells, indicating each vector that was transfected and the time point at which samples were extracted and analysed. PCR results are presented where a 'v' denotes that the Kanamycin sequence was amplified and a band was present; Cell Counts indicates the percentage of positive cells as counted from immunostained images of transfected populations; Flow Cytometry results presented include the percentage of positive cells per transfected population, the mean fluorescence intensity (MFI) of the positive cells, and the total amount of expression which was the product of expression and number of positive cells; the Immunostaining data is presented by '-' indicating the absence of positive cells, '+' indicating the presence of less than 10 positive cells, and '++' indicating the presence of more that 10 positive cells; the FISH data is presented as the number of slides which suggested an integration event having occurred out of the total number of slides analysed on Day 120 only, a 'v' in the episomal column indicated the presence of episomal copies of the vector observed on the slides, and a '-' indicates none was observed.

3.3. Discussion

3.3.1 pEPI-eGFP showed no evidence of integration into HeLa, C2C12, or HepG2 host genome

Consistent with other studies conducted, where a β-IFN S/MAR vector was transfected into replicating cells and integration of the vector into the host genome was investigated (as covered in Chapter 1: Introduction, Section 1.10), the S/MAR vector pEPI-eGFP was found not to integrate in the host genome of any of the three cell lines tested in this study, indicating that the S/MAR element may have had a protective role against integration of the vector.

3.3.2 Levels of transgene expression varied between transfected cells in HeLa, C2C12, and HepG2 cells

The expression studies indicated that the levels of expression of the eGFP transgene differed between cells of the same line in a population transfected with the same vector. In each population, some cells were brightly eGFP fluorescent whereas others were barely visibly fluorescent. Where some cells may have contained enough copies which led to high levels of eGFP expression, thus making the eGFP more easily detectable in positive cells, other cells may have contained such few copies that eGFP was not detectable, but the copy number was sufficient for the expression of adequate levels of Neomycin resistance gene transcripts, thus conferring resistance and enabling survival under selection. This is plausible considering that these transfected populations were multiclonal, and is supported by a study by Papapetrou *et al* where it was found that, although an S/MAR episome can be passed down to progeny cells, it is not divided amongst them equally, and the equal distribution of episomes between daughter cells is not regulated by the cell (Papapetrou *et al*., 2006). Therefore, where vectors had remained episomal, in this case it was the pEPI-eGFP vector in the HeLa, C2C12, and HepG2 cells and the eGFP-C1 vector in the HepG2 cells as shown by FISH, and the episome numbers were variable between cells, variable total transgene expression resulted per cell. And where the vector sequences were present as integrants, in this case the eGFP-C1 and pEPI-M18 vectors in HeLa and C2C12 cells,

variable amounts of integration events may have occurred in each cell, also leading to variable transgene expression per cell.

3.3.3 The comparison of flow cytomtery to cell counting of eGFP positive immunostained cells in HeLa, C2C12, and HepG2 transfected populations

It can be noted that the percentages of positive cells in each HeLa and C2C12 transfected population, as quantified by cell counting of immunostained cell images, varied from those percentages obtained by flow cytometry analysis, as the former values were generally found to be lower. Considering that flow cytometry involves the use of a detector which is able to detect fluorescence at levels which may not be within a range that can be detected by the naked human eye alone, a greater percentage of low eGFP-expressing cells can be identified, leading to greater values of total positive cells being counted within a population in comparison to those obtained by the method of cell counting. However, a different case was observed with the HepG2 cells, which may be attributed to the method by which the HepG2 cell type grows in culture, differing to that of the HeLa and C2C12 cells. The difference lies in the fact that HepG2 cells tend to grow in clumps or tight clusters, as can be observed in the immunostained images (see Figure 3.12), and therefore the clusters may have been present in the samples being quantified for eGFP positive cells. In a flow cytometer, such clusters which contained positive cells may have been quantified as a single cells, whereas in cell counts of immunostained images, the DAPI-stained nuclei clearly define each individual cell, leading to the quantification of a greater number of positive cells within clusters, and hence a greater total percentage of positive cells.

3.3.4 Most eGFP-C1, pEPI-eGFP, and pEPI-M18 transfected cells which survived selection were not eGFP positive in transfected HeLa and C2C12 populations

A decline in the percentage of eGFP positive cells and transgene expression in eGFP-C1 and pEPI-eGFP transfected Hela cell populations and eGFP-C1, pEPI-eGFP, and pEPI-M18 transfected C2C12 cell populations was observed over time. This decline was, surprisingly, noted from Day 1 to Day 21, after a 21 day period of selection, where

100% of cells surviving selection must have contained vector copies so that expression of the Neomycin resistance gene was at a level sufficient for survival.

One factor that may have accounted for the decline observed in eGFP expression in the transfected HeLa and C2C12 cells was the status of the transfected vector within the cells. If integrated into the genome, as was suggestively the case for the eGFP-C1 and pEPI-M18 vectors, the site of integration of the vector would have determined the levels of expression of the transgene. Schubeler *et al* explained that when an S/MAR was located at a distance of 4kb from the transcription start site within genomic DNA expression was sustained, whereas a distance less than 2.5kb led to the suppression of gene expression (Schubeler *et al*., 1996). Therefore, if the vector had integrated into a region of euchromatin, where the DNA was frequently transcribed, and was within optimal proximity to a genomic S/MAR, leading the vector sequence to be in a position easily transcribed by the transcription machinery at the nuclear matrix, then expression would have been expected to be high (Schubeler *et al*., 1996). If it had integrated into heavily packaged DNA such as heterochromatin, where expression was low and infrequent, and/or at a position that was at a non-optimal distance from a genomic S/MAR, then vector transgene expression would have been expected to be low (Schubeler *et al*., 1996).

In addition to the position of the integrants within the genome, the orientation in which the vector had integrated would also have had a significant effect on expression. When a plasmid is integrated, certain elements within it may get disrupted as it changes from a circular to a linear form. In the case where any part of the Neomycin resistance cassette was disrupted by integration and the cells were unable to express the gene, these cells would have perished under selection and would have ceased to exist within the cell population. Integration that occured where any part of the eGFP expression cassette was disrupted, such as the eGFP gene, its promoter, or its polyadenylation signal, would have led to the disruption of eGFP transcription, and could thus have led to Neomycin resistant cells that did not express eGFP, as eGFP expression was not essential for cell survival under selection.

As for the non-integrating vector pEPI-eGFP, the decline in expression and percentage of positive cells can be accounted for differently. This decline had also been observed in the study by Papapetrou *et al* using pEPI-eGFP in K562 and MEL cells (as described in Chapter 1: Introduction section 1.10), where low copy numbers of the vector were

present within the transfected populations, yet expression levels had declined dramatically (Papapetrou, *et al*., 2006).

One factor which may have contributed to this was the use of selection. Schubeler *et al* explained that with the use of selection a selection bias is created where cells expressing below a threshold level of the protein required for resistance will be eliminated, and those where expression is initially very high may have expression silenced and would thus also be eliminated from selection (Schubeler *et al*., 1996). It is plausible that the decline in intensity observed in these populations from Day 1 to Day 21 was due to the elimination of many cells that may have expressed eGFP but did not express sufficient Neomycin resistance protein to survive selection, and were thus eliminated from the population.

Another factor may be that the CMV promoter driving eGFP transgene expression had been silenced.

3.3.5 CMV promoter may be responsible for transgene silencing in HeLa and C2C12 cells

One possibility for the observed decline in eGFP expression despite cell survival throughout selection is transgene silencing. The PCR and FISH results showed that copies of the vectors were still present within cells of the transfected populations by the final day of selection, however the flow cytometry and immunostaining showed that a very low proportion of the population were expressing eGFP. Yet, these cells were still able to express Neomycin in order to survive selection. This indicated that a certain factor was specifically disrupting eGFP transgene expression, which may have been a cellular-mediated silencing of the CMV promoter.

The silencing of gene expression can be a result of the methylation of a promoter by mammalian cells, and in this case, the methylation of the CMV promoter driving the eGFP transgene. In the study by Argyros *et al* the CMV promoter was analysed and was found to contain 3 main CpG-rich regions which form a 444bp island (Argyros *et al*., 2008). In addition to this island were other CpG rich areas within this promoter that were equally able to be methylated (Argyros *et al*., 2008). More disruptive to expression is the recruitment of methyl-DNA binding proteins and histone deacetylases that can further change chromatin structure and block transcription (as

reviewed in Chapter 1: Introduction, Section 1.7.2). Thus methylation and its implications may, wholly or in part, be the cause for the observation that a large proportion of Neomycin resistent cells were not simultaneously eGFP positive, both in cases where vector had or had not integrated. One way to investigate this would have been to add 5'-Azacytidine, a methyltransferase inhibitor, to the medium of each cell population, at a time point where the cells were still Neomycin resistant but had lost eGFP fluorescence, such as several days before the final day of selection. The inhibition of DNA methylation would have led to an observable increase in eGFP transgene expression if the cause of the repression had been, at least in part, due to the methylation of the CMV promoter.

It has been proposed that CMV promoter activity is meant to be a transient one, and that this characteristic is what allows CMV to be so ubiquitously expressed (Mocarski *et al*., 1990). It is also known that the CMV promoter can be downregulated, even where stably integrated (Mocarski *et al*.,1990). This downregulation is due, at least in part, to cellular factors which influence CMV-driven expression, and could have contributed to the decline observed in eGFP expression. Different cell types express different factors which lead them to express a given transgene at varying efficiencies, sometimes as a result of the promoter driving transgene expression. This was shown in the study previously described by Papapetrou *et al* where different types of cells were transfected with the same vectors and it was found that silencing occurred in one cell type, the MEL cells, but not in the K562 or $CD34^+$ enriched cells (Papapetrou *et al*., 2006) (as described in Chapter 1: Introduction, section 1.10). In the case of the CMV promoter, studies have shown that the human CMV virus has a limited number of tissues that it infects, and the cells in which it can replicate must express the transcription factors necessary to transcribe its genes, namely the CMV immediate-early enhancer/promoter (Koedood *et al*., 1995). In cells where the replication of the virus cannot be sustained, the transcription factors required for this promoter are lacking, and without the transcripts controlled by this promoter, no replication can occur (Mocarski *et al*., 1990). In an experiment by Mocarski transgenic mice were created expressing a CMV-driven transgene (Mocarski *et al*., 1990). This served as a representative of the tissue types the CMV virus can naturally replicate in, and would explain the variation observed in expression between cell types and tissues when using the CMV immediate-early enhancer/promoter. Some of the tissues which Mocarski

found it to infect included endothelial cells, smooth muscle cells of the arteries, the inner ear and eye, nerve cells, kidneys, the pancreas, and some parts of the liver and spleen. It was found that the amount of expression varied between tissues, which showed that CMV-driven expression was cell-type dependent.

The CMV promoter is made up of 4 types of repeat elements which have been found to bind specific nuclear proteins and form complexes (Boshart *et al*., 1985; Ghazal *et al*., 1987). These repeats are the 17bp, 18bp, 19bp, and 21bp repeats, each found 3-5 times in the promoter/enhancer region of CMV. The 18bp element has a consensus sequence which allows the Nuclear Factor κB (NFκB) to bind. NFκB is a transcription factor that is known to play a role in the expression of several cellular and viral genes (Baeuerle and Henkel, 1994). The 19bp repeats can bind the cAMP Response Element-Binding (CREB)/Activating Transcription Factor (ATF). The binding of these factors has been shown to activate and enhance CMV activity (Hunninghake *et al*., 1989; Stamminger *et al*., 1990), and it is thought that for CMV to achieve its full activity the transcription factors NFκB and CREB/ATF must be present. It is interesting to note that CREB and ATF are believed to be important transcription factors involved in the early expression of viral genes (Adam *et al*., 1996), especially considering that CMV is a viral promoter. In a study by Niller *et al* several agents were used to stimulate the production of these factors in transfected cells containing a transgene expression cassette controlled by the CMV promoter (Niller *et al*., 1991). It was found in MRC5 and HeLa cell cultures that CMV had become strongly re-activated (Niller *et al*., 1991). Thus the decrease in CMV-driven expression observed in the HeLa cells may be attributed, in part, to insufficient amounts of NFκB, the gene of which is almost inactive in HeLa cells (Niller *et al*., 1991) and CREB/ATF. Although C2C12 myoblasts have also been found to express NFκB (Guttridge *et al*., 1999) and CREB (Mori *et al*., 1993), as do HeLa cells, it is possible that low amounts are expressed, which may have contributed to the decline in CMV-driven eGFP expression in these cells.

3.3.6 Long term CMV-driven transgene expression was observed in HepG2 cells

The HepG2 cells transfected with the eGFP-C1 and pEPI-eGFP plasmids exhibited a remarkably different pattern of expression than the HeLa and C2C12 populations, as was shown by the increase in mean fluorescence intensity of the positive cells in each

population with time. This may, in part, be due to cellular factors, as it has been found that NFkB is expressed in HepG2 cells, and is present in a constitutively active form (Pogliaghi *et al*., 1995). CREB-1 has also been found to be highly expressed in HepG2 cells, and ATF-1 has also found to be expressed (Whetstine and Matherly, 2001). Hence the presence of these factors may have account for the constitutive expression of CMV-driven eGFP transgene expression within this cell line.

Interestingly, the CMV promoter driving the expression of a luciferase transgene was found to be silenced *in vivo* in mouse liver over time in a study by Argyros *et al,* whereas this was not the case in the study in this chapter involving the human hepatocyte cell line HepG2 (Argyros *et al*., 2008). This discrepancy may be explained by the availability of different factors in HepG2 cells which may not be found in liver cells *in vivo,* as it was found that in rat liver cells, although NFkB was found to be present, it was in an inactive form where its activation required induction by stress factors and inflammatory cytokines (Pogliaghi *et al*., 1995; Grimm and Baeuerle, 1993).

3.3.7 Differentiation of C2C12 myoblasts may ameliorate CMV-driven eGFP expression

Although it was difficult to determine whether expression was higher in the myotubes than it was in the myoblasts based on the immunostaining images, the images did reveal positive myotubes in the pEPI-eGFP transfected population in the Day 21 sample where no positive myoblasts had been detected prior to differentiation. This could be explained in one of two ways.

The first is that, in addition to factors that are required for continuous CMV-driven expression that bind consensus sequences on the promoter, there also exist factors which negatively influence promoter activity which may stop being expressed by the cell upon differentiation, leading to enhanced expression. Kothari *et al* believe that the 21bp repeat in the CMV promoter binds to factors that negatively regulate the promoter's activity, such as the transcriptional activator Multiprotein Bridging Factor 1 (MBF1) and the transcription factor YY1 (Kothar *et al*.,. 1991), and repress expression, leading to silencing (Kothari *et al*., 1991; Sinclair *et al*., 1992). When these 21bp repeats were deleted in an experiment by Kothari *et al*, an increase in expression was observed in undifferentiated T2 (teratocarcinoma) cells (Kothari *et al*., 1991). Kothari

et al also observed the presence of MBF1 in undifferentiated monocytes (Kothari *et al*., 1991). Once these cells were differentiated, MBF1 was undetectable in the cells and CMV activity was amplified. In differentiated cells the amount of MBF1 decreases and expression is seen to increase. Gossett *et al* showed that in C2C12 cells MBF1 is expressed in myoblasts and, upon differentiation, its expression is significantly reduced (Gossett *et al*., 1989). Gossett *et al* also found that myoblasts could not express this factor at higher levels, as this would interfere with their ability to differentiate (Gossett *et al*., 1989). This may suggest that eGFP transgene expression had been ameliorated by differentiation due to the reduction in MBF1. Studies by Loser *et al* suggest this is plausible as he observed that CMV-driven expression was improved in his studies involving the differentiation of several different cell types (Loser *et al*., 1998).

The second possibility for the observed improvement in expression may have also, in part, been due to an accumulation of eGFP within the terminally differentiated C2C12 myotubes which had not been diluted by cell division.

3.3.8 The pEPI-M18 vector containing the 'mini-S/MAR' showed evidence of integration into the C2C12 host genome

Although it has been shown in several studies (Manzini *et al*., 2006; Jenke *et al*., 2005; Schaarschmidt *et al*., 2004; Papapetrou *et al*., 2006) that a plasmid containing an S/MAR sequence within the open reading frame of a gene prevents the plasmid from integrating into the host genome, and as was also demonstrated in this study by the use of FISH on the Day 60 samples of the pEPI-eGFP transfected HeLa, HepG2, and C2C12 populations, FISH conducted on the sample derived at Day 60 of the pEPI-M18 transfected C2C12 population showed evidence that the plasmid containing a fragment of the S/MAR, termed 'mini-S/MAR', within the eGFP open reading frame, appeared to have integrated into the host genome. If truly integrated, then one possible explanation for this is that the 'mini-S/MAR' sequence did not include one or more essential elements present in the β-IFN S/MAR which gave the β-IFN S/MAR the ability to confer protective qualities to its vector, such as the prevention of integration. The results obtained in a study by Jenke *et al* may assist in explaining this finding, where a fragment of the original S/MAR element was inserted into a plasmid as a dimer or as a tetramer (Jenke *et al*., 2004). The aim of their study was to identify the

minimal essential part of the S/MAR that would function as well as the full length sequence. In the study, a 155bp sequence was isolated, chosen on the basis that it was the core unwinding element of the full length S/MAR where specific factors that also associate with the nuclear matrix are able to bind. It was inserted within the transcription unit in the place of the full length S/MAR as a dimer or tetramer sequence. Southern blotting and FISH indicated that all the vectors containing the dimer sequence had intergrated into the genome, whereas those carrying the tetramers did not and remained episomal. Jenke *et al* believed that they were able to discover the minimal sequence necessary for the S/MAR to function, and that SAF-A, which is a DNA-matrix binding protein, is involved in the recognition of this minimal sequence (Jenke *et al*., 2004). Kip *et al* and Fackelmayer and Richter explained that SAF-A can bind to DNA at its AT-rich domains (Kipp *et al*., 2000; Fackelmayer and Richter, 1994). When a cluster of such a domain is available, it allows the protein to bind more strongly to the sequence. Jenke *et al* concluded that multimers can be recognised and bound to SAF-A successfully only if they are of a minimum critical length, below which they will not be recognised (Jenke *et al*., 2004). This length requirement was fulfilled by the tetramer but not by the dimer. In the study, the dimer was 310bp, the tetramer was 465bp. The 'mini-S/MAR' in this study was 733bp and was above said critical length and was therefore not expected to integrate. Yet, the FISH data showed potentially that integration had occurred. Therefore, it is possible that the 'mini-S/MAR' did not include the specific 155bp sequence necessary to be recognised by SAF-A, which were included in the dimer and tetramer elements in the study by Jenke *et al,* thus leading to integration (Jenke *et al*., 2004). To date no studies have been conducted on the mini-S/MAR element and it is possible that the sequence does not contain essential elements of the β-IFN S/MAR, such as the SAF-A binding site, that may confer or contribute to the S/MAR's ability to prevent integration and allow episomal retention of a transfected vector. The potential integration of the pEPI-M18 vector further highlights the importance of certain regions within the S/MAR element which contribute to this ability, without which this ability would be lost.

3.3.9 Episomal copies of each vector were found in transfected cells at the final day of the experiment in HeLa, C2C12, and HepG2 populations

Positive cells analysed by FISH showed that copies of each vector were present as episomes within transfected cells in HeLa and C2C12 cells 60 days post-transfection, and in HepG2 cells after 120 days. It is interesting to note that even in the cases where evidence had suggested vector integration into the host genome, such as eGFP-C1 or pEPI-M18, episomal forms of the vectors were still present within cells by the final day of the experiment. Although these vectors may have been getting diluted out of the populations by cell division, this shows that foreign plasmid DNA can remain within cells for many generations as separate extrachromosomal units without necessarily integrating under selection. A method for quantifying the number of vector copies within eGFP positive cells is Quantitative PCR (Q-PCR), and would have been a useful tool in the analysis of vector copy number per cell in these experiments.

3.3.10 The amount of eGFP present in HepG2 cells transfected with eGFP-C1 or pEPI-eGFP increased with time

The increase observed in the mean fluorescence intensity in the transfected HepG2 cells was progressive. This suggests that eGFP expression had accumulated, and eGFP protein was passed on to daughter cells. Daughter cells which inherited the vector sequence continued to express more eGFP transgene, which lead to a further increase in the amount of the protein, which was once again passed down to the progeny cells, and so on, leading to a cumulative increase in the mean fluorescence intensity with time. An interesting experiment to investigate this would be the use of Q-RT-PCR in order to quantify the amount of eGFP mRNA transcripts present within the cell populations.

3.3.11 Transfection of HepG2 cells with pEPI-eGFP lead to increased numbers of positive cells over time

In the pEPI-eGFP transfected HepG2 population the percentage of positive cells rose continually over time. The increase observed in the percentage of positive cells after a

period of selection in this population was due to the death and elimination of cells not containing the vectors. Furthermore, seeing as the number of eGFP positive cells and the total eGFP increased in the pEPI-eGFP transfected population much more significantly than in the eGFP-C1 cells from Day 1 to Day 21, it can be inferred that the pEPI-eGFP vector copies were retained in a manner which allowed better expression, and/or were of a higher average vector copy number per cell.

It is difficult to explain the rise in the percentage of positive cells observed between Days 21 and 120 where no selection pressure was applied and therefore the cells that may have lost copies of the vector would have been able to survive and replicate. A possible explanation for this is that when the number of positive cells was quantified at the end of the selection period many cells negative for eGFP were still within the culture and were still in the process of dying. Following their death most of the population was then eGFP positive, and copies of the vector were very effectively retained and expressed from Day 21 until Day 120, thus indicating that the vector was effectively retained and passed on to daughter cells as an episome, considering no evidence of integration was found when investigated using FISH.

Another explanation for the increase in the percentage of positive cells is that during selection, by retaining the vector sequence, the positive cells had also acquired a growth advantage besides the growth/survival advantage conferred by the Neomycin expression cassette, which led to increased levels of proliferation in these cells within the population, which were then able to dominate the population due to the speed of their growth, even after selection was removed. Considering that the cells were not under selection between Days 21 to 120 it is difficult to infer as to the nature of this growth advantage.

3.3.12 eGFP-C1 was lost with HepG2 cell divisions over time leading to a decrease in positive cells

As with the transfected HeLa and C2C12 populations, the percentage of eGFP-C1 transfected HepG2 cells declined over time. The fact that the mean fluorescence intensity continued to rise over time in spite of the decline in the percentage of positive cells within the population could denote that the decline in the percentage of

positive cells could be attributed to a loss in episome copies with cell division, and not due to transgene silencing.

3.3.13 eGFP-C1 does not always integrate into the host genome under selection

It is also interesting to note that evidence towards eGFP-C1 integration does not always occur in every population transfected with it and grown under selection, as was the case with the HepG2 cells. Although the possibility of low frequency integration cannot be excluded, it was not detected. It is possible that the slower rate of cell division of these cells led to a slower rate of vector dilution, where the Neomycin expression cassette was transcribed at an adequate amount in positive cells during selection, which eliminated the need for vector integration into the host genome in order to survive selection.

3.3.14 pEPI-eGFP was effectively retained in HepG2 cells due to the presence of the S/MAR element in the vector

It is clear that the pEPI-eGFP vector allowed continuous, long term transgene expression, without CMV promoter silencing, and with no evidence of integration into the host genome in HepG2 cells. It is also evident that without the S/MAR element, the vector would not have been retained as well within the cells, and the number of positive cells within the population would have declined, as was shown with the eGFP-C1 vector. This could be attributed to potential of the S/MAR element to be epigenetically marked, leading to retention within the cells and their replication and passing on to daughter cells (as shown in studies described in Chapter 1: Introduction section 1.10).

It has been shown that S/MAR vector transgene expression and the efficiencies of being retained in replicating cell lines can be variable according to the cell line (as described in Chapter 1: Introduction section 1.10). Yet, S/MAR elements found intrinsically in cells all serve similar functions, regardless of the cell type. Therefore, it is possible that certain qualities of HepG2 cells may have contributed to the success of pEPI-eGFP retention within these cells over many generations. One main and obvious

difference between the HepG2 cells and the C2C12 and HeLa cells is their rate of division. It is possible that the longer length of the cell cycle in HepG2 cells allowed the reporter gene to have more time to be expressed and accumulated thus leading to high levels within the cells. The slower rate of division may have also carried other benefits, such as allowing more time from the moment the S/MAR plasmid entered the nucleus to the time of cell division for the plasmid to be epigenetically marked, as the cell would normally mark endogenous DNA, or allowing the cell more time to mark a larger amount of vector before replication. This may have led to a higher number of vector copies retained per cell, and thus a higher level of total expression.

In a study by Courbet *et al* different rates of DNA replication, and hence fork rate progression, were found to influence the number of origins that were attached to the nuclear matrix and, consequently, those which were initiated during replication in different cell lines (Courbet *et al*., 2008). This is significant considering that the S/MAR element has been found to attach to the nuclear matrix and act as an origin of replication (as reviewed in Chapter 1: Introduction, Section 1.9). Thus, the efficiency of the initiation of replication of pEPI-eGFP in this study may have been influenced by the rate of DNA replication in each cell line used in this experiment.

In the study by Courbet *et al* it was found that initiation sites were determined in G1 (Courbet *et al*., 2008), a step called origin decision point (Li *et al*., 2003), and that the attachment of origins of replication to the matrix may be required for the initiation of replication from these origins (Courbet *et al*., 2008). Courbet *et al* used cell lines 471 and 474, isolated from Chinese hamster lung fibroblast GMA32 cells, where each line was found to have a different rate of DNA synthesis during replication (Courbet *et al*., 2008). In 471 cells, the replication fork was found to travel quickly, at an average speed of 1.3kb/min, whereas in 474 cells the average rate was 0.6kb/min. These differences in speed, it was found, were directly related to the density of initiation events. In the 'fast' 471 cells, initiation events occurred at a low density, where 80% of initiation events in a region of DNA occurred at previously characterised origin of replication, ori*GNAl3,* and the remaining events were distributed between 5 other separate origins. In the 'slow' 474 cells, a higher density of events were observed, where initiation had occurred evenly between ori*GNAl3* and the other 5 origins. Hence, during fast conditions, large replicons were formed due to the anchorage of origins with high affinities to the nuclear matrix, and during slower conditions, smaller

replicons were formed as a larger number of origins, which may not necessarily have had high affinities to the nuclear matrix, were also anchored and initiated.

HepG2 cells have a relatively long doubling time in comparison to those of HeLa and C2C12 cells. The doubling time for HepG2 cells is ~48hrs (www.cell-lines-services.de), for HeLa is ~22hrs (Mu *et al*., 1997), and for C2C12 is ~12hrs (Kitzmann *et al*., 1998). It has also been shown that the rate of DNA synthesis in HeLa cells in S phase is 3.6kb/min (Collins 1978; Painter and Schafer, 1969), which is even quicker than the 'fast' cell line 471 in the study by Courbet *et al* (Courbet *et al*., 2008). Considering that HeLa cells have been found to maintain a low copy number of pEPI-eGFP (Schaarschmidt *et al*., 2004) with such a fast doubling time, and that only very low levels of transgene expression were observed by the final day of the experiment in this study, and that HepG2 cells transfected with pEPI-eGFP were shown to have a large amount of transgene expression with no evidence of vector integration by the final day, it is possible that the longer cell cycle of the HepG2 cells indicates a slower rate of fork progression, which in turn may have played a role in the attachment of the S/MAR vector to the nuclear matrix. This may have led to a greater amount of pEPI-eGFP anchoring to the matrix, and thus a greater incidence of replication initiation from the S/MAR element within this vector.

3.3.15 Fluorescent In-Situ Hybridisation (FISH) Method To Show Vector Status In Transfected Populations Of Three Cell Lines

Several methods to determine the episomal/integrant status of the transfected vectors at the final day of the experiments were attempted before choosing fluorescent in-situ hybridisation (FISH), such as Southern blotting and plasmid rescue. However, it has these methods were insufficiently sensitive for the purposes of this thesis, as was shown experimentally and by the hypothetical calculation, where it was found that the sensitivity of the Southern blotting would have to be in the picograms, and that of the plasmid rescue in femtograms, in order for these techniques to work in this setting. Considering the approach taken in conducting the experiments in this chapter, where populations were transfected with vector and then selected, without the picking of clones, a greater difficulty in vector status determination would ensue. The study by Papapetrou *et al* (described in Chapter 1: Introduction, Section 1.10) demonstrated that the detection of plasmid sequences transfected into heterogenous uncloned

populations, where no single clones where subsequently isolated, by Southern blotting or plasmid rescue in order to verify the status of the vector can be ineffective when the percentage of positive cells is very low, and a more sensitive method would be required, as was the case in this thesis investigation (Papapetrou *et al*., 2006). Papapetrou *et al* therefore attempted to verify the episomal state of the vector by amplifying the entire vector sequence by using closely apposed primers to yield a band of the vector's size if present in a circular form (Papapetrou *et al*., 2006). However, they found that the efficiency of PCR amplification decreased when a vector the size of pEPI-eGFP (6.7kb) was attempted to be amplified.

FISH is an extremely sensitive method used to locate a specific sequence within genomic DNA and can also be used to identify plasmid DNA that has been transfected into cells, thus indicating episomal/integrant status. A large number of cells can be screened and the sequence being probed can be visualised under a microscope. As this has been used to successfully distinguish between episomes and integrants in transfected multiclonal populations in studies of S/MAR vector (Stehle *et al*., 2007; Baiker *et al*., 2000), FISH was chosen as the most suitable and appropriate method to utilise for the purposes of the experiments conducted in this study.

Although FISH effectively managed to demonstrate vector status in the samples tested in this study, a disadvantage was that, considering multiclonal populations of cells were being screened, even where no integration was detected in a sample and the vector was found in an episomal state, the possibility of low levels of integration could not be totally excluded. Another disadvantage is that even if fluorescence spots indicating the presence of the vector sequence were observed at similar positions on both arms of a given chromosome, the possibility that that these spots were episomal copies present at these regions due to the plasmids preferentially binding to certain sequences present on both arms of each chromatid, as opposed to conclusively being there due to the integration of the vector sequences, cannot be excluded. This method is also limited in that plasmid copy number cannot be quantified, and thus only qualitative observations can be made on the amounts of copies present per cell. In order to quantify plasmid copy number, a method such as Q-PCR would be required to be used.

Unfortunately, only a small number of positive spreads could be analysed by this method in these experiments, owing to the fact that very few positive spreads were

observed per slide for each population. In order to have been able to make more definitive conclusions on the integration status of each vector in each transfected population, a minimum of 20 positive spreads would have been required to be screened in order to be able to have a confidence level of 95%. Other sensitive methods that could be utilized to detect integration in multiclonal populations that could have been applied in this study include Linear Amplification-Mediated (LAM) PCR, which involves not only the detection of integrants, but can also identify the location of integration within the host genome. Another such method is pyrosequencing, which allows the sequencing of DNA and hence the detection of integrants and their positions within the genome.

It was observed that the HeLa and C2C12 cells that were analysed by FISH 24 hours post-transfection contained large areas where the probe had hybridised to large amounts of clustered plasmid vector. However, it cannot be determined by FISH alone whether these large clusters were plasmid that had already entered the nucleus, whether these clusters represented plasmid that had been present within the cells but had not yet entered the nucleus, or a combination of both. Nonetheless, it is clear that large amounts of vector had been taken up by these cells 24 hours after transfection with lipofection.

In evaluating the quality of the FISH images and metaphase spreads obtained, it is clear that in some cases the spreading of the chromosomes was not optimal. This could have been improved by the exposure of the slides to hot water vapour before adding the cells onto the slides, then exposing them again to the vapour followed by drying them on a hot surface (Henegariu *et al*., 2001). The addition of a few drops of acetic acid to the slides after adding the cells could also have improved chromosome spreading (Henegariu *et al*., 2001).

Some chromosomes also appeared ‘puffy’ or distorted in architecture. Large changes in the thermal temperature of slides can lead to the distortion of chromosomes and their architecture (Henegariu *et al*., 2001). Furthermore, the use of a thermocycler heating block could have improved the quality of the chromosomes (Henegariu *et al*., 2001) by adding the solutions directly to the slides, such as the denaturant solution when denaturing the chromosomes, and the pre-hybridisation buffer when applying probe to the slides, and then gradually increasing the temperature of the slides. Unfortunately, a thermocycler heating block was not available and, although care was

taken to gradually increase/decrease the temperatures of the slides during the experiment using water baths at different temperatures (as described in Chapter 2: Materials & Methods), this may have contributed to the distortion observed in several of the spreads.

3.4. Concluding Remarks

The findings suggesting the integration of eGFP-C1 in the HeLa and C2C12 cell lines, and the lack of evidence of integration of pEPI-eGFP in the HeLa, C2C12, and HepG2 cell lines are in line with what was observed in several studies (as reviewed in Chapter 1: Introduction, Secction 1.10), including the study by Baiker *et al,* where FISH was conducted on CHO cells transfected with eGFP-C1 or pEPI-eGFP vectors, showing that, after 100 cell generations, eGFP-C1 had integrated into variable sites in the host genome, whereas pEPI-eGFP had not (Baiker *et al*., 2000). HeLa cells were used in this study as a control cell line, and the evidence towards integration of eGFP-C1 and against that of pEPI-eGFP in HeLa cells is also as was revealed in a study by Schaarschmidt *et al* where the episomal status of the same vectors were investigated in HeLa cells (Schaarschmidt *et al*., 2004).

Although copies of both eGFP-C1 and pEPI-M18 were present as episomes in the cells on the final day of the experiment, as were copies of pEPI-eGFP, they were also shown that they had potentially integrated into the host genome, therefore making them potentially unsafe candidates for use as vectors in gene therapy. However, if pEPI-eGFP were to be a potential candidate for use in muscle cells for gene therapy for Muscular Dystrophy, then transgene expression must essentially be improved, as the vector would be required to express a significant enough amount of protein in order to have an effect on improving the diseased condition.

One potential method of improving expression would be to change the promoter driving the transgene. Furthermore, the sequence of the expression cassette in a vector can have an effect on different characteristics of the vector besides the rate of expression, such as DNA conformation, vector stability, and ease of base unpairing and DNA unwinding, which would affect nuclear matrix binding, replication, and transcription (Papapetrou *et al*., 2006). Using a different promoter could alter these properties. Therefore, it was essential to test another promoter which may have enabled a more open conformation at the promoter region of the eGFP transgene, which in itself could significantly improve the expression of pEPI-eGFP. Another promoter may also avoid silencing due to methylation, which would also result in longer-term transgene expression. Thus, a different promoter was tested in C2C12 cells in Chapter 5, after the generation of the vectors, described in Chapter 4.

Chapter 4: The Generation of Vectors CeGFP-C1, CpEPI-eGFP, CMYC-pEPI, and the Ideal Minicircle S/MAR Vector

4.1 Introduction:

In addition to the vectors that were purchased or kindly donated to be studied for the purposes of this thesis, several expression vectors were also required to be constructed in order to be tested for episomal retention and transgene expression in the C2C12 murine skeletal muscle cell line. These newly constructed vectors included CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI. For vectors CeGFP-C1 and CpEPI-eGFP, the CAGG promoter was inserted into the vectors and replaced the CMV promoter in eGFP-C1 and pEPI-eGFP. For vector CMYC-pEPI, the murine *c-myc* 1.7kb S/MAR element was inserted 3' to the eGFP transgene in CeGFP-C1. Several other vectors were also required to be constructed in order to enable the construction of these vectors (Table 4.1).

An ideal vector was also designed and attempted to be constructed. This vector was an S/MAR-containing minicircle. Chen et al developed a new plasmid vector, devoid of bacterial backbone, based upon the *Streptomyces* temperate phage integrase ØC31 – mediated site-specific intramolecular recombination technology (Chen et al., 2003). This construct, designed by Chen et al, was used in this thesis in order to attempt to produce minicircles. The parent plasmid undergoes recombination at two specifically inserted sites, *att*B and *att*P, resulting in a minicircle containing the transcription unit only, and a miniplasmid (maxicircle) containing the bacterial backbone. The principle is based on the destruction of the unrecombined parent plasmid, and miniplasmid after recombination has taken place, by the insertion of the endonuclease I-SceI gene and I-SceI recognition sequences outside the minicircle unit in order to linearize the maxicircle and unrecombined parental plasmids, which would eventually be digested by exonucleases within the cell, leaving the minicircle intact (Figure 4.1).

The limitation that exists to using minicircles as vectors is that they are only well sustained in non-dividing cells and may not be passed on to daughter cells over many generations. In order to overcome this problem and optimize minicircle performance,

maintenance in cells, and transgene expression, an S/MAR element can be inserted into minicircle. In a study by Nehlsen et al using a 4kb minicircle made by Flp recombinase technology, containing only a transcription unit and an S/MAR element, it was found that the vector was established without selection pressure and was stably propagated over many cell generations without integration (Nehlsen et al., 2006). The study showed that S/MAR-containing minicircles are potentially very powerful and effective tools if used as gene delivery vectors, as they appeared to overcome many of the obstacles that currently limit the field of gene therapeutics, and can even be considered to be 'ideal vectors'. Therefore, S/MAR-containing minicircles were attempted to be constructed to be tested in C2C12 cells in this thesis, containing only an eGFP transgene expression cassette, driven by the CMV promoter, and the human β -IFN S/MAR element.

Vector Name	Transgene Promoter	Transgene	S/MAR	Length (kb)
eGFP-C1	CMV	eGFP	none	4.7
CeGFP-C1	CAGG	eGFP	none	5.0
pEPI-eGFP	CMV	eGFP	B-IFN S/MAR	6.7
CpEPI-eGFP	CAGG	eGFP	B-IFN S/MAR	7.0
eGFP-N3	CMV	eGFP	none	4.7
CeGFP-N3	CAGG	eGFP	None	5.0
pEPI-N3	CMV	eGFP	B-IFN S/MAR	6.7
CMYC-pEPI	CAGG	eGFP	c-MYC S/MAR	6.7
p2θC31	empty	empty	none	9.7
p2θC31.LacZ	CMV	LacZ	none	14.0
p2θC31.eGFP	CMV	eGFP	none	11.3
p2θC31.eGFP.S/MAR	CMV	eGFP	B-IFN S/MAR	13.3

Table 4.1 Vectors used and created in this chapter, indicating name, the transgene promoter, the transgene used for assessment studies, the presence of an S/MAR element, and the length (kb) of each construct.

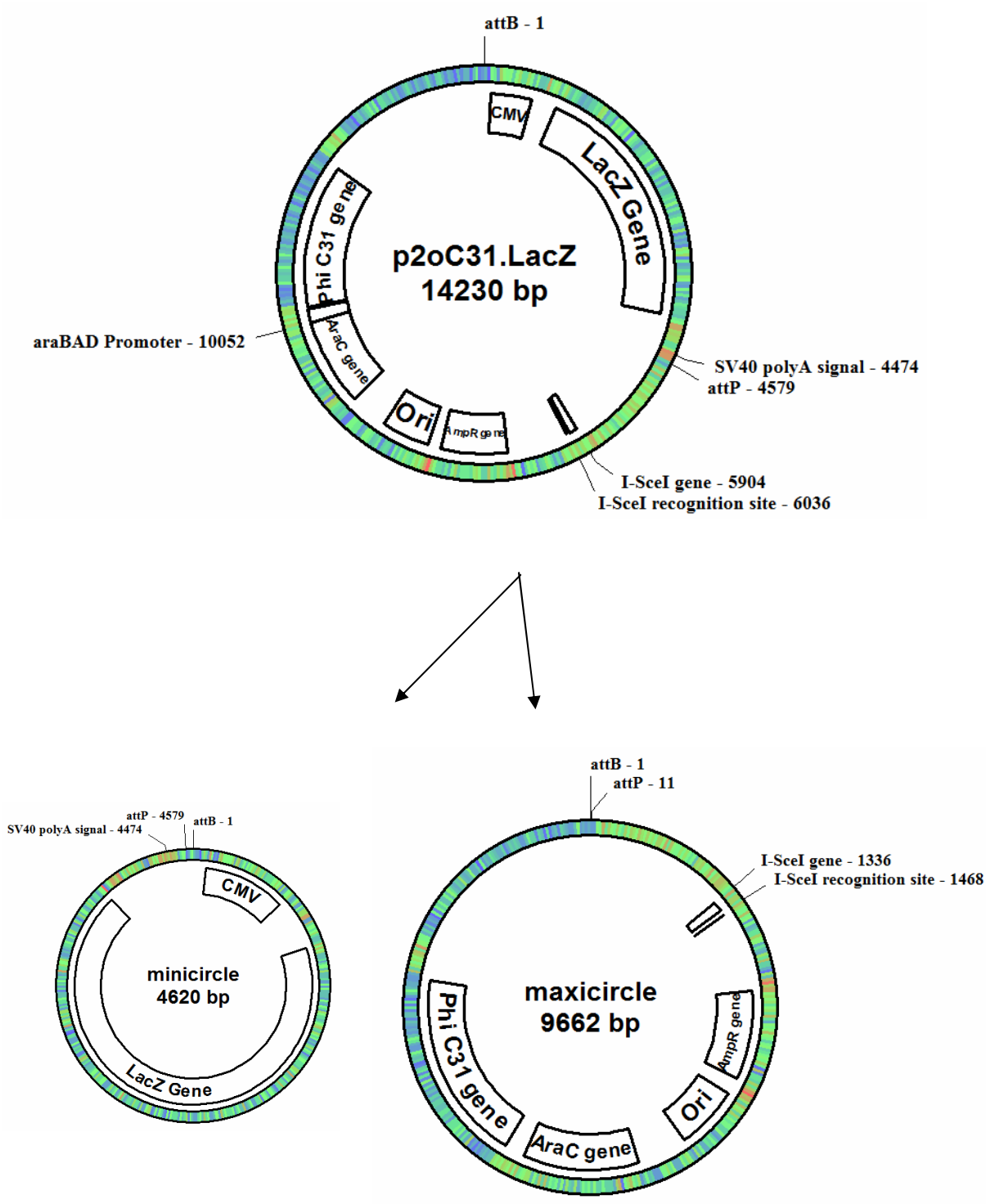


Figure 4.1 The 14kb minicircle parental plasmid p2ØC31.LacZ expressing the LacZ transgene is shown above. The BAD promoter induced by L-arabinose, added into the LB media, initiates the transcription of the C31 phage recombinase, which leads to the formation of the 4.6kb minicircle (lower left vector) containing the transcription unit, and the 9.7kb maxicircle (lower right vector) containing the backbone including all the bacterial sequences, which is then linearized by I-SceI, and eventually degraded by bacterial exonucleases.

4.2 Results:

4.2.1 The insertion of the CAGG promoter into eGFP-C1 and pEPI-eGFP by molecular cloning

The CAGG promoter was excised from CeGFP-N3 in order to replace the CMV promoter in vectors eGFP-C1 and pEPI-eGFP. This was conducted by the digestion CeGFP-N3 with *SalI,* blunting with T4 DNA Polymerase (Materials & Methods), then digesting with *ApaLI* to yield the promoter, which was a fragment size of 1288bp, and the 3718bp plasmid backbone fragment (Figures 4.2 and 4.3).

eGFP-C1 was digested with *AfeI*, blunted, then digested with *ApaLI* to yield the 3763bp backbone fragment into which the CAGG promoter was to be inserted, and the 968bp CMV promoter fragment to be discarded (Figures 4.2 and 4.3). pEPI-eGFP was digested as eGFP-C1, yielding the 5726bp backbone into which the CAGG promoter was to be inserted, and the 968bp CMV promoter fragment to be discarded (Figures 4.2 and 4.4).

The 1288bp CAGG promoter was ligated into the 3763bp eGFP-C1 plasmid backbone, and in a separate reaction with the 5726bp pEPI-eGFP plasmid backbone, as sticky-blunt directional ligations (see Methods and Materials). The ligated vectors were transformed by electroporation into Top10 electrocompetent *E.coli* which were then plated onto Kanamycin agar plates. Colonies were picked from the selective agar plates, grown up as mini-preps, then plasmid was extracted and digested in order to identify the clones which contained the vectors CAGG-eGFP-C1 and CAGG-pEPI-eGFP (termed CeGFP-C1 and CpEPI-eGFP).

Plasmids extracted from the mini-preps were digested with *AseI* and *EcoRI*. Expected fragment sizes for CeGFP-C1 were 1468bp and 3552bp, and 1481bp, 1951bp, and 3551bp for CpEPI-eGFP if the cloning had been successful (Figures 4.5 and 4.6). The successful clones were identified, grown as maxi-preps, and plasmid was extracted in preparation for transfection. Further digests were conducted after growing the clones as maxi-preps in order to re-confirm the plasmid sequences. They were digested with *Eco010901* and *MluI*, with expected fragment sizes for CeGFP-C1 of 2212bp and 2808bp, and for CpEPI-eGFP of 2211bp and 4772bp (Figures 4.5 and 4.7).

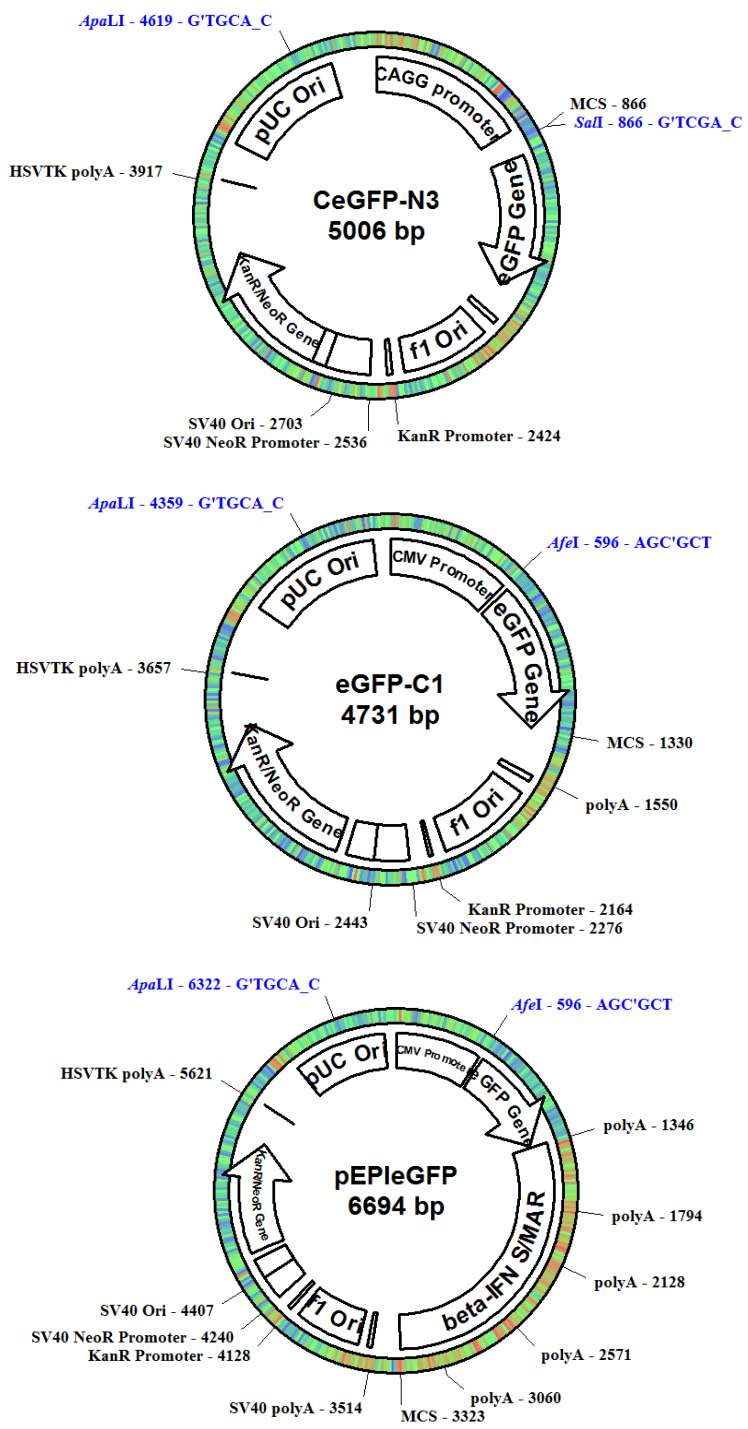


Figure 4.2 Schematic diagram of plasmid maps used to clone CeGFP-C1 and CpEPI-eGFP, indicating restriction sites utilised. Plasmid CeGFP-N3 was digested with *SalI* and *ApaLI*. Plasmids CeGFP-C1 and CpEPI-eGFP were digested with *AfeI* and *ApaLI*. The CAGG promoter insert was then ligated into the CeGFP-C1 and CpEPI-eGFP backbones by a sticky-blunt directional ligation.

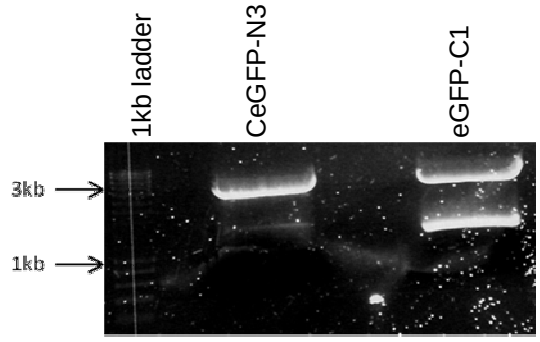


Figure 4.3 A 0.8% agarose gel of the plasmid CeGFP-N3 (middle lane) digested by *SalI*, blunted, then digested by *ApaLI* to yield the CAGG promoter insert fragment of 1288bp, then excised for ligation, and the 3718bp backbone fragment. eGFP-C1 (right lane) was digested by *AfeI*, blunted, then digested with *ApaLI* to yield the 968bp CMV promoter fragment and the 3763bp backbone, excised for ligation. The left lane contained a 1kb ladder.

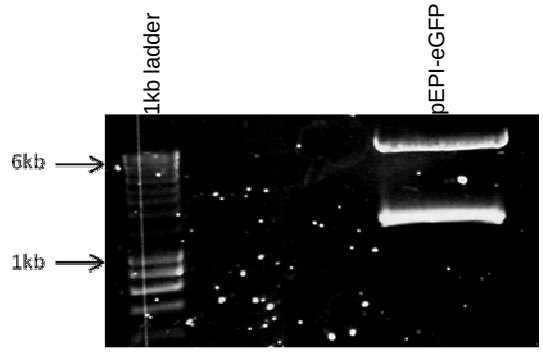


Figure 4.4 A 0.8% agarose gel of pEPIeGFP (right lane) digested by *AfeI*, blunted, then digested with *ApaLI* to yield a 968bp CMV promoter fragment and the 5726bp backbone, excised for ligation. The left lane contained a 1kb ladder.

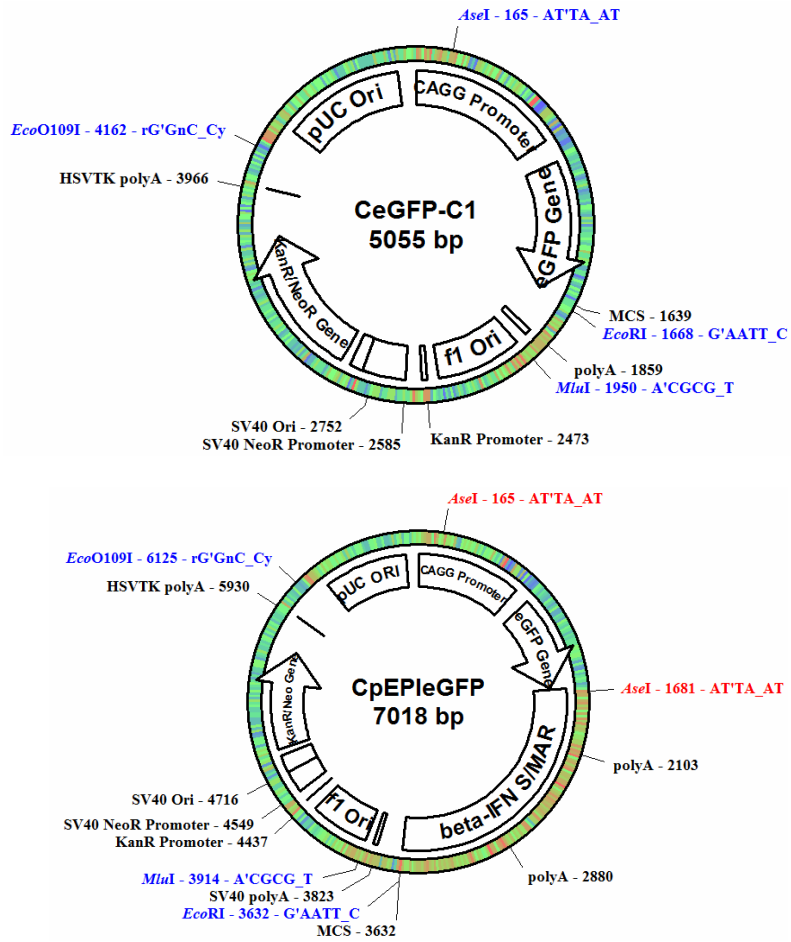


Figure 4.5 The maps of constructs CeGFP-C1 and CpEPI-eGFP indicating restriction enzyme sites *AseI, MluI, EcoRI, and Eco01091* which were used to confirm the successful cloning of these vectors.

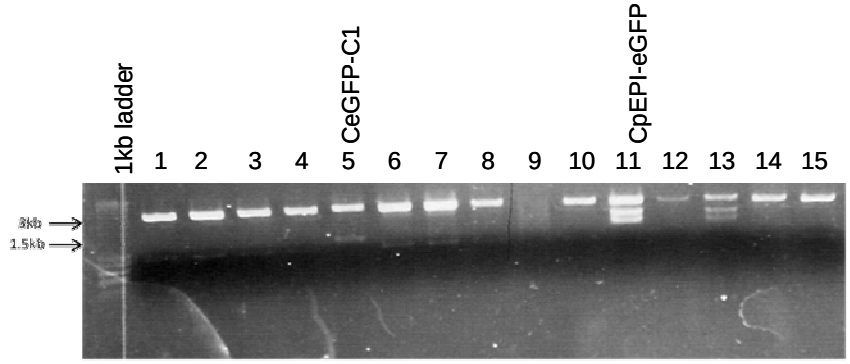


Figure 4.6 A 0.8% agarose gel of mini-prep clones digested by *AseI* and *EcoRI* to identify the successfully cloned CeGFP-C1 and CpEPI-eGFP. Clones 1-8 were potential CeGFP-C1 vectors, and the expected sizes of the fragments were 1468bp and 3552bp if successful. Clone #5 was chosen as the successfully created CeGFP-C1 vector. Clones 9-15 were potential CpEPI-eGFP vector, and the expected sizes of the fragments were 1481bp, 1951bp, and 3551bp if successful. Clone #11 was chosen as the successfully created CpEPI-eGFP vector. The left lane contained a 1kb ladder.

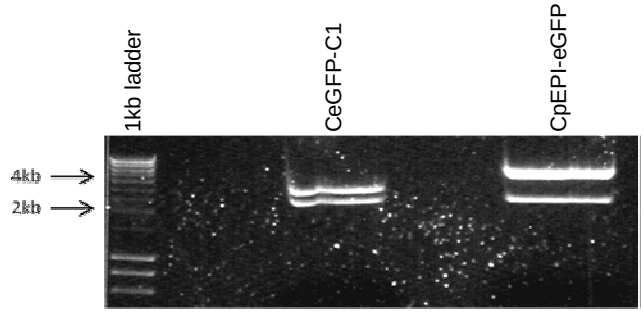


Figure 4.7 A 0.8% agarose gel of CeGFP-C1 (middle lane) and CpEPI-eGFP (right lane) plasmids digested with *Eco010901* and *MluI* to re-confirm the vector sequences. Expected fragment sizes for CeGFP-C1 were 2212bp and 2808bp, and for CpEPI-eGFP were 2211bp and 4772bp, which was confirmed by the gel. The left lane contained a 1kb ladder.

4.2.2 The insertion of the *c-myc* S/MAR element into CeGFP-C1 by PCR and molecular cloning

In order to insert the *c-myc* S/MAR element into CeGFP-C1, the sequence was first identified in the S/MART database (http://smartdb.bioinf.med.uni-goettingen.de/). The BAC clone containing the sequence was obtained from the Wellcome Trust Sanger Institute, and primers were designed to amplify the sequence by PCR (Materials & Methods). The primers contained overhangs of restriction sites so that the PCR product would be flanked by a *XhoI* restriction site at its 5' end and an *EcoRI* site at its 3'end. CeGFP-C1 also contained *EcoRI* and *XhoI* sites, both located at its MCS site downstream of the eGFP gene (Figure 4.8). Both the product and the plasmid were digested with *EcoRI* and *XhoI,* and the c-myc S/MAR was then inserted by a directional sticky-sticky ligation into the CeGFP-C1 backbone to create CMYC-pEPI (Figure 4.9).

The correct CMYC-pEPI clone was identified by digesting 8 minipreps with *AseI* and *BamHI,* where the correct clone yielded three fragments of sizes 1067bp, 2145bp, and 3521bp (Figure 4.10).

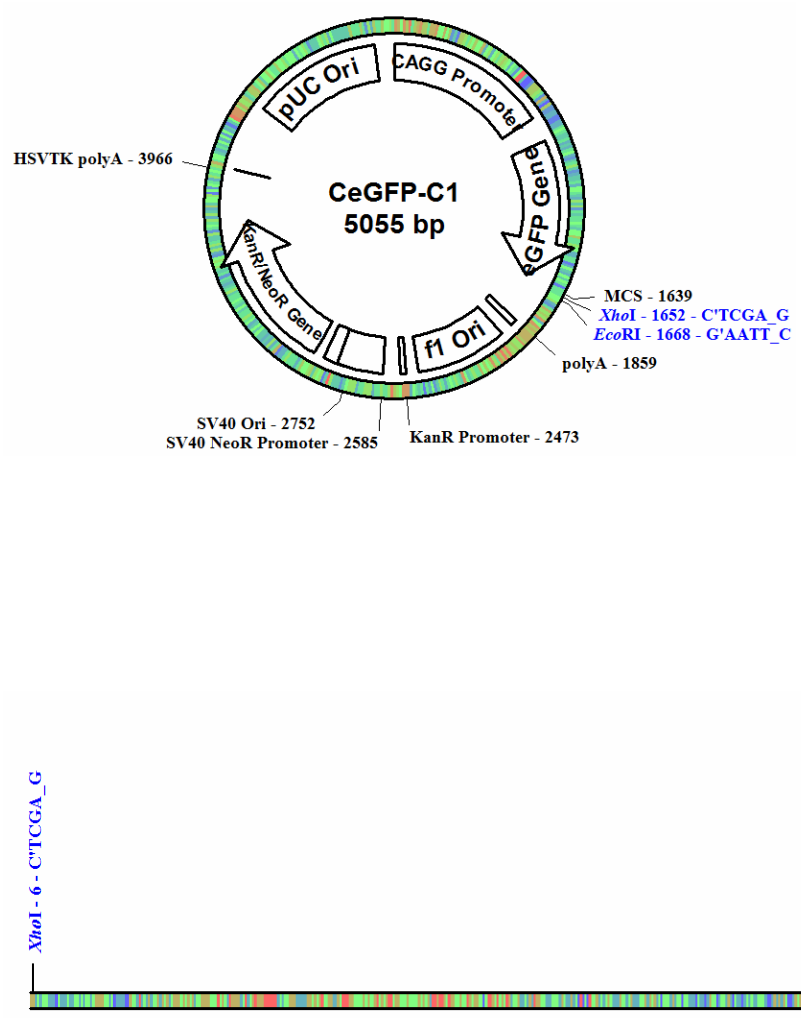


Figure 4.8 Schematic diagram of plasmid map of CeGFP-C1 (top image) and the c-myc S/MAR insert fragment (lower image), indicating restriction sites utilised to construct CMYC-pEPI. The c-myc S/MAR element was ligated into the CeGFP-C1 backbone by a sticky-sticky directional ligation.

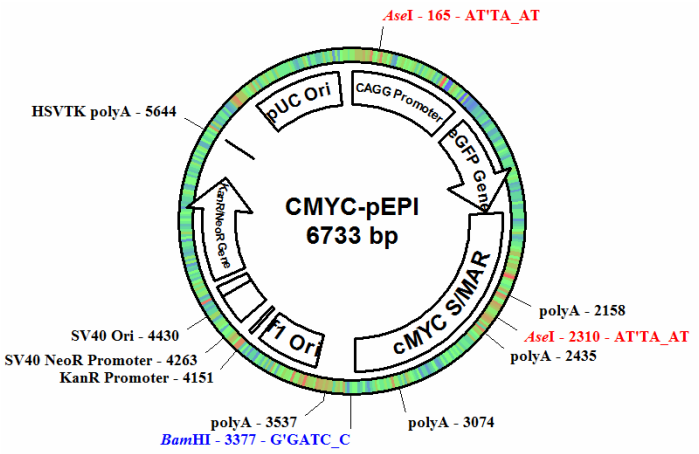


Figure 4.9 The map of constructs CMYC-pEPI indicating restriction enzyme sites *AseI* and *BamHI,* which were used to confirm the successful cloning of this vector.

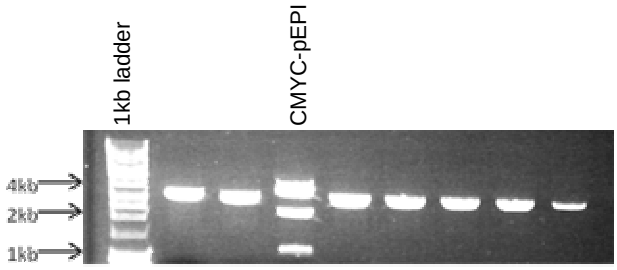


Figure 4.10 A 0.8% agarose gel of 8 minipreps for the potential CMYC-pEPI vector, digested with *AseI* and *BamHI*. Fragment sizes expected for the correct clone were 1067bp, 2145bp, and 3521bp. The correct clone was chosen as indicated. The left lane contained a 1kb ladder.

4.2.3 The generation of eGFP-expressing minicircles/ S/MAR minicirlces

The parental plasmid p2θC31, which is the plasmid used to generate minicircles, contained the lacZ transgene (p2θC31.lacZ) driven by the CMV promoter when it was received. Cloning strategies were designed in order to insert the eGFP transgene into the vector, replacing the lacZ transgene. This construct (p2θC31.eGFP) could then be used to generate eGFP expressing minicircles. An insert containing the eGFP transgene followed by the β-IFN S/MAR element was also planned to be cloned into the p2θC31 parental plasmid (p2θC31.eGFP.S/MAR) in order to generate eGFP expressing S/MAR-containing minicircles. These constructs were to be transfected into C2C12 cells to assess expression and episomal status, and the S/MAR-containing minicircles were to be tested to determine their ability to replicate as episomes within dividing cells.

Construct 1: p2θC31.eGFP

The cloning of p2θC31.eGFP involved digesting eGFP-N3 with *AflII*, blunting with T4 DNA polymerase, then digesting with *XhoI*, generating a 3.7 kb backbone to be discarded, and a 1kb insert which included the eGFP transgene and the SV40 polyA sequence. p2θC31.LacZ was digested with *ApaI*, blunted, then digested with *XhoI*, yielding a 10.3kb backbone, and a 3.9kb fragment containing the lacZ transgene and the SV40 polyA signal, which was to be discarded (Figures 4.11 and 4.12). The 1kb insert from eGFP-N3 was ligated by a sticky-blunt directional ligation into the p2θC31.LacZ backbone, creating the construct p2θC31.eGFP (Figure 4.13). The correct ligation of construct p2θC31.eGFP was confirmed by obtaining the correct restriction pattern when digesting separately with BcgI and with BsrGI (Figure 4.14). The digest with *BcgI* of the correct construct should yield 5.6kb, 3.4kb, 1.9kb, and 0.4kb fragments, and with *BsrGI* should yield 9.8kb and 1.5kb sized fragments. All these fragments were observed when the construct was digested, with the exception of the 0.4kb fragment which was too small and faint to appear on the gel. Another gel with a greater amount of BcgI-digested p2θC31.eGFP DNA could have been run on a gel containing a greater percentage of agarose, such as 1.2%, in order to enable visualisation of this band.

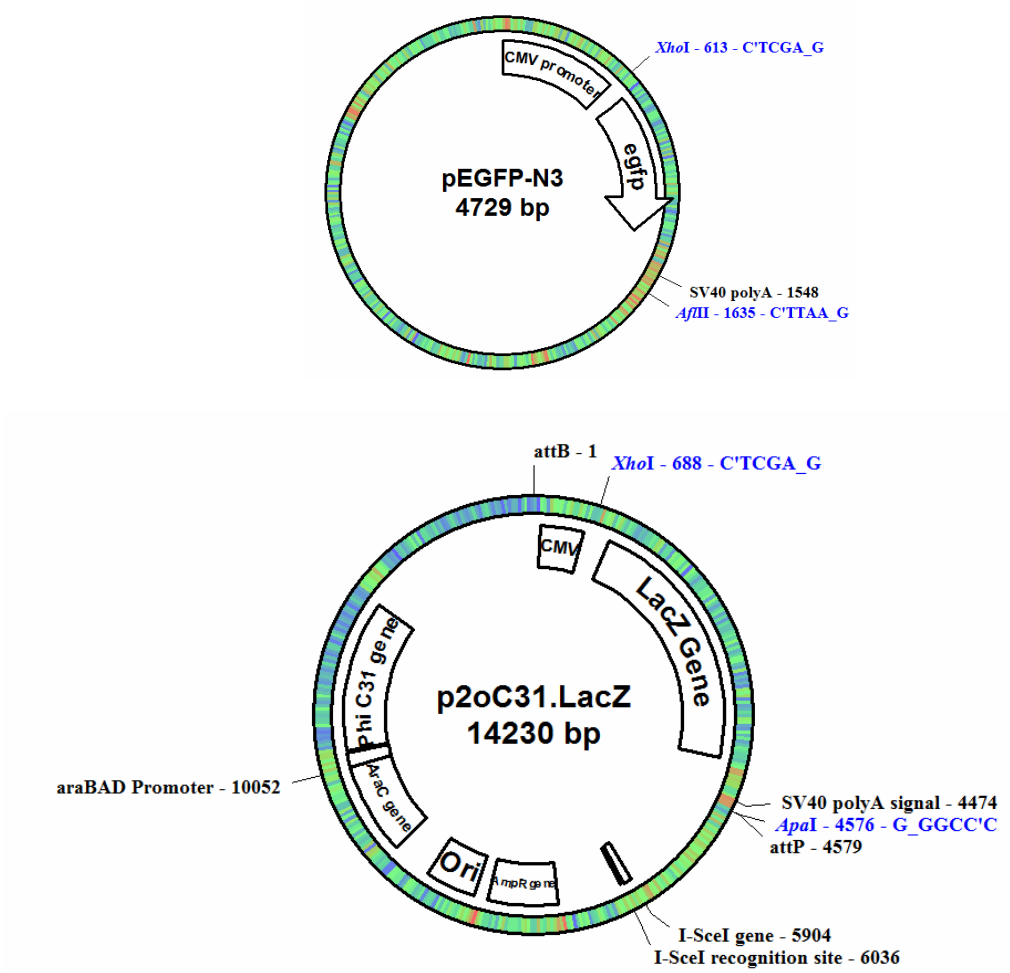


Figure 4.11 Schematic diagram of plasmid maps used to clone p2θC31.eGFP, indicating restriction sites utilised. Plasmid eGFP-N3 was digested with *AflII*, blunted, then digested with *XhoI*. Plasmid p2θC31.LacZ was digested with ApaI, blunted, then digested with *XhoI*. The insert was then ligated into the p2θC31.LacZ backbone by a sticky-blunt directional ligation.

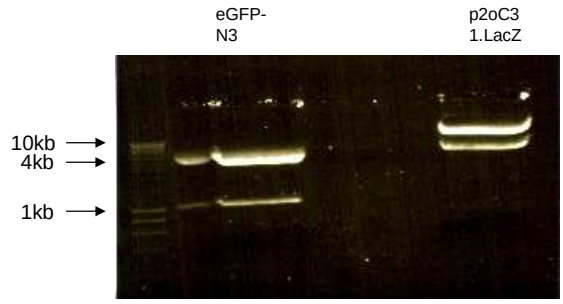


Figure 4.12 A 0.8% agarose gel indicating the fragments to be cloned to create vector p2θC31.eGFP. eGFP-N3 was digested with AflII, blunted, then digested with XhoI yielding a 3.7 kb backbone to be discarded, and a 1kb insert which included the eGFP transgene and the SV40 polyA sequence. p2θC31.LacZ was digested with ApaI, blunted, then digested with XhoI, yielding a 10.3kb backbone, and a 3.9kb fragment containing the lacZ transgene and the SV40 polyA signal, which was to be discarded. The left lane contained a 1kb ladder.

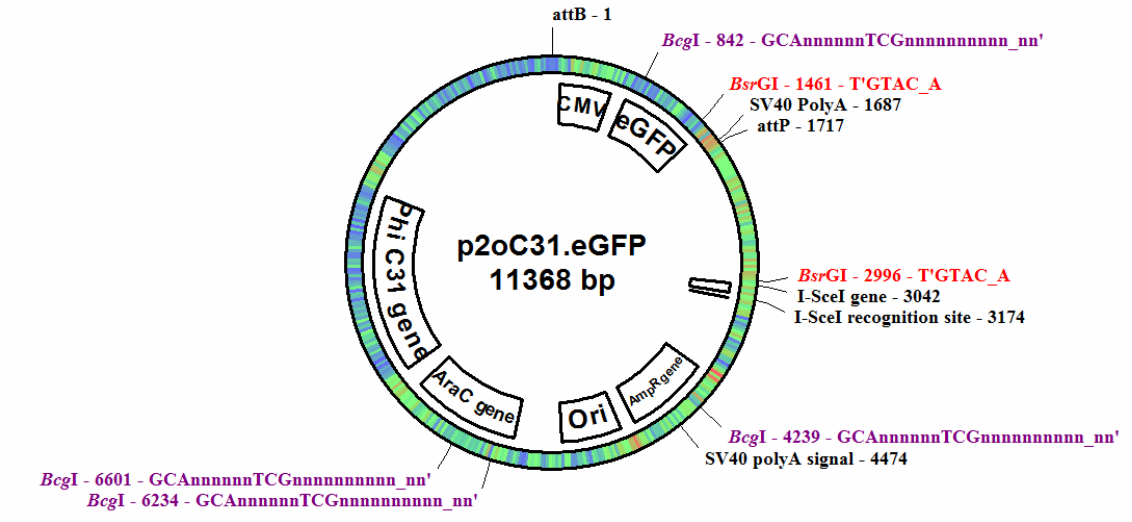


Figure 4.13 The map of construct p2θC31.eGFP indicating restriction enzyme sites *BcgI* and *BsrGI* which were used to confirm the cloning of the eGFP transgene and SV40 polyA into the p2θC31 backbone.

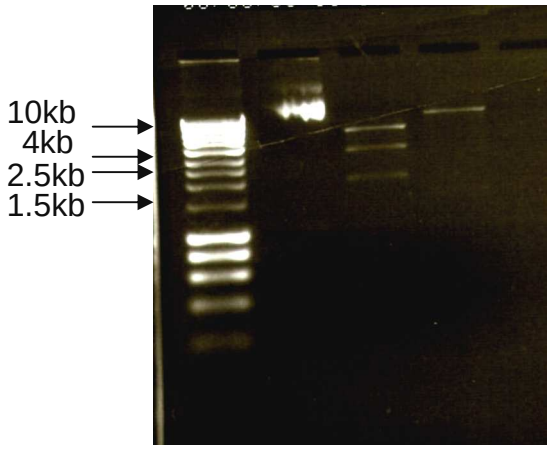


Figure 4.14 A 0.8% agarose gel confirming the correct cloning of p2θC31.eGFP. From left to right, lane 1: 1kb ladder; lane 2: undigested p2θC31.eGFP; lane 3: p2θC31.eGFP digested with *BcgI* showing fragments of sizes 5.6kb, 3.4kb, and 1.9kb; lane 4: p2θC31.eGFP digested with *BsrGI* showing fragments of sizes 9.8kb and 1.5kb.

Construct 2: p2θC31.eGFP.S/MAR

Unfortunately, although several cloning strategies were devised in order to construct p2θC31.eGFP.S/MAR, none had proven successful. Several of the strategies employed are be demonstrated here.

Strategy 1:

The plasmid pEPI-eGFP was digested fully and linearized with *MluI*, then partially digested with *AseI* (Figure 4.15) at different concentrations of *AseI* in order to find the optimal concentration of enzyme which would yield a maximum amount of the fragment required to be inserted into the p2θC31 backbone, containing the CMV promoter, eGFP transgene, S/MAR element, and SV40 polyA signal, which was 3.6kb. The fragment sizes yielded from this partial digest were 5.3, 4.4, 3.6, 3.0, 2.2, and 1.3kb (Figure 4.16). Once the 3.6kb fragment was excised from the gel it was blunted with T4 DNA polymerase. p2θC31 was linearized with *XhoI* and blunted to yield a backbone of 9.7kb into which the insert would be ligated in a blunt-blunt non-directional ligation. However, many attempts using this strategy did not lead to the successful of p2θC31.eGFP.S/MAR. An example gel (Figure 4.17) demonstrated how the backbone seemed to keep recircularising without the insert being ligated, as digestion of the correct clone by AseI would have yielded 9kb and 4.2kb fragments or 8.2kb and 5 kb fragments in the case of an inverted insert. Yet, only single fragments of approximately 9.7kb were shown. De-phosphorylation of the backbone was also attempted prior to ligation using shrimp alkaline photphatase, and the ratios of vector:insert for the ligation reaction were altered and varied (Materials& Methods), but still with no success.

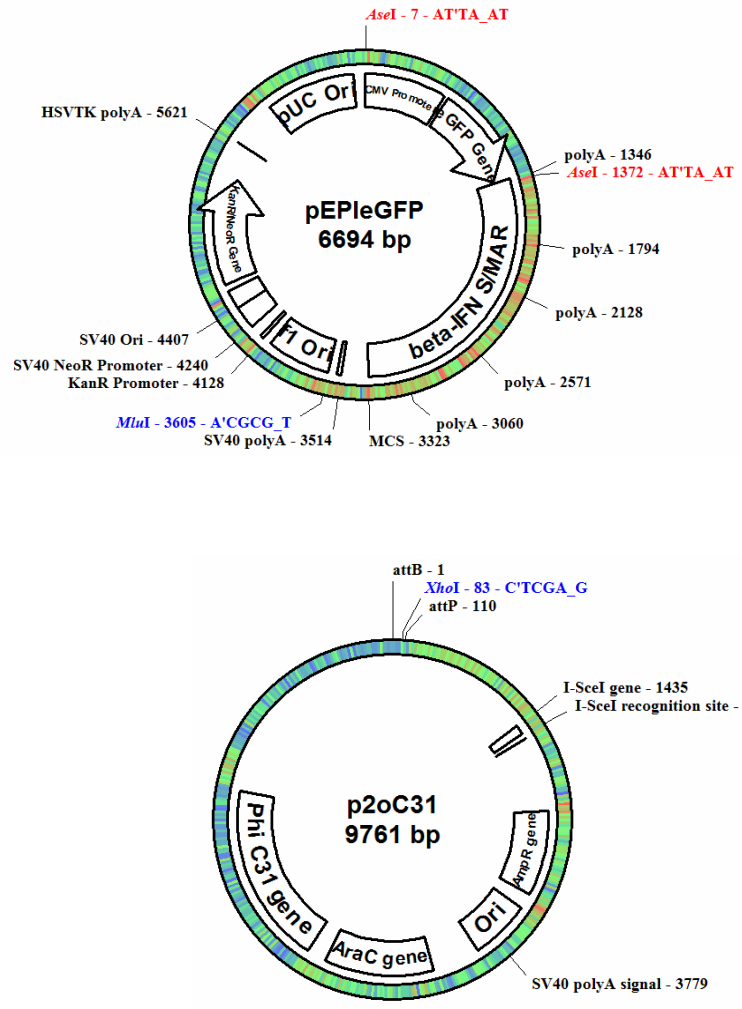


Figure 4.15 Schematic diagram of plasmid maps used to clone p2θC31.eGFP.S/MAR, indicating restriction sites utilised. Plasmid pEPI-eGFP was digested with *MluI* to completion, then partially digested with *AseI*. p2θC31 was linearised with *XhoI*.

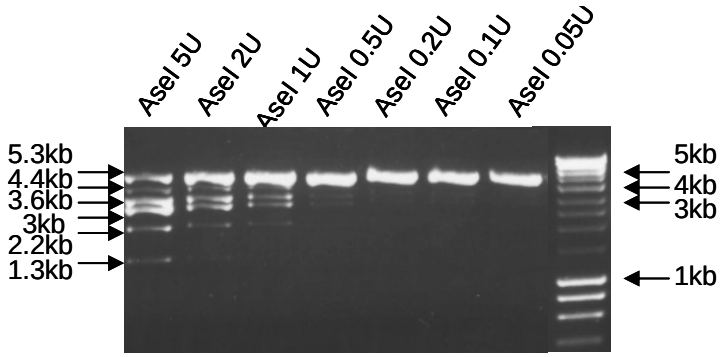


Figure 4.16 A 0.8% agarose gel of 12.5µg of pEPI-eGFP digested to completion with *MluI*, and then partially digested with variable amounts (0.05-5U) of *AseI* to determine the optimal units of enzyme required to be used in order to yield the greatest amount of the 3.6kb fragment to be used in cloning p2θC31.eGFP.S/MAR. The fragment sizes yielded from this partial digest were 5.3, 4.4, 3.6, 3.0, 2.2, and 1.3kb. As determined from this gel, the optimal amount of units was 5U. The right lane contained a 1kb ladder.

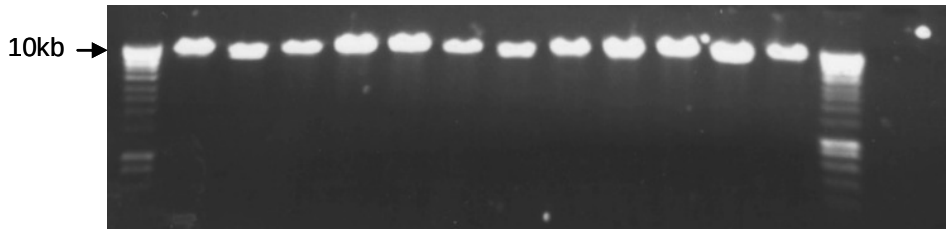


Figure 4.17 A 0.8% agarose gel of plasmid extracted from minipreps after cloning attempt of p2θC31.eGFP.S/MAR, digested with *AseI*. The large fragments observed in each lane indicate variable amounts of the 9.7kb p2θC31 which seems to have recircularized without the ligation of the insert, then linearised by digestion with AseI. A construct with the insert digested with *AseI* would have yielded 9kb and 4.2kb fragments or 8.2kb and 5 kb fragments in the case of an inverted insert. The first and last lanes contained a 1kb ladder.

Strategy 2:

Considering the cloning strategy employed to create p2θC31.eGFP had been successful, it was adapted in order to attempt to create p2θC31.eGFP.S/MAR. Considering that the β-IFN S/MAR-containing vector available in the lab was pEPI-eGFP, which was essentially the S/MAR element inserted into the eGFP-C1 plasmid, and in order to create p2θC31.eGFP the eGFP-N3 vector had been utilised due to the convenient location of the restriction enzymes sites which had made the directional cloning possible, the S/MAR element would first need to be inserted into the eGFP-N3 backbone. This was conducted by digesting pEPI-eGFP plasmid, that had been transformed into Dam negative *E.coli,* with *BsrGI* and *XbaI*, as *XbaI* is subject to Dam methylation. The resulting fragments were the 4.6kb backbone to be discarded, and the 2kb insert to be ligated into the eGFP-N3 backbone, which had also been transformed into Dam negative *E.coli* and digested with *BsrGI* and *XbaI* to yield a 4.7kb backbone fragment, and a 23bp fragment which was too short to detect on a 0.8% agarose gel (Figure 4.18).

Minipreps of the ligations were grown, and the successful clone, pEPI-N3, was isolated and digested with *AseI* to show that the ligation had been successful (Figure 4.19). Expected fragment sizes were 5.3 and 1.4kb from the *AseI* digest, indicating that the cloning had been a success (Figure 4.20).

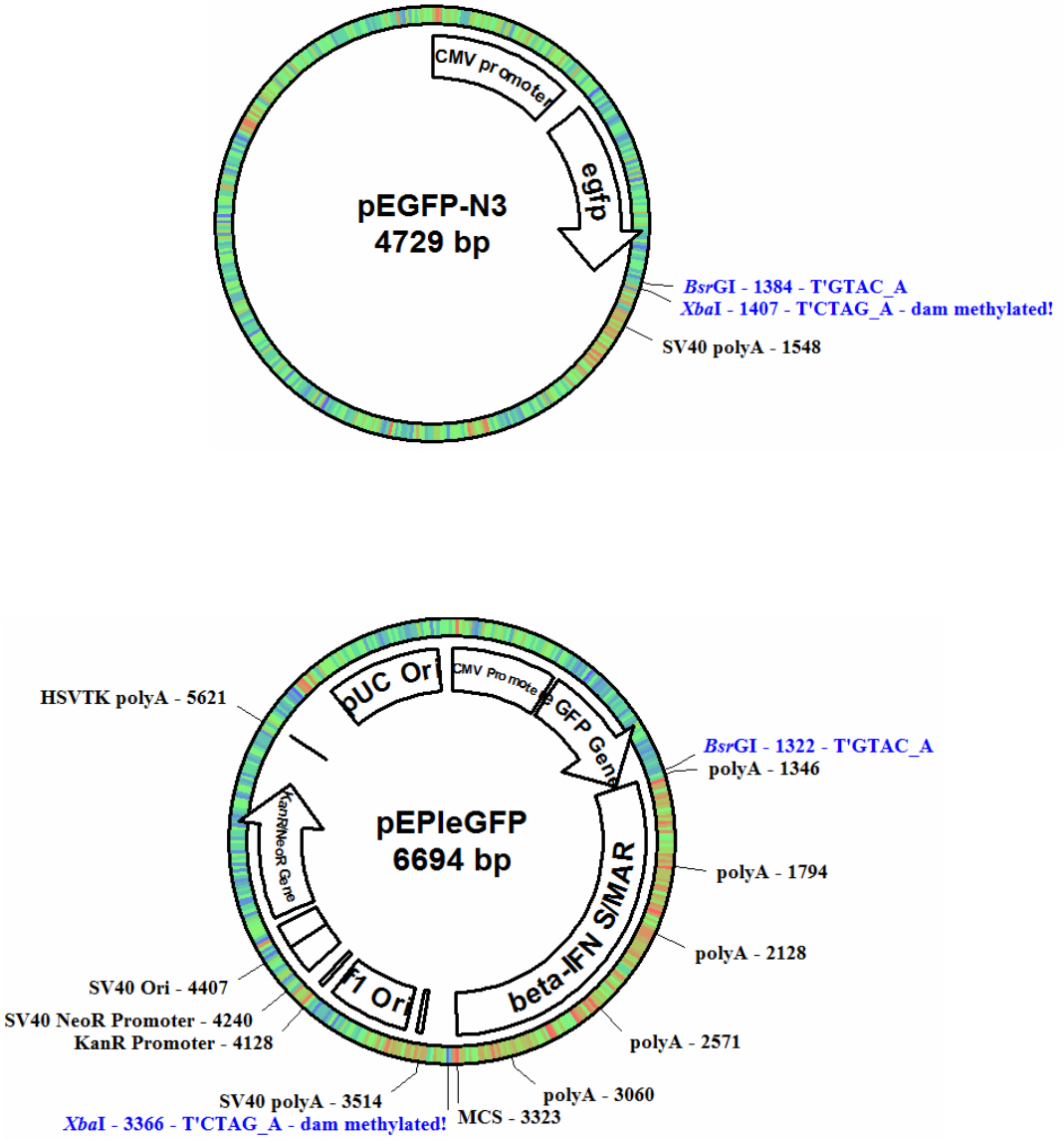


Figure 4.18 Schematic diagram of plasmid maps used to clone pEPI-N3, indicating restriction sites utilised. Both plasmids eGFP-N3 (top image) and pEPI-eGFP (lower image) were digested with *BsrGI* and *XbaI.*

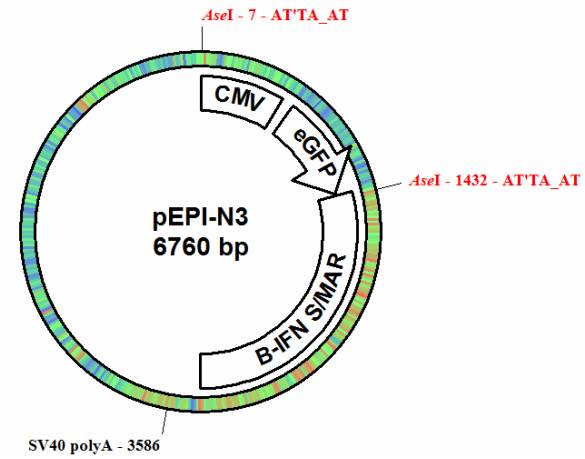


Figure 4.19 The map of construct pEPI-N3 indicating restriction enzyme sites for *AseI*, which was used to confirm the cloning of the S/MAR element into the eGFP-N3 backbone.

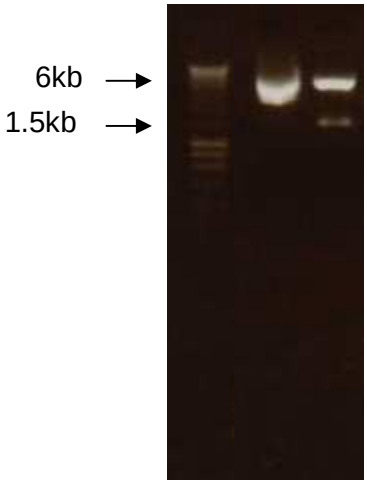


Figure 4.20 A 0.8% agarose gel confirming the successful cloning of pEPI-N3. From left to right, lane 1: 1kb ladder; lane 2: undigested pEPI-N3; lane 3: pEPI-N3 digested with *AseI* yielding 5.3 and 1.4kb fragments.

Next, the same strategy applied for the cloning of p2θC31.eGFP was applied. pEPI-N3 was digested with *AflII*, blunted with T4 DNA polymerase, then digested with *XhoI*, generating a 3.7kb backbone to be discarded, and a 3kb insert which included the eGFP transgene, the S/MAR element, and the SV40 polyA sequence. p2θC31.LacZ was digested with *ApaI*, blunted, then digested with *XhoI,* yielding a 10.3kb backbone, and a 3.9kb fragment containing the lacZ transgene and the SV40 polyA signal, which was to be discarded (Figures 4.21 and 4.22). The 3kb insert from pEPI-N3 was ligated by a sticky-blunt directional ligation into the p2θC31.LacZ backbone, in order to create the construct p2θC31.eGFP.S/MAR.

The correct ligation of construct p2θC31.eGFP.S/MAR was attempted to be confirmed by digesting with *AseI*. A correctly cloned p2θC31.eGFP.S/MAR vector digested with *AseI* should yield fragment sizes of 8 and 5kb (Figures 4.23 and 4.24). *BsrGI* was also used to attempt to confirm this cloning. A correctly cloned p2θC31.eGFP.S/MAR vector digested with *BsrGI* should yield fragment sizes of 9.8 and 3.5kb (Figures 4.24 and 4.25). Unfortunately, many attempts at cloning p2θC31.eGFP.S/MAR using this strategy were also unsuccessful, even when de-phosphorylation of the backbone was attempted prior to ligation using shrimp alkaline photphatase, and the ratios of vector:insert for ligation were altered and varied (Materials & Methods).

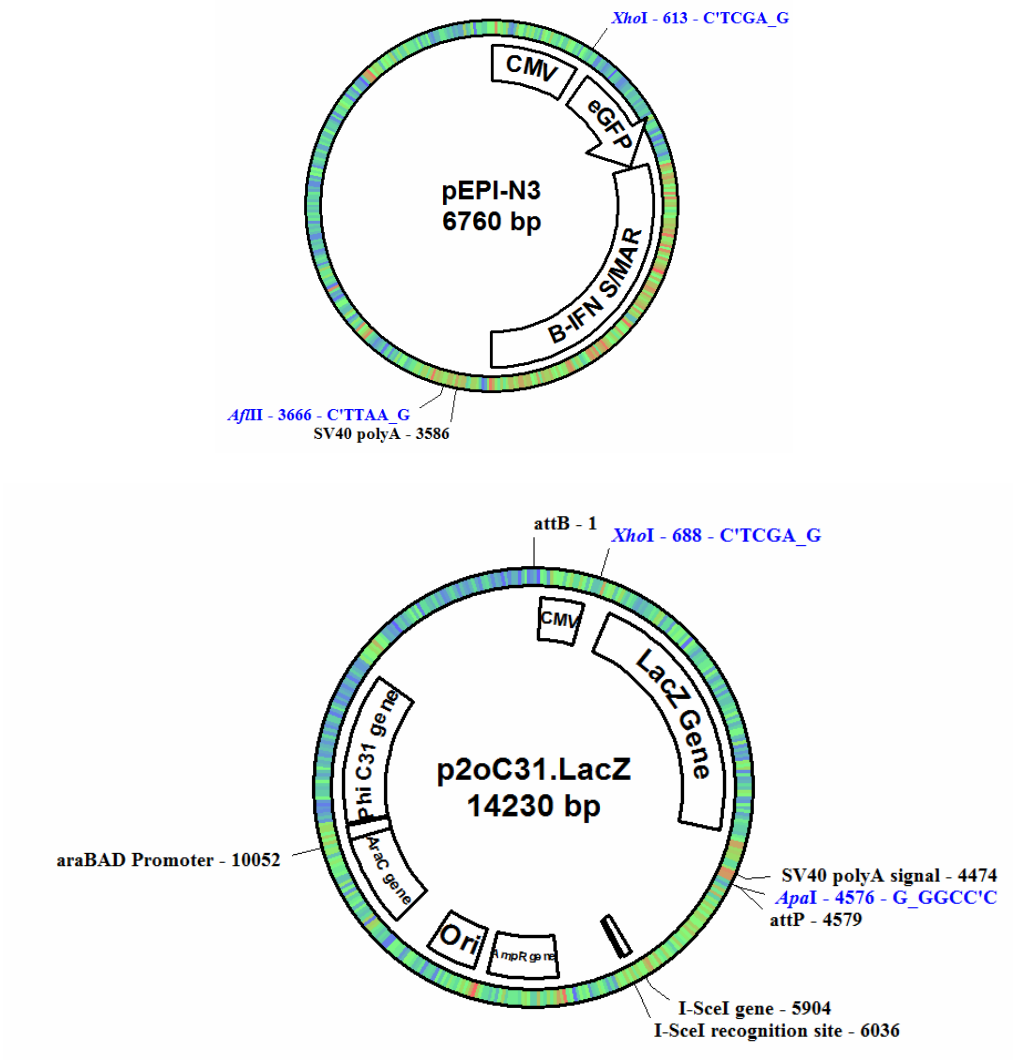


Figure 4.21 Schematic diagram of plasmid maps used to clone p2θC31.eGFP.S/MAR, indicating restriction sites utilised. Plasmid pEPI-N3 (above) was digested with *AflII*, blunted, then digested with *XhoI*. Plasmid p2θC31.LacZ (below) was digested with *ApaI*, blunted, then digested with *XhoI*. The insert was then attempted to be ligated into the p2θC31.LacZ backbone by a sticky-blunt directional ligation.

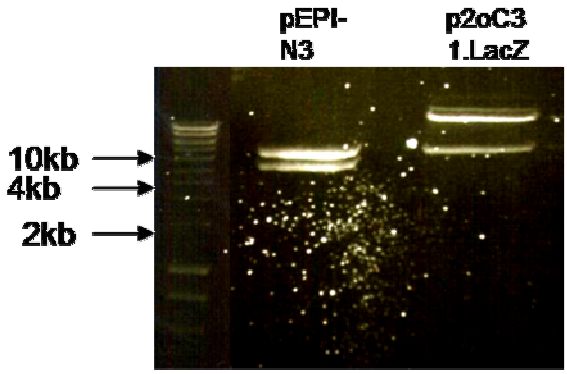


Figure 4.22 A 0.8% agarose gel indicating the fragments to be excised and cloned in order to create vector p2θC31.eGFP.S/MAR. From left to right: lane 1: 1kb ladder; lane 2: pEPI-N3 digested with *AflII* and *XhoI* yielding a 3kb insert to be excised from the gel, and a 3.7kb backbone to be discarded; lane 3: p2θC31.LacZ digested with *ApaI* and *XhoI*, yielding a 10.3kb backbone, and a 3.9kb fragment to be discarded.

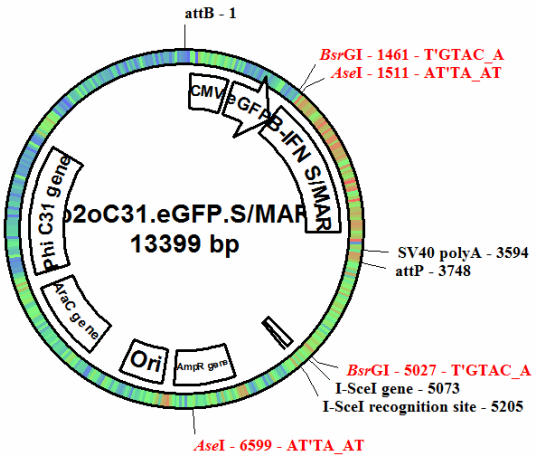


Figure 4.23 The map of construct p2θC31.eGFP.S/MAR indicating restriction enzyme sites *AseI* and *BsrGI*, which were used to confirm the cloning of this vector.

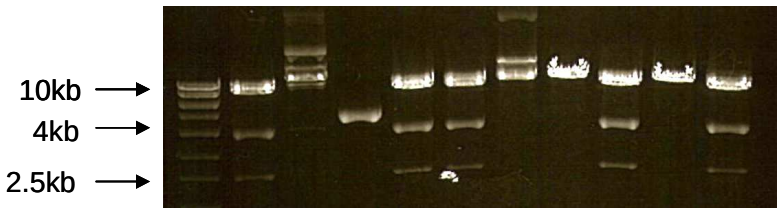


Figure 4.24 A 0.8% gel of 10 different miniprep samples of transformants grown after transformation with potentially cloned p2θC31.eGFP.S/MAR vector, digested with *AseI*. A correctly cloned p2θC31.eGFP.S/MAR vector digested with *AseI* should yield fragment sizes of 8 and 5kb. As shown, none of the samples digested contained a correctly cloned p2θC31.eGFP.S/MAR sequence, and indicated unexpected fragment sizes. Left lane contained a 1kb ladder.

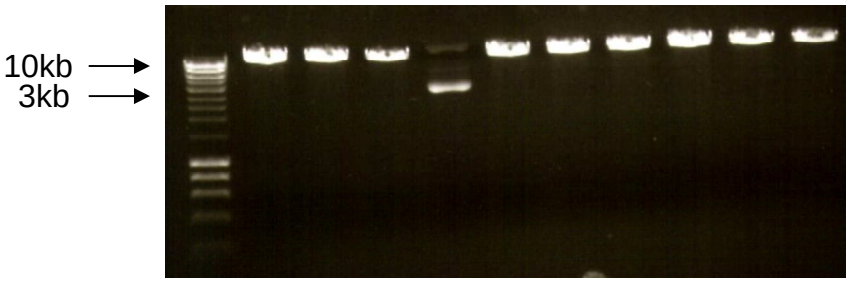


Figure 4.25 A 0.8% gel of 10 different miniprep samples of transformants grown after transformation with potentially cloned p2θC31.eGFP.S/MAR vector, digested with *BsrGI.* A correctly cloned p2θC31.eGFP.S/MAR vector digested with *BsrGI* should yield fragment sizes of 9.8 and 3.5kb. As shown, none of the samples digested contained a correctly cloned p2θC31.eGFP.S/MAR sequence, and indicated unexpected fragment sizes. Left lane contained a 1kb ladder.

4.3 Discussion:

The constructs CeGFP-C1, CpEPI-eGFP, CMYC-pEPI, and p2θC31.eGFP were all successfully cloned. In the cases of CeGFP-C1, CpEPI-eGFP, CMYC-pEPI, these vectors will all be transfected into C2C12 cells, and expression and episomal status will be assessed in subsequent chapters to ascertain the efficiency of the CAGG promoter as opposed to the CMV, in the cases of CeGFP-C1, CpEPI-eGFP, and to investigate the efficiency of the novel vector CMYC-pEPI.

The successful cloning of p2θC31.eGFP meant that eGFP-expressing minicircles could be generated. However, minicircles are unable to replicate within dividing cells (Chen *et al*., 2003; Aryros *et al*., 2011), which would lead to a rapid loss of vector, and hence expression, within a dividing C2C12 population transfected with these minicircles. The vector p2θC31.eGFP was created in order to make minicircle vectors, which would act as one of the controls required for a study which would have involved the evaluation of an S/MAR-containing eGFP-expressing minicircle transfected into a replicating population of C2C12 cells, whose persistence within dividing cells was theorised to be superior to its non-S/MAR counterpart (Argyros *et al*., 2011). However, considering that, despite many repeated attempts using several different strategies had resulted in an inability to successfully clone p2θC31.eGFP.S/MAR, S/MAR-containing eGFP-expressing minicircles could not be generated using this minicircle system, and hence the study was not undertaken.

It is difficult to ascertain the reason behind the failure to clone p2θC31.eGFP.S/MAR. Using strategy 1 where both ends of the insert and vector had been blunted may have posed a challenge for the insert to ligate into the backbone. Despite de-phosphorylation and varying the ratio of vector:insert, most of the restriction patterns indicated the re-circularisation of the parental plasmid p2θC31. Subsequently, a new strategy was designed and attempted. Yet, after many repeated attempts it was still found that p2θC31.eGFP.S/MAR was not being created, and the restriction digests of the plasmid extracted from numerous minipreps indicated, in addition to re-circularisation of the backbone without the inclusion of the insert, some strange restriction patterns. A possible explanation was that p2θC31.eGFP.S/MAR may have been ligating successfully, but that the competent *E.coli* cells into which they were

being electroporated were rearranging the DNA, or that the DNA created was unstable. Several studies have indicated that DNA with repetitive and/or high AT contents, such as S/MAR elements, pose a great difficulty in being cloned (Glockner *et al*., 2002; Gardner 2001). Furthermore, plasmid supercoiling could lead to the formation of DNA secondary structures with elements of such characteristics, which would lead to their rearrangement or deletion (Leach and Lindsey, 1986; Malagon and Aguilera, 1998).

Other minicircle systems have been developed in order to create minicircles, and could be used in future studies to create S/MAR-containing minicircles. One such system, which has already been used to successfully create an S/MAR-containing minicircle, involves the use of the Cre-recombinase system (Argyros *et al*., 2011). In this system, an expression cassette within a parental minicircle plasmid is flanked by loxP sites. These are the sites where intramolecular recombination occurs by the Cre recombinase upon its expression. As with the construct produced by Chen et al, the recombination leads to the formation of a maxicircle containing all the bacterial sequences, and a minicircle containing only the expression cassette and the loxP sites (Argyros *et al*., 2011). Other minicirlce-creating systems have also been created, utilising other recombinases such as the λ integrase (Darquet *et al.*, 1999), or the Flp recombinase (Nehlson *et al*., 2006). Such systems could also be used in order to create S/MAR-containing minicircles in future studies.

Chapter 5: Expression and Episomal Status Assessment of Control and S/MAR Vectors Driven By The CAGG Promoter

5.1 Introduction

In this chapter, a different promoter was tested, replacing the Cytomegalovirus (CMV) promoter used to drive eGFP expression from the eGFP-C1 and pEPI-eGFP vectors in Chapter 3, in an attempt to improve the transgene expression in C2C12 cells. This promoter was the hybrid CMV enhancer/chicken β-actin (CAGG) promoter. The rationale behind choosing this promoter was that several studies have shown it to be ubiquitously expressed (as described in Chapter 1: Introduction, Section 1.8), and it may be less prone to silencing by methylation, as the promoter itself is of mammalian, and not of viral, origin. Moreover, it is possible that the CAGG promoter would also alter the conformation of the eGFP-C1 or pEPI-eGFP vectors which could, according to Stehle *et al* facilitate access of the ORC and initiator proteins, resulting in improved transgene expression and/or vector maintenance within the cells (Stehle *et al*., 2003),.
In this study the expression of the eGFP transgene in eGFP-C1 and pEPI-eGFP was tested under the control of the CAGG promoter (vectors named CeGFP-C1 and CpEPI-eGFP, respectively) (vector maps in Chapter 2: Materials & Methods). The analysis conducted on the cell samples derived included PCR in order to investigate the presence of the vector sequences within each population. Flow cytometry and immunostaining allowed the assessment of the expression of the eGFP reporter gene over time in each population, the trends observed in each population in comparison to the other, and therefore, ultimately, an indication of the efficacy of each vector tested at expressing the transgene over time. Flow cytometry was conducted with the transfected myoblast populations, and immunostaining on transfected myoblasts differentiated into myotubes. The investigation of expression in myotubes was important considering that, *in vivo,* transfected myoblasts eventually differentiate and fuse together to form myotubes, and the expression of a transgene within myotubes is essential for the amelioration of diseases such as Muscular Dystrophy. Fluorescent In-

Situ Hybridisation (FISH) was used to investigate the status of the vectors within the cells as episomal vectors and/or as integrants.

Table of plasmid vectors tested in this chapter:

Vector Name	Vector Feature	Promoter-Transgene	Antibiotic Resistance	Vector Length (kb)
CeGFP-C1	No S/MAR element	CAGG-eGFP	Kan/Neo	5.0
CpEPI-eGFP	Full-length 2kb β-IFN S/MAR element	CAGG-eGFP	Kan/Neo	7.0

Table 5.1 Vectors used in this chapter, indicating the name, the presence of an S/MAR element, the transgene promoter, the transgene used for assessment studies, the antibiotic resistance gene, and the length (kb) of each construct.

5.2 Results

5.2.1 A Comparison Of Stress-Induced Duplex Destabilisation (SIDD) Profiles Generated For CMV-Driven Vectors eGFP-C1 and pEPI-eGFP, and CAGG-Driven Vectors CeGFP-C1 and CpEPI-eGFP

The Stress-Induced Duplex Destabilisation (SIDD) profiles of the CeGFP-C1 and CpEPI-eGFP plasmids were generated in order to assess the conformations of the vectors and identify the areas that were easily base-unpaired. These profiles were compared to those of the CMV-driven eGFP-C1 and pEPI-eGFP vectors. The SIDD profiles of the eGFP-C1, CeGFP-C1, pEPI-eGFP, and CpEPI-eGFP plasmids were calculated by the WebSIDD program designed by Dr. Craig Benham of the UC Davis Genome Centre (http://genomics.ucdavis.edu/benham/sidd/). This program contains an algorithm designed to calculate and predict the areas and amounts of double stranded DNA duplex destabilisation when put under superhelical stress based on the DNA's specific base pair sequence (Bode *et al*., 2006). In superhelical DNA there are sites at which strand separation occurs. The unpairing at these sites is usually quite necessary, such as at promoter regions where transcription proteins bind to initiate transcription, at replication sites, or at scaffold attachment sites (S/MARs) where DNA binds to the nuclear matrix via proteins (as detailed in Chapter 1: Introduction, Section 1.9). The algorithm is able to predict the amount of energy required at each unpairing region for the destabilisation and denaturation to occur. The energy calculated includes the chemical energy required for the paired bases to be separated, the tortional energy which is the twisting of the single strands after separation, and the free energy, which is the supercoiling that occurs in the surrounding areas of the destabilised region (Bi and Benham, 2004). It also takes into account the negative superhelicity that is imposed upon the plasmid sequence as a result of the base unpairing that occurs at easily destabilised regions (Bi and Benham, 2004). A profile is then produced for the entire DNA sequence that, when aligned with a map of the DNA sequence, denotes regions such as open reading frames, promoters, and matrix attachment regions etc. that can destabilise and unpair, and the energy required for them to do so.

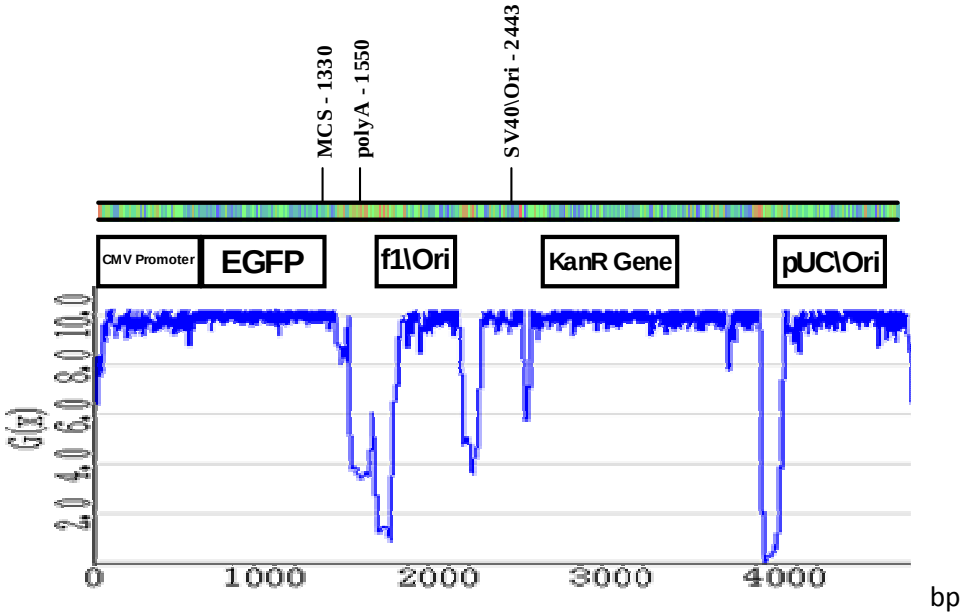
In the S/MAR vector studies conducted by Giannakopoulos *et al*, the SIDD profiles of vectors were calculated using the WebSIDD program (Giannakopoulos *et al*.,2009).

Similarly, the four plasmids in this study were calculated in order to compare the ease with which the promoters destabilise, as this may give an indication of transcription efficiency.

The resulting profiles generated showed that the unpairing at the promoter regions occured at the start of the CMV and CAGG promoters. CMV in eGFP-C1 required approximately 6.0 kcal/mol ($G_{(x)}$) for strand separation to occur, whereas CAGG in CeGFP-C1 required approximately 3.0 kcal/mol (Figure 5.1); and CMV in pEPI-eGFP required approximately 9.0 kcal/mol, whereas CAGG in CpEPI-eGFP required approximately 8.0 kcal/mol (Figure 4.2). As can be seen, the unpairing that occured in the CMV and CAGG promoters of eGFP-C1 and CeGFP-C1, respectively, required less energy than was required for the CMV and CAGG promoters in the S/MAR-containing pEPI-eGFP and CpEPI-eGFP vectors, respectively. This was due to the presence of the S/MAR element that is so easily destabilised throughout the entire 2.0kb region that it induces negative supercoiling in the surrounding regions, which leads to an increase in the amount of energy needed for other areas to denature (Giannakopoulos *et al*., 2009). This was also observed in the unpairing regions surrounding the origins of replication when comparing the control plasmids to the S/MAR-containing plasmids, where the same regions were destabilised in each vector, but the control plasmids required less energy for strand separation than the S/MAR-containing plasmids did.

It was also clear that the CMV-driven eGFP-C1 and pEPI-eGFP vectors had a higher energy requirement to base-unpair than their corresponding CAGG promoter-driven plasmids. Furthermore, the CAGG promoter contained more destabilised regions within its sequence than the CMV promoter, which was demonstrated in both the control and S/MAR-containing plasmids (Figures 5.1 and 5.2). This indicated that the CAGG promoter had a more open conformation than the CMV promoter in these plasmids, which suggested the possibility of superior expression of the reporter gene when driven by the CAGG promoter.

(A) eGFP-C1 Vector



(B) CeGFP-C1 Vector

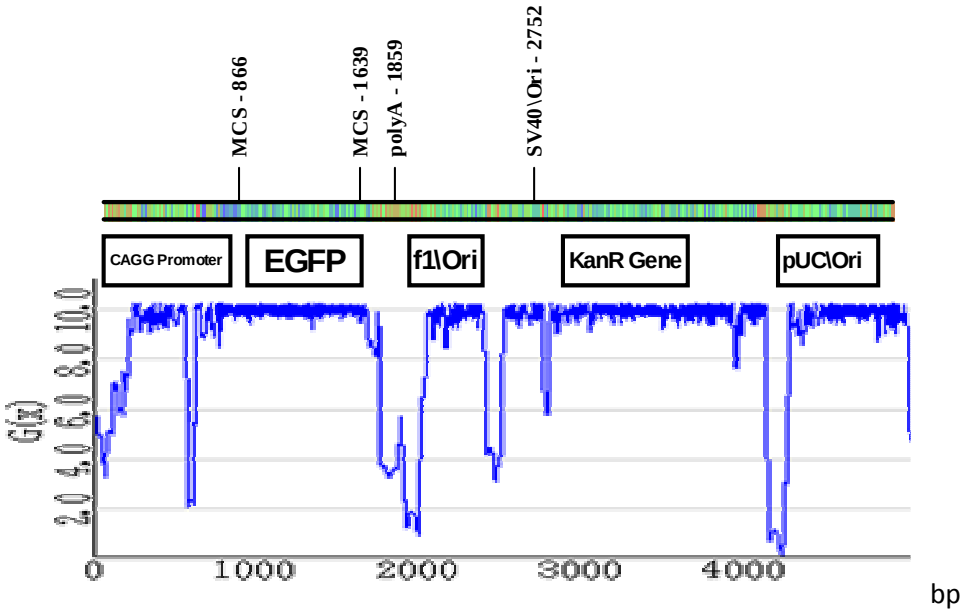
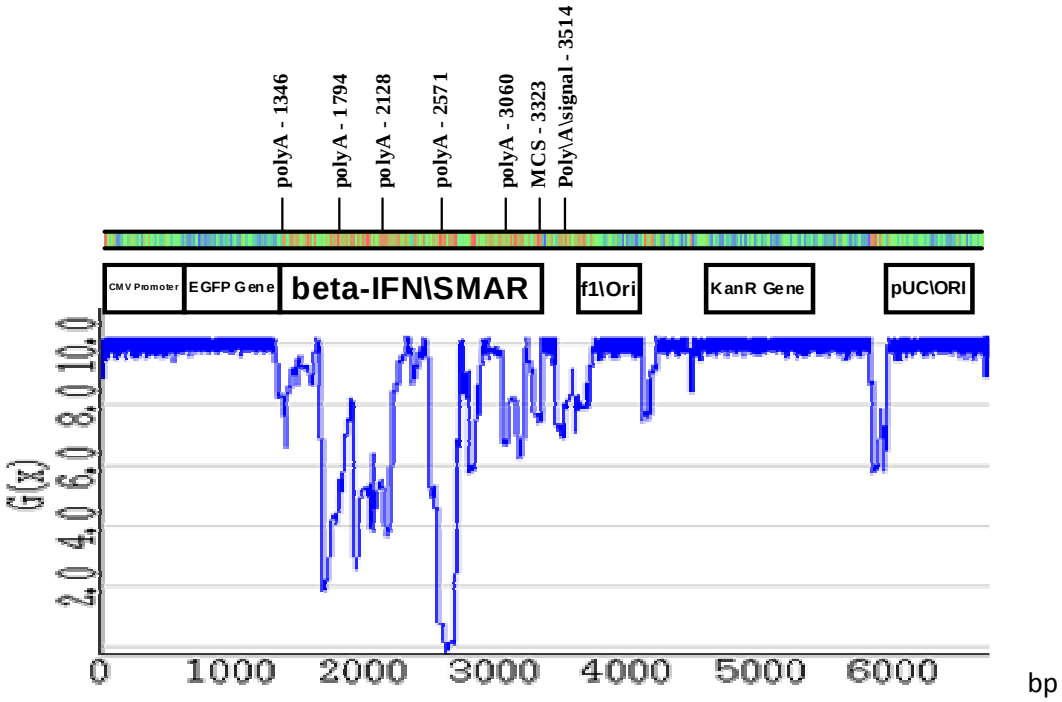


Figure 5.1 SIDD profile of eGFP-C1 with the CMV promoter (A) and of CeGFP-C1 with the CAGG promoter (B), calculated using WebSIDD program. Regions requiring lower amounts of energy ($G(_x)$[kcal/mol]) to base unpair were centred around the origins of replication, and the SV40 promoter region drivng the Neomycin resistance gene. The start of the promoter region of eGFP-C1 ~6.0 kcal/mol for denaturation whereas that of CeGFP-C1 required ~3.0 kcal/mol for denaturation, which was half of that needed for denaturation of the promoter region of eGFP-C1, indicating a more open DNA conformation. The remainder of the promoter region of eGFP-C1 required higher amounts of energy for unpairing, whereas the CAGG promoter in CeGFP-C1 contained other regions of high base unpairing potential, which may allow easier access for transcription factors to bind.

(A) pEPI-eGFP Vector



(B) CpEPI-eGFP Vector

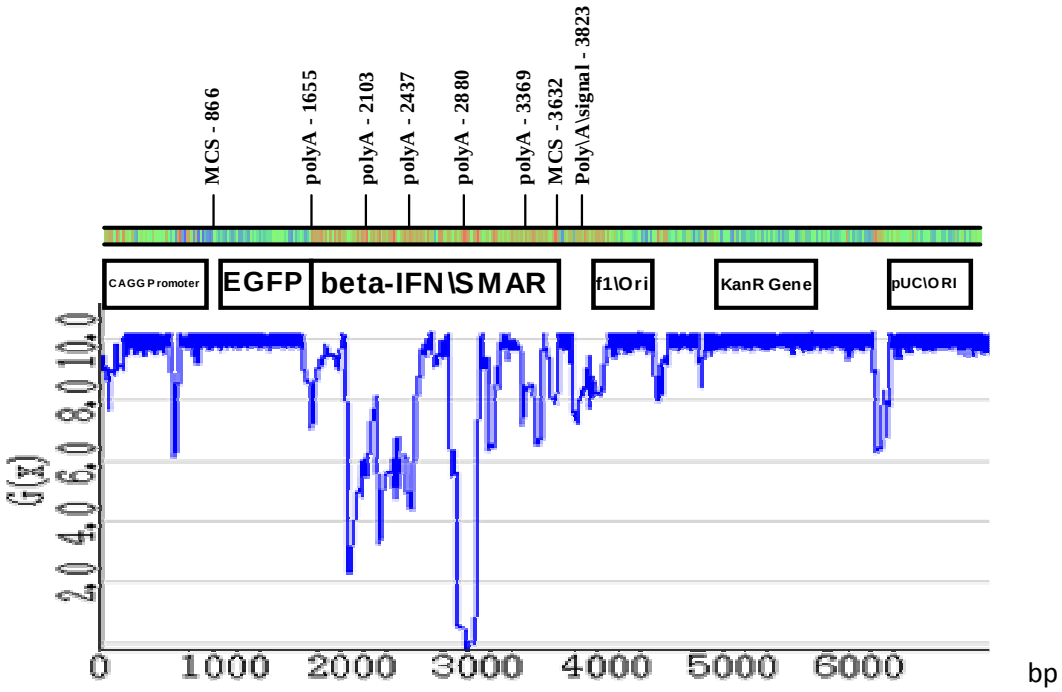


Figure 5.2 SIDD profile of pEPI-eGFP with the CMV promoter (A) and of CpEPI-eGFP with the CAGG promoter (B), calculated using WebSIDD program. Regions of base unpairing which required relatively less energy ($G(_x)$ [kcal/mol]) to destabilise were at the origins of replication, the SV40 promoter region drivng the Neomycin resistance gene, and the S/MAR region. The 2kb S/MAR region required such low amounts of energy to unpair that it induced supercoiling in the surrounding regions, leading to an increase in the energy required for other areas to denature. Therefore, the 5' region of the CMV promoter required ~9.0 kcal/mol for denaturation, and the remainder of the promoter region required greater energy in the pEPI-eGFP vector. This may impair transcription factor access which could lead to

less effective expression of the reporter gene. The 5’ region of the CAGG promoter in CpEPI-eGFP required less energy for denaturation than that of pEPI-eGFP, which was ~8.0 kcal/mol. The CAGG promoter also contained a second region of destabilisation not observed in the CMV promoter, meaning its propensity to unpair was greater than that of the CMV promoter, which may allow easier access for transcription factors to bind.

5.2.2 Experimental Design

In order to test CAGG promoter driven expression in C2C12 cells and compare the S/MAR-containing vector CpEPI-eGFP with the control, CeGFP-C1, the experiment was conducted as follows. $2.5x10^5$ C2C12 cells were plated into $25cm^2$ flasks. After 24 hours the cells had reached ~70% confluency, and were transfected with control vector CeGFP-C1, or CpEPI-eGFP, by lipofection (as described in Chapter 2: Materials & Methods). 24 hours after transfection each population of cells was analysed under a fluorescent microscope to confirm the presence of eGFP positive cells, which indicated that the cells had been successfully transfected with plasmid. Four replicates were conducted, and cell samples were collected from each population for analysis, with the remaining cells from each population seeded into fresh $25cm^2$ flasks at a density of $5x10^4$ cells. At this point the cells were put under selection by the addition of Geneticin (G-418) to the cultures' media. Fresh media containing G-418 was added to each culture every 48 hours, and the cells were passaged when culture confluency reached ~70%. Selection was maintained for 21 days, and on the final day of selection cell samples were again collected from each population for analysis. The remaining cells were seeded again into fresh flasks, and the cells were passaged for a further 39 days. On the final day of the experiment, Day 60, cell samples were again collected from each population for analysis.

5.2.3 Detection of CeGFP-C1 and CpEPI-eGFP Vector Sequences In Transfected C2C12 Cells By Polymerase Chain Reaction (PCR)

PCR was conducted on samples taken on Days 1, 21, and 60 for C2C12 cells transfected with CeGFP-C1 or CpEPI-eGFP. 50ng of total DNA extracted from each sample was used per PCR reaction. The primers used were designed by Primer3 and lie within the Kanamycin resistance gene present in the transfected vectors. The PCR was run for 35 cycles and the expected product size was 503bp. Products were run on 0.8% agarose gels counterstained with ethidium bromide and visualised under UV light. PCR results confirmed the presence of CeGFP-C1 or CpEPI-eGFP vector sequences within the transfected populations at each time point tested (Figure 5.3).

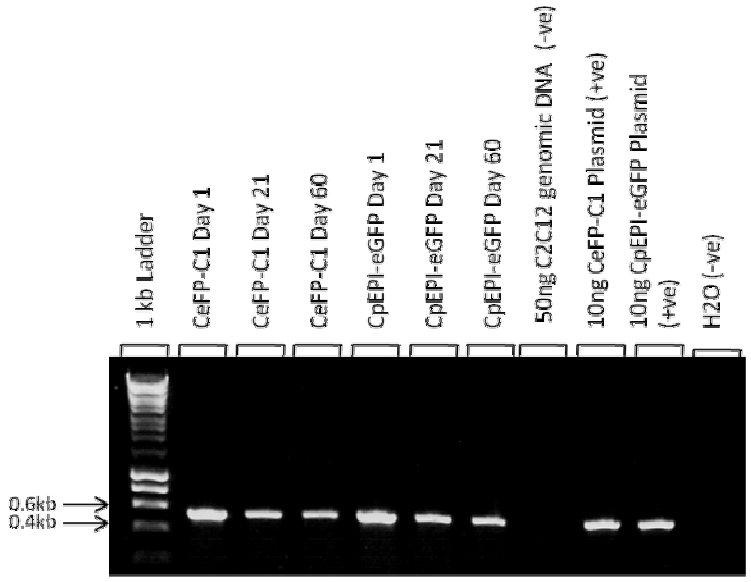


Figure 5.3 PCR conducted on 50ng of total C2C12 DNA transfected with CeGFP-C1 or CpEPI-eGFP plasmid run on samples taken at Days 1, 21, and 60. The positive controls were 10ng of CeGFP-C1, and 10ng of CpEPI-eGFP, and, and the negative controls were 50ng untransfected C2C12 genomic DNA, and H_20. The primers used were within the Kanamycin resistance gene, and product size expected was 503bp. The PCR was run for 35 cycles and products were run on a 0.8% agarose gel. Copies of the vectors were present within cells of the transfected populations at each time point tested.

5.2.4 Expression Analysis of CeGFP-C1 and CpEPI-eGFP Transfected C2C12 Myoblasts Using Flow Cytometry

Expression of $1x10^5$ C2C12 myoblasts transfected with either CeGFP-C1 or CpEPI-eGFP plasmid were analysed by flow cytometry at Day 1 (24 hours post-transfection), Day 21 (final day of a 21 day selection period), and Day 60 (final day of experiment after 39 day of cell proliferation without selection pressure after the 21 day period of selection) to determine the mean percentage of positive cells and the mean fluorescence intensity of each population. The percentages, mean fluorescence intensity, and total eGFP ± the standard error of the means (SEM) were plotted (*n* = 4) (Figures 5.4-5.6, respectively). Statistical tests were applied using NCSS statistics program to determine significance. The test used to compare the transfected populations to each other at each time point, and to compare all of the groups at of the time points was One-Way ANOVA. The P-value was set to 0.05 and indicated a significant difference between groups if $P<0.05$. The post-hoc Tukey-Kramer test was used to determine specifically which groups differed significantly from other groups.

At Day 1 the flow cytometry analysis indicated that CeGFP-C1 transfected a significantly higher percentage of C2C12 cells than CpEPI-eGFP (means of 63.1 ± 0.4% positive CeGFP-C1 transfected cells and of 23 ± 0.3% CpEPI-eGFP transfected cells; P = 0.03) (Figure 5.4). The CeGFP-C1 transfected population also had a higher mean fluorescence intensity (mean of 9803 ± 455.7 for CeGFP-C1 and of 950.8 ± 26.4 for CpePIeGFP; P = 0.03) (Figure 5.5) and hence a higher total expression (means of 618803.2 ± 31925 for CeGFP-C1 and of 22506.3 ± 900.1 for CpEPI-eGFP; P = 0.03) (Figure 5.6).

By Day 21, there was no significant difference found between the percentages of eGFP positive cells of the CeGFP-C1 and CpEPI-eGFP transfected populations (means of 0.3 ± 0.03% and 0.3 ± 0.03 respectively; P = 1.0), the mean fluorescence intensities (means of 694 ± 107.1 and 476.3 ± 63.5 respectively; P = 0.2), or the total expression (means of 174.7 ± 31.7 and 122 ± 26.9 respectively; P = 0.3). However, the decline in the percentage of positive cells for both transfected populations was significant between Days 1 and 21 as was indicated by the Tukey-Kramer test. The decline in the mean

fluorescence intensity was significant between Days 1 and 21 in the CeGFP-C1 transfected population, however it was not significant between these time points in the CpEPI-eGFP transfected population, also as was indicated by the Tukey-Kramer test.

Again, there was no significant difference found between the CeGFP-C1 and CpEPI-eGFP transfected populations by Day 60 in the mean percentage of positive cells (means of 0.01 ± 0% and 0.03 ± 0.03% respectively; P = 1), the mean fluorescence intensity (means of 316.3 ± 316.3 and 305.3 ± 305.3 respectively, P = 0.79), and hence the total eGFP expression (means of 11.1 ± 11.1 and 30.5 ± 30.5 respectively; P = 0.79). The decline was not significant for either of the transfected populations in the percentage of positive cells, mean fluorescence intensities, or total eGFP expression between Days 21 and 60, as was indicated by the Tukey-Kramer test.

It can be seen from these results that although the two populations contained a large proportion of eGFP positive cells at Day 1, expression and percentage of positive cells declined markedly with time, and the levels between the two populations were comparable by Day 60. However, eGFP positive cells could still be detected within the populations by the final day of the experiment, indicating that a small percentage of the populations were able to continually express eGFP protein and contained copies of the vectors over time.

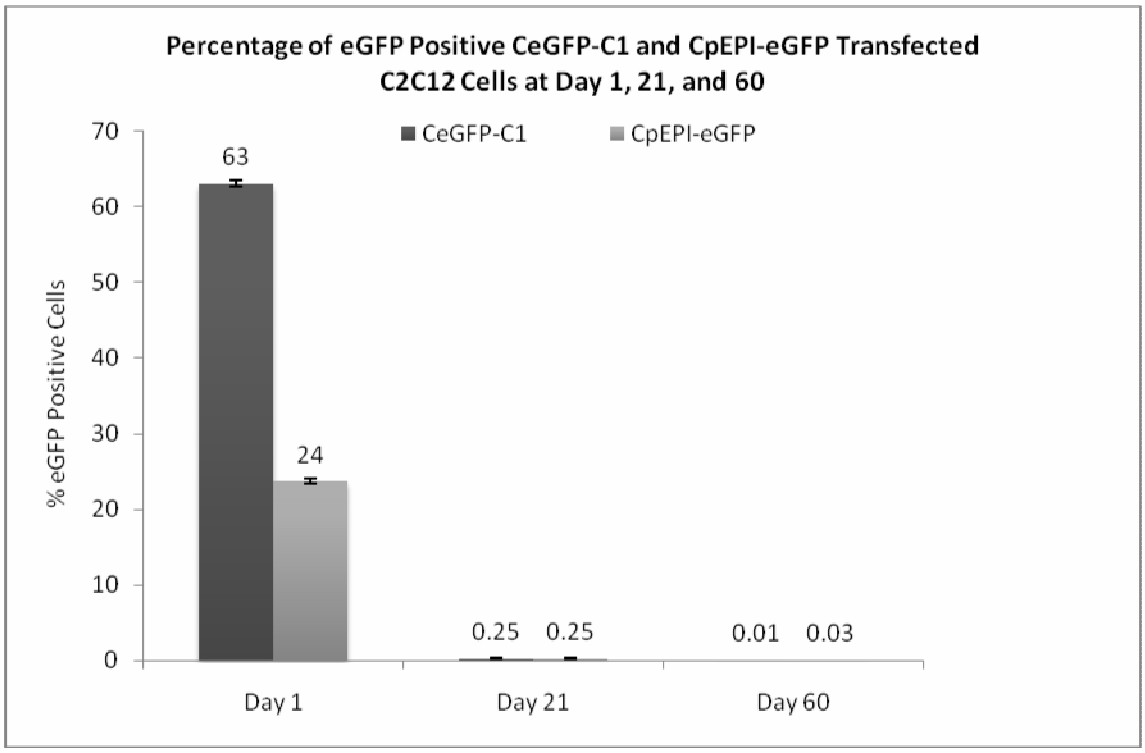


Figure 5.4 Flow cytometry analysis of percentage of eGFP positive C2C12 cells transfected with CeGFP-C1 or CpEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. The mean value is shown above the columns and the error bars represent ± SEM (*n* = 4). At Day 1 the CeGFP-C1 cells had a significantly larger amount of positive cells than the CpEPI-eGFP. By Day 21 the percentages had decreased significantly and the amounts for both populations were not significantly different to each other. A non significant decrease was observed from Day 21 to Day 60, and the percentages of positives in the populations remained comparable.

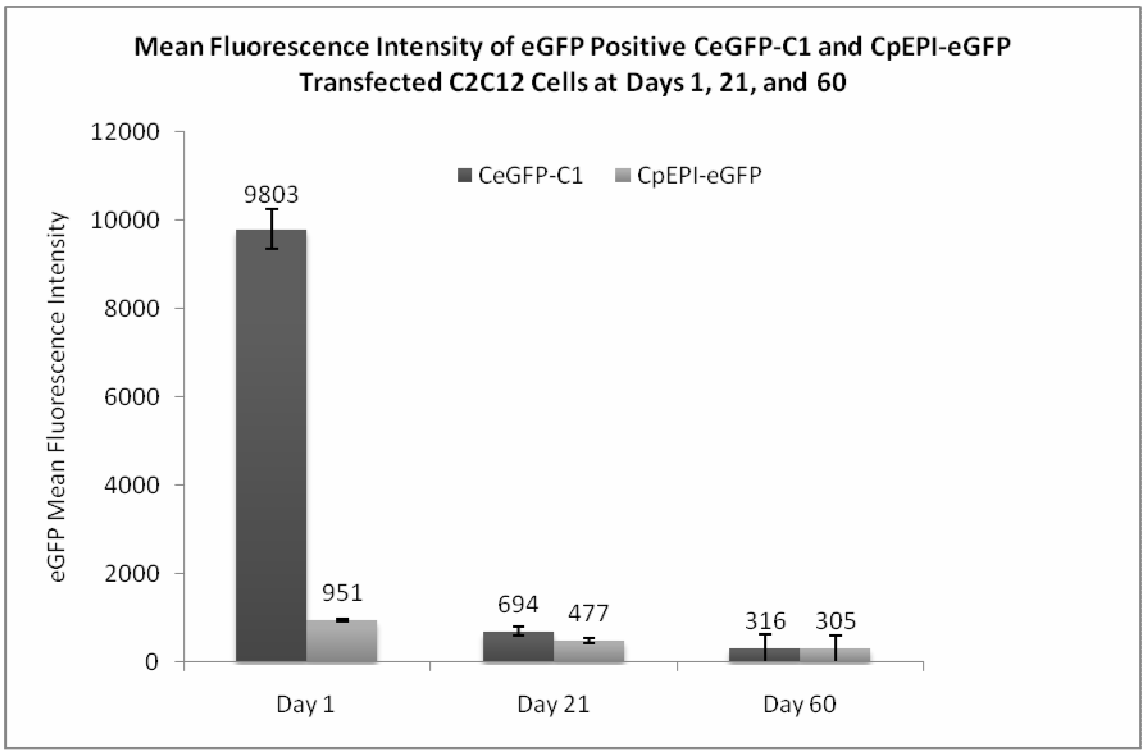


Figure 5.5 Flow cytometry analysis of the mean fluorescence intesity of eGFP positive C2C12 cells transfected with CeGFP-C1 or CpEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. The mean value is shown above the columns and the error bars represent ± SEM (*n* = 4). On Day 1 the CeGFP-C1 cells had a significantly higher mean expression than the CpEPI-eGFP. By Day 21 the mean had significantly decreased in the CeGFP-C1 population but had not significantly decreased in the CpEPI-eGFP population, and the expression levels for both populations were not significantly different. A non significant decrease was observed by Day 60, and the mean expressions of both populations remained comparable to each other.

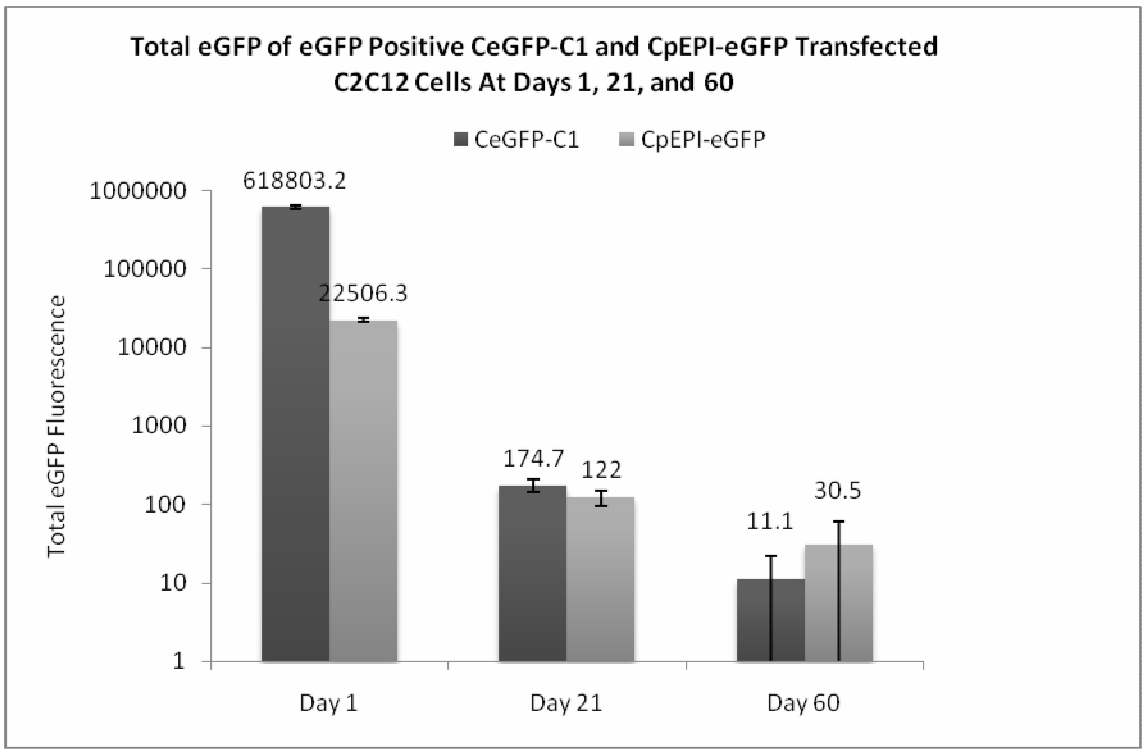


Figure 5.6 Total eGFP of eGFP positive C2C12 cells transfected with CeGFP-C1 or CpEPI-eGFP vector at 24 hours post transfection (Day 1), the final day of selection after a 21 day selection period (Day 21), and the 39th day of proliferation without any selective pressure post-selection (Day 60), where 1x10^{5} cells were analysed per sample. The mean value is shown above the columns and the error bars represent ± SEM (*n* = 4). The total eGFP fluorescence values were converted to log as the data was spread over a wide range. As can be seen on Day 1 the CeGFP-C1 cells had a significantly higher mean expression than the CpEPI-eGFP. By Day 21 the total expression had decreased markedly for both populations, and the total eGFP in the populations were not significantly different to one another. A non significant decrease was observed by Day 60, and the total expression of both populations remained comparable.

5.2.5 Expression Analysis Of CeGFP-C1 and CpEPI-eGFP Transfected C2C12 Cells Differentiated Into Myotubes By Immunostaining

The differentiation of transfected myoblasts into myotubes may have an effect upon transgene expression of the CeGFP-C1 and CpEPI-eGFP plasmids. $5x10^4$ cells were isolated from each transfected population at Day 1 (24 hours post-transfection), Day 21 (final day of a 21 day selection period), and Day 60 (final day of experiment after 39 day of cell proliferation without selection pressure after the 21 day period of selection) and seeded into $1cm^2$ chamber slide wells. The cells were differentiated by changing the medium from growth medium to differentiation medium for 5 days. The cells were then fixed in 1% PFA/PBS and immunostained for eGFP (as described in Chapter 2: Materials & Methods).

The results of the staining suggested that the CeGFP-C1 transfected myotubes, differentiated from Day 1 myoblasts, expressed more eGFP than the equivalent CpEPI-eGFP transfected myotubes, as indicated by the brightness and intensity of eGFP fluorescence (Figure 5.7 (A) and (B)). Myotubes, which were differentiated from a sample of transfected myoblasts derived at Day 21, indicated that eGFP positive myotubes were present in both transfected populations, however the amount of expression was, again, greater in the CeGFP-C1 transfected population than in the CpEPI-eGFP transfected population as was subjectively noted, as the cells appeared brighter, though only marginally (Figure 5.7 (C) and (D)). The CeGFP-C1 myotubes expressed a lower amount of eGFP on Day 21 than they did on Day 1, whereas the CpEPI-eGFP cells expressd more than they did on Day 1, judging by the brightness of fluorescence. No eGFP was detected by immunofluorescence in either the CeGFP-C1 or the CpEPI-eGFP myotubes that were differentiated from myoblasts derived from the Day 60 cultures (Figure 5.7 (E) and (F)), suggesting silencing of the transgene, a loss of vector, or insufficient sensitivity of this method to detect lower levels of transgene expression.

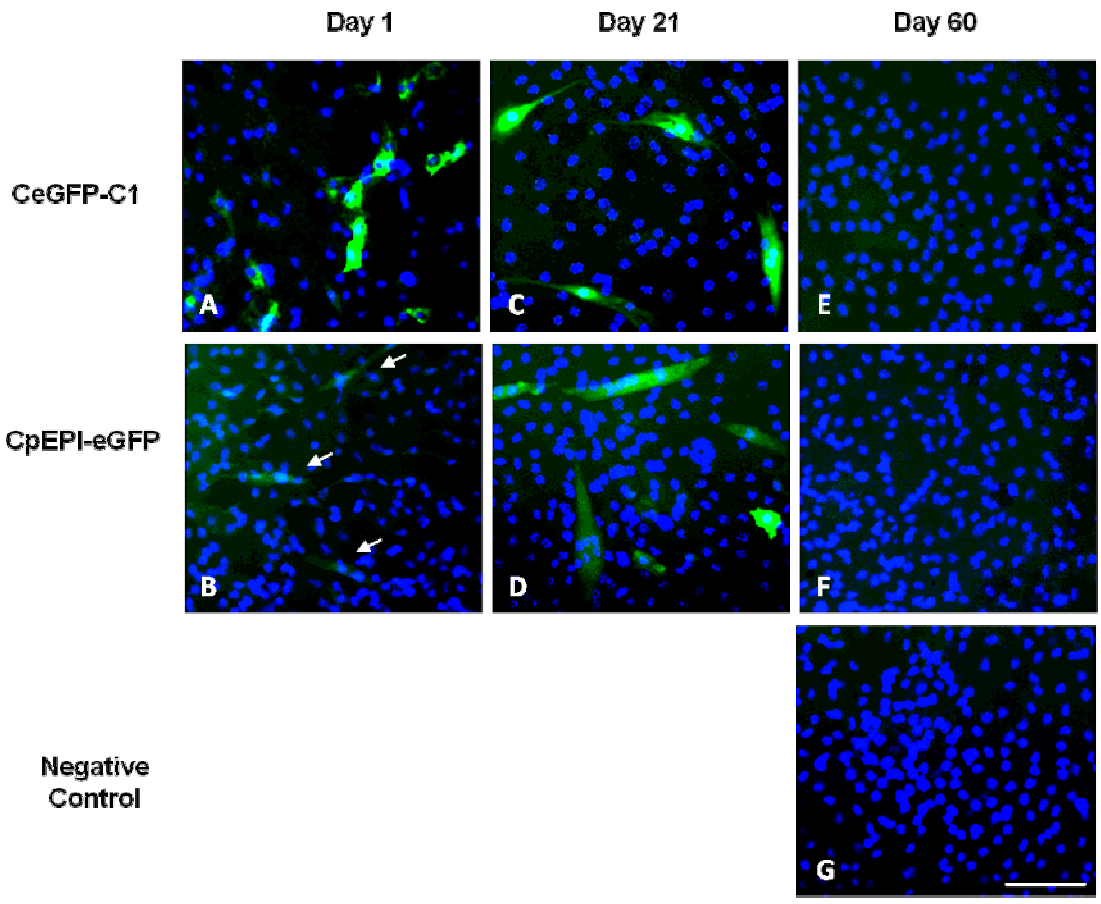


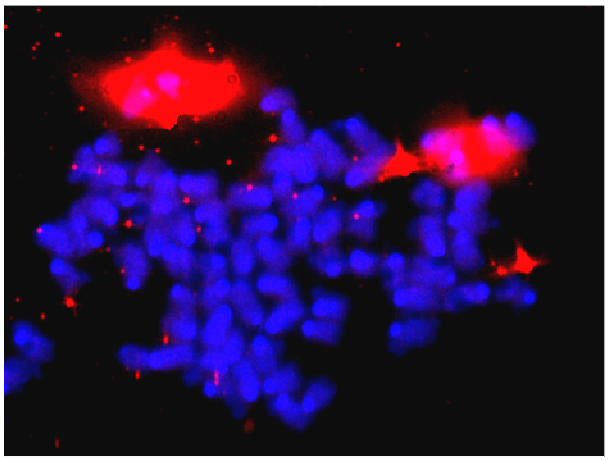
Figure 5.7 eGFP immunostaining of C2C12 myotube cultures transfected with CeGFP-C1 or CpEPI-eGFP vector and differentiated into myotubes for 5 days before fixing. Samples from the transfected populations were differentiated at three time points. Day 1 indicates the CeGFP-C1 and CpEPI-eGFP transfected populations that were put into differentiation medium 24 hours post transfection (images A and B respectively), Day 21 indicates the populations that were put under differentiation medium on the final day of selection after a 21 day selection period (images C and D, respectively), and Day 60 indicates the populations that were put into differentiation medium on the 39th day of proliferation without any selective pressure post-selection (images E and F, respectively). The negative control was untransfected C2C12 myotubes (image G). Green indicates eGFP positive cells and blue is the DAPI staining of the nuclei. The images indicate that CeGFP-C1 myotubes expressed more eGFP at Day 1 than the CpEPI-eGFP myotubes. At Day 21 eGFP positive myotubes were detected in both populations, however there was more eGFP protein in the CeGFP-C1 myotubes than the CpEPI-eGFP, subjectively, as was judged by the brightness of the signal. Expression in the myotubes increased from Day 1 to Day 21 in the CpEPI-eGFP transfected population, as the Day 21 myotubes were brighter. At Day 60 no eGFP positive myotubes were detected in the CeGFP-C1 or the CpEPI-eGFP transfected populations, denoting expression had either been silenced or expressed at levels too low to be detected by this method. Bar, 90µm.

5.2.6 Investigation Of Episomal/Integrant Status Of CeGFP-C1 and CpEPI-eGFP Vectors In Transfected C2C12 Cells By Fluorescent In-Situ Hybridisation (FISH) Analysis On Final Day Samples

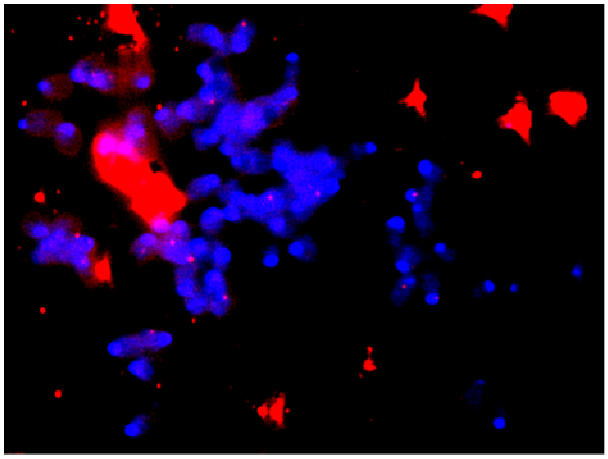
FISH was used to determine the episomal and/or integrant status of the transfected vectors (as conducted in Results Chapter 3, and as described in Chapter 2: Materials & Methods). The positive controls used in this experiment were C2C12 cells 24 hours post transfection with CpEPI-eGFP or CeGFP-C1 (Figure 5.8 (A) and (B), respectively). The negative controls in this experiment were untransfected C2C12 cells (Figure 5.8 (C)). From the images taken under the fluorescence microscope it can be seen that a large amount of plasmid was present within the positive control cells, as the fluorescence signal was very strong. It was also clear that much of the plasmid observed was clustered together.

FISH analysis showed no evidence of the CpEPI-eGFP plasmid integration into the host genome (0/6 positive spreads) (Figure 5.9 (A)). However, the data is suggestive of CeGFP-C1 plasmid integration into the genome (4 integrant spreads out of 5 positive spreads) (Figure 5.9 (B)). FISH has also shown that copies of CpEPI-eGFP and CeGFP-C1 were also still present as episomes in their respective populations after many rounds of replication.

(A)C2C12 cells 24 hr after transfection with CpEPI-eGFP



(B)C2C12 cells 24 hr after transfection with CeGFP-C1



(C)C2C12 cells untransfected

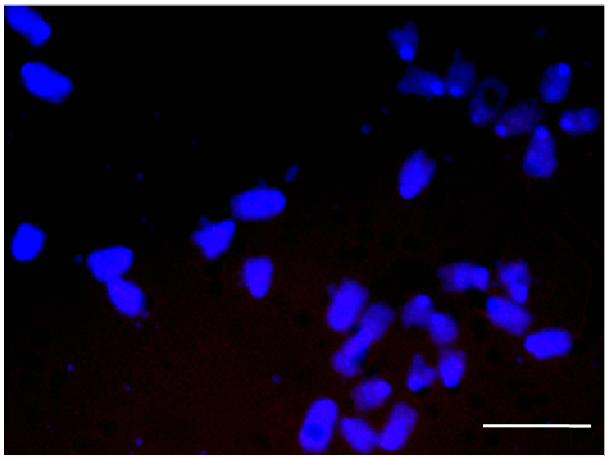
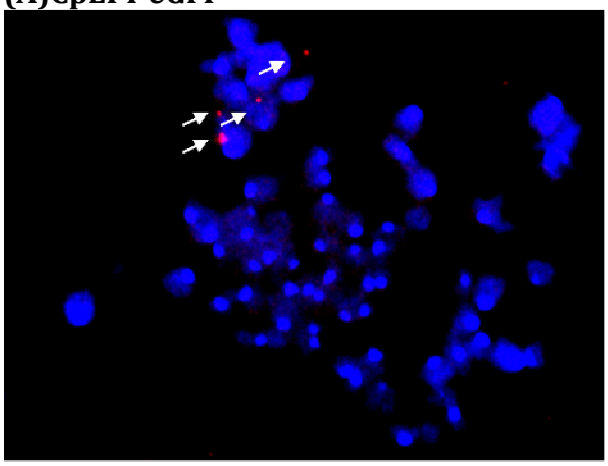


Figure 5.8 Fluorescent In-Situ Hybridisation positive control of C2C12 cells 24 hours post transfection with the CpEPI-eGFP plasmid (A) and CeGFP-C1 plasmid (B). Vector copies can be seen as the bright red fluorescing red spots. Genomic DNA was counterstained with Dapi (blue). A large amount of vector was present within the transfected cells, and the plasmid was found mostly clustered together. The negative control was untransfected C2C12 cells (C) indicating no background. Bar, 5µm.

Transfected C2C12 cells Day 60

(A)CpEPI-eGFP



(B)CeGFP-C1

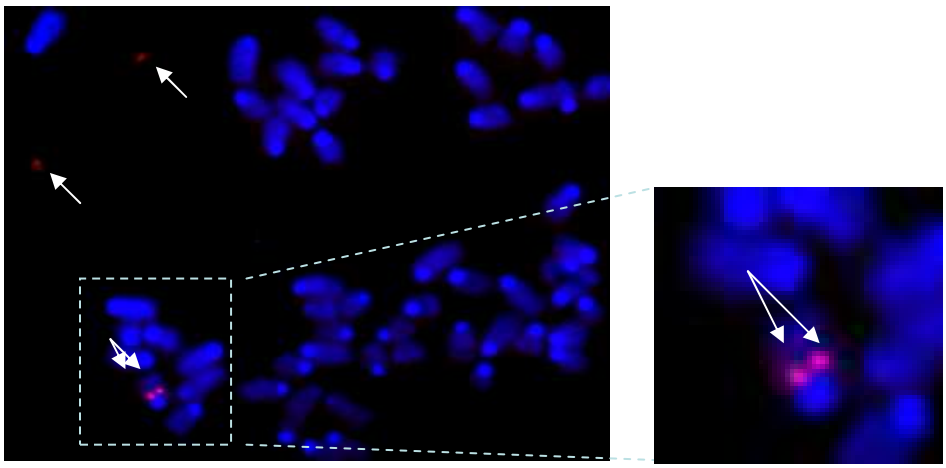


Figure 5.9 Fluorescent In-Situ Hybridisation of C2C12 cells transfected with CpEPI-eGFP plasmid (A) or CeGFP-C1 plasmid (B), of samples collected at Day 60. Single white arrows indicate red fluorescence spots where the probe has hybridised to plasmid sequence and denotes the episomal status of the plasmid. The double arrows mark spots of potential integration. Genomic DNA was counterstained with Dapi (blue). From these images it can be seen that CpEPI-eGFP plasmid has been retained as an episome with no evidence of integration (0/6 positive spreads). CeGFP-C1 appears to have integrated into the C2C12 host genome (4 integrant spreds out of 5 positive spreads) in addition to being retained as an episome by Day 60.

5.2.7 Summary of Results:

		PCR	Flow Cytometry			Immunostaining of Myotubes	FISH	
			% positive cells	MFI	Total		Integration	Episomal
Vector	Day							
CeGFP-C1	1	√	63.1 ± 0.4	9803 ± 455.7	618803.2 ± 31925	+	0/6	√
	21	√	0.3 ± 0.03%	694 ± 107.1	174.7 ± 31.7	+		
	60	√	0.01 ± 0%	316.3 ± 316.3	11.1 ± 11.1	-		
CpEPI-eGFP	1	√	23 ± 0.3	950.8 ± 26.4	22506.3 ± 900.1	+	4/5	√
	21	√	0.3 ± 0.03%	476.3 ± 63.5	122 ± 26.9	+		
	60	√	0.03 ± 0.03%	305.3 ± 305.3	30.5 ± 30.5	-		

Table 5.2 Table summarising the results obtained in this chapter for transfected C2C12 cells, indicating each vector that was transfected and the time point at which samples were extracted and analysed. PCR results are presented where a 'v' denotes that the Kanamycin sequence was amplified and a band was present; Flow Cytometry results presented include the percentage of positive cells per transfected population, the mean fluorescence intensity (MFI) of the positive cells, and the total amount of expression which was the product of expression and number of positive cells; the Immunostaining data is presented by '-' indicating the absence of positive myotubes, '+' indicating the presence of less than 10 positive myotubes, and '++' indicating the presence of more that 10 positive myotubes; the FISH data is presented as the number of slides which suggested an integration event having occurred out of the total number of slides analysed on Day 60 only, a 'v' in the episomal column indicated the presence of episomal copies of the vector observed on the slides, and a '-' indicates none was observed.

5.3 Discussion:

5.3.1 Transgene expression driven by the CAGG promoter from CeGFP-C1 and CpEPI-eGFP vectors declined with time

The aim of the experiments presented in this chapter was to investigate the expression of the CAGG driven eGFP reporter gene in the CeGFP-C1 and CpEPI-eGFP vectors in C2C12 cells, and to investigate whether these vectors integrate into the host genome. SIDD profiles were generated of the plasmids in order to investigate the regions within the DNA sequences that have a high propensity for base unpairing, and the amount of energy required for denaturation to occur. The CMV promoter had less regions of destabilisation than the CAGG promoter, and required a greater amount of enery for strand separation, as was shown by comparing the control plasmids to one another and the S/MAR-containing plasmids to one another. This lower energy requirement of the CAGG promoter led to the hypothesis that replacing the CMV promoter with the CAGG promoter to drive the reporter gene may result in improved expression, especially in the S/MAR-containing vector where the destabilisation of the S/MAR element led to negative supercoiling in the surrounding region which led to the higher energy requirements for CAGG promoter strand separation in relation to the control vector CeGFP-C1.

Considering that the CMV-containing vectors were not tested at the same time as the CAGG-containing vectors in C2C12 cells, expression from the two different promoters cannot be directly compared. However, it is clear that the expression profile over time of the C2C12 cells transfected with CeGFP-C1 and CpEPI-eGFP in this chapter show a similar trend as the C2C12 cells transfected wtih eGFP-C1 and pEPI-eGFP in Chapter 3 in terms of the decline in expression. At Day 1 cells were eGFP positive but by Day 21 the percentage of positive cells and expression had declined in both populations, as indicated by flow cytometry. By Day 60 the percentage of positive cells and the expression of both populations were comparable and at very low levels, as observed in the eGFP-C1 and pEPI-eGFP transfected C2C12 cells in Chapter 3. Likewise, the immunostaining of the myotubes showed a similar trend where no eGFP positive myotubes could be detected by Day 60 in either population.

However, an important difference was observed, albeit subjectively, which was the increase in the brightness of the eGFP positive myotubes from Day 1 to Day 21 in the CpEPI-eGFP transfected population, which may indicate superior long term expression of the CAGG promoter from the S/MAR-containing plasmid than that of the CMV promoter in the S/MAR-containing plasmid. Although eGFP positive myotubes were present in the pEPI-eGFP transfected population at Day 21 in Chapter 3, a decrease and not an increase in fluorescence was observed between Days 1 and 21. It is interesting to note, however, that this increase in fluorescence intensity was not found in the C2C12 myoblasts transfected with either of the vectors CeGFP-C1 or CpEPI-eGFP by flow cytometry, and may have occurred exclusively in the myotubes due to the different factors present within differentiated cells as compared to undifferentiated ones, or an accumulation of eGFP, as discussed in Chapter 3 Discussion.

5.3.2 Total eGFP expression and percentage of positive cells of CeGFP-C1 transfected C2C12 cells were greater than those Of CpEPI-eGFP transfected cells at Day 1

The flow cytometry data demonstrated that CeGFP-C1 transfected a larger number of cells and expressed eGFP more efficiently than CpEPI-eGFP, as was shown by the percentage of positive cells, the mean fluorescence intensity, and thus the total eGFP expression of the samples derived at Day 1. This difference in the mean percentage of positive cells and expression on Day 1, and thus the transfection efficiency, may be attributed to plasmid size. Kreiss *et al* found that transfection efficiency is inversely proportional to plasmid size where the greater the size of the plasmid, the lower the efficiency of transfection and thus the total expression (Kreiss *et al*., 1999). CeGFP-C1 (5kb) is smaller in size than CpEPI-eGFP (7kb), which may explain the difference found between the transfection efficiencies of the two vectors.

The supercoiling of a plasmid is another factor that can influence transfection efficiency (Maucksch *et al*., 2009). A study by Cherng *et al* confirmed that the transfection efficiency with the use of polyplexes of supercoiled DNA is greater than that of open-circular and/or heat-denatured plasmid DNA, whose efficiencies are similar to one another (Cherng *et al*., 1999). It is possible that the lack of an S/MAR

element in eGFP-C1 allowed the plasmid to be more tightly supercoiled as it was being produced in bacteria. The S/MAR element, as previously explained, has AT rich tracts that easily unpair, which may have led to a more open DNA conformation such as open-circular, hence leading to a lower transfection efficiency.

Another factor which may have influenced transgene expression by CpEPI-eGFP was the negative supercoiling in the surrounding regions of the S/MAR, namely at the CAGG promoter region. As the S/MAR element must be inserted within an open reading frame of a transcribed gene for it to function as an element that prevents the plasmid from being integrated into the genome and maintains it as an episome, as reported in the literature (Chapter 1: Introduction, Section 1.10), the destabilisation at the site of the S/MAR, as shown by the SIDD profiles generated, led to negative supercoils being introduced further down in the plasmid in the promoter region, which resulted in the CAGG promoter in CpEPI-eGFP to require a relatively greater amount of energy for strand separation than the same promoter in the CeGFP-C1 plasmid. This phenomenon was demonstrated in the study by Giannakopoulos *et al* (Giannakopoulos *et al*., 2009). If the unwinding of DNA requires more energy, then the frequency of transcription would be lower than that of DNA that can unwind in the necessary places with more ease and less energy. And, as concluded by Giannakopoulos *et al,* supercoiling at the promoter region would lead to the cell's transcription factors to have more difficulty in expressing the gene due to impaired access to the promoter (Giannakopoulos *et al*., 2009). Thus, this may have contributed to the lower mean fluorescence intensity of the CpEPI-eGFP transfected cell population compared with the control vector CeGFP-C1.

An additional factor that may also have accounted for the differences in mean expression resulting from the two vectors was the efficiency of the polyadenylation signal at the 3' end of the eGFP transgene. Considering that the S/MAR element was located within the open reading frame of a transcribed gene and before the SV40 polyadenylation signal, the S/MAR was also transcribed, and it has been observed that the β-IFN S/MAR has 5 polyadenylation stop codons within it (vector map in Chapter 2: Materials & Methods). Different types of polyadenylation sequences exist, some more efficient than others, which can have an effect on gene expression (Azzoni *et al*., 2007). A study done by Azzoni *et al* comparing the three polyA signals BGH, Synt, and SV40 showed that each one had a different amount of purine-rich sequences which gave

each signal a different capacity to form single-stranded DNA structures, which in turn influenced transcription efficiency (Azzoni *et al*., 2007). Based on this, it is possible that cells transfected with CeGFP-C1 expressed greater amounts of eGFP than the CpEPI-eGFP transfected cells was due to the presence of the efficient SV40 polyA signal as part of the eGFP expression cassette in CeGFP-C1. CpEPI-eGFP contained the S/MAR sequence directly after the eGFP gene at its 3' end and prior to the SV40 signal at its 5' end, and the possibility that the stop signals within the S/MAR element were not as efficient as the SV40 signal cannot be excluded.

It is also possible that the resulting mRNA transcript which included part, or all, of the S/MAR sequence was not as stable a transcript as that transcribed from CeGFP-C1, which led to lower protein expression, as was found in a study by O'Rourke *et al,* where the sequence of a transcribed transgene affected the stability and half-life of the mRNA transcript, and thus the level of reporter protein within cells (O'Rourke *et al*., 2002).

5.3.3 Evidence of CeGFP-C1 integration into the C2C12 host genome was found whereas none was found of CpEPI-eGFP

In investigating the presence of the plasmid sequences using PCR it was found that a copy of each plasmid was present at every time point tested within their respective populations. The FISH data showed that there was no evidence of CpEPI-eGFP integration into the C2C12 host genome by Day 60 and that it was found as an episome, whereas the CeGFP-C1 plasmid may have integrated in addition to being present extrachromosomally. As mentioned in Chapter 3, not having observed any integration events within the CpEPI-eGFP transfected population does not rule out the possibility that integration may still have occurred, however this would indicate that integration, if it did occur, happened at a lower frequency than that of the CeGFP-C1 plasmid. Hence, the S/MAR element could have a protective role which keeps its plasmid mostly, if not completely, episomal in status within C2C12 cells.

5.4 Concluding Remarks

The potential integration of CeGFP-C1 makes this vector an unsuitable candidate for use in gene therapy due to safety issuses, as was found with eGFP-C1 in Chapter 3. However, there was no evidence of integration of CpEPI-eGFP, and episomal copies were present within the cells by Day 60, again indicating the possibility of a protective role of the β-IFN S/MAR element in preventing integration of the vector into the host genome. Nonetheless, transgene expression levels and the percentage of positive cells in the CpEPI-eGFP transfected population were very low at the end of the experiment, as was observed in Chapter 3 where CMV was used to drive eGFP expression. It is possible that the low levels in total expression may have, in part, been due to a a relatively low amounts of episomal vector copies retained per cell with time, and not as a result of transgene silencing, when considering the fact that there was no significant difference or decline in the mean expression of this population between Days 1 and 60. Therefore, an improvement in vector retention and maintenance is clearly essential before CpEPI-eGFP can be used as a tool for gene therapy. Improving vector retention was the focus of the following chapter: Chapter 6.

Chapter 6: Analysis Of Long-Term Transgene Expression and Vector Status Of CpEPI-eGFP and CeGFP-C1 In C2C12 Cells Held In G0/G1 Phase of The Cell Cycle/Quiescence Post-Transfection

6.1. Introduction

As concluded in Chapter 5, it is important to find a mechanism which allows a larger amount of plasmid copies to be retained episomally in order to have a greater amount of reporter protein expression. As was observed in Chapter 3, the HepG2 cells, which are slow dividing cells with a relatively long cell cycle, retained pEPI-eGFP episomally after many cell divisions, and expression was at high levels by the final day of the experiment. As was discussed in Chapter 3 discussion, the long cell cycle may indicate a slower rate of replication fork progression, which in turn may influence the attachment of the S/MAR vector to the nuclear matrix. A slower rate of fork progression leads to a greater amount of replicons during DNA replication, and hence a greater amount of attachment points between origins of replication and the nuclear matrix (Courbet *et al*., 2008). In view of the fact that the S/MAR element has been proposed to also act as an origin of replication in pEPI-eGFP within mammalian cells (as reviewed in Chapter 1: Introduction, Section 1.9), this may have led to a greater amount of pEPI-eGFP S/MAR-matrix binding, and thus a greater amount of pEPI-eGFP may have been anchored to the matrix, leading to an increased incidence of replication initiation from the S/MAR element, and a larger amount of vector being retained within the cells. A relatively faster rate of replication fork progression in C2C12 cells would indicate fewer points of attachment, where those of relatively lower affinity to the matrix remaining unattached (Courbet *et al*., 2008). This would lead to attached origins initiating before unattached origins (Courbet *et al*., 2008) and, in considering the speed of the cell cycle, may not allow the binding of sufficient amounts of pEPI-eGFP to the matrix, leading to fewer copies of vector being replicated and retained within the cells.

It was found that initiation sites are determined in the G1 phase of the cell cycle, a step called origin decision point (Li *et al*., 2003), and that pre-initiation complexes are assembled on all origins, dominant or latent, in G1 (as reviewed in Chapter 1: Introduction, Section 1.9.1). A longer cell cycle, and hence a relatively slower rate of division, may have allowed the transfected HepG2 cells to have had more time from the moment pEPI-eGFP entered the nucleus for the plasmid to be epigenetically marked. It is this epigenetic marking which allows the vector to be continually propagated with divisions, expressed, and retained, as the marking occurs as a result of the S/MAR's presence within the plasmid (Bode *et al*., 2006). It is plausible that by attempting to arrest cells that have a relatively short cell cycle, such as C2C12 cells (Kitzman *et al*., 1998), in quiescence or in the G0/G1 phase for a period of time, that this may achieve several things. First, that it may lead to a decrease in the replication fork rate during the first S phase once the cells have been switched media as they prepare to arrest, leading to more origins, with various affinities to the matrix, binding to the matrix. Second, that during the extended arrest time in G0/G1 that the cells would be given more time for more S/MAR vector (CpEPI-eGFP) copies to anchor to the nuclear matrix, for pre-initiation complexes to be assembled on these vectors, and for these vectors to be epigenetically marked (Chapter 1: Introduction, Sections 1.7-1.9). And third, that once these cells were released from arrest, considering that the organisation of replicons and hence origin initiation, is carried out in the same manner as it is in the previous S phase, before having to revert back to the usual replicon organisation characteristic of the cell during replication (Courbet *et al*., 2008), that the vectors bound to the matrix would be replicated and consequently marked by the cell. Such marks may include termination marks which remain present on replicated DNA until mitosis (Courbet *et al*., 2008), hence leading to further replication of the vectors in daughter cells for many generations.

The G0 phase of the cell cycle and quiescence can both be defined as "a reversible withdrawal from the cell cycle" where the cells are not proliferating and have stepped out of the cell cycle (Blow and Hodgson, 2002), and the terms 'quiescence' and 'G0' are often used interchangeably. For the aims of this chapter, C2C12 was an ideal cell line to use as the cells can be put into quiescence or G0/G1 phase of the cell cycle, as demonstrated by Yoshida *et al* and Kitzman *et al* (Yoshida *et al*., 1998; Kitzmann *et al*.,

1998) (as described in Chapter 1: Introduction, Section 1.3.3). Both methods were used to induce cells post-transfection into quiescence (as described in Chapter 2: Materials & Methods,) and were compared in this chapter.

In this study, C2C12 cells were transfected with CeGFP-C1 or CpEPI-eGFP plasmid. The CAGG promoter was used to drive transgene expression instead of the CMV promoter in these vectors for two reasons. The first was that, although there did not appear to be a large difference between CAGG- or CMV-driven expression as expression was extremely low and eGFP positive myotubes were not detected by immunostaining by Day 60, the immunostaining in Chapter 4 suggested an increase in eGFP expression in the positive myotubes in the CpEPI-eGFP population from Day 1 to Day 21, which was not observed in the pEPI-eGFP transfected population in Chapter 3. The second reason was that the SIDD profiles generated in Chapter 4, where the CMV- and the CAGG-driven constructs were compared, indicated a more open DNA conformation at the site of the CAGG promoter constructs CeGFP-C1 and CpEPI-eGFP and, in accordance with Riu *et al*'s study (as described in Chapter 1: Introduction, Section 1.7), a more open conformation and an active DNA state could lead to the DNA being marked with euchromatin-associated markers (Riu *et al*., 2007).

The aims of this chapter were to transfect cycling C2C12 cells with CeGFP-C1 and CpEPI-eGFP, then hold them in quiescence or G0/G1 using two separate experimental methods, followed by analysing and comparing long term expression as a result of the arrest. Overall transgene expression was analysed at several time points, the expression resulting from the groups held in quiescence/G0/G1 by two separate methods were compared, and the expression profiles resulting from the S/MAR containing vectors (CpEPI-eGFP) in each group were compared to those resulting from the control vectors of each group (CeGFP-C1). Furthermore, the integrant/episomal status of each vector was investigated. The analysis conducted on the cell samples in order to fulfil the aims of this study included PCR to confirm the presence of the vector sequences within each population. Flow cytometry and immunostaining allowed the assessment of the expression of the eGFP reporter gene over time in each population and the trends observed in each population in comparison to the others. Finally, FISH was used to investigate the status of the vectors within the cells as episomal vectors and/or as integrants.

Table of plasmid vectors tested in this chapter:

Vector Name	Vector Feature	Promoter-Transgene	Antibiotic Resistance	Vector Length (kb)
CeGFP-C1	No S/MAR element	CAGG-eGFP	Kan/Neo	5.0
CpEPI-eGFP	Full-length 2kb β-IFN S/MAR element	CAGG-eGFP	Kan/Neo	7.0

Table 6.1 Vectors used in this chapter, indicating the name, the presence of an S/MAR element, the transgene promoter, the transgene used for assessment studies, the antibiotic resistance gene, and the length (kb) of each construct.

6.2. Results

6.2.1 Experimental Design

The experiment was conducted as follows (Figure 6.1). Two 185cm^2 flasks were seeded with 18.5x10^5 C2C12 cells. After 24 hours, when the confluency had reached ~70%, one flask was transfected with CeGFP-C1 and the other with CpEPI-eGFP, by lipofection (as described in Chapter 2: Materials & Methods). 24 hours post transfection each population of cells was viewed under a fluorescent microscope to confirm the presence of eGFP positive cells, which indicated that the cells had been successfully transfected with plasmid. Cell samples were collected from each of the two transfected populations for analysis. Next, the remaining cells of the CeGFP-C1 and CpEPI-eGFP populations were each seeded into eight 25 cm^2 flasks, at 5x10^4 cells per flask. Four flasks of the CeGFP-C1 population and four flasks of the CpEPI-eGFP population were cultured in 2% horse serum DMEM medium for 10 days in order to differentiate the myoblasts and yield a culture of myotubes and quiescent reserve cells. The myotubes were separated from the reserve cells (as described in Chapter 2: Materials & Methods). In order to confirm that reserve cells were C2C12 myoblasts that had been induced into quiescence, samples were fixed immediately after isolation from the myotubes, as any delay in this process could lead to reserve cell reactivation, then immunostained for the presence of lysenin, as conducted in a study by Nagata *et al* (Nagata *et al*., 2006). Nagata *et al* had found that quiescent muscle satellite cells, in addition to quiescent reserve cells isolated from a differentiated culture of C2C12 cells, contain a high level of the lipid sphingomyelin on the plasma membrane (Nagata *et al*., 2006). Upon activation, the levels of sphingomyelin decrease. Lysenin is a protein that binds specifically to sphingomyelin (Yamaji *et al*., 1998). In their study, Nagata *et al* used lysenin to label these quiescent cells, then used a fluor-conjugated antibody against lysenin to stain and thus identify quiescent cells within a population (Nagata *et al*., 2006). The reserve cells in this study were identified using the same method (as described in Chapter 2: Materials & Methods). The cells were also simultaneously stained for eGFP in order to confirm that of the reserve cells isolated, some were eGFP

positive and therefore had contained the transfected vectors within their nuclei whilst in quiescence.

The reserve cells isolated from the myotubes in each transfected population were then trypsinised and re-seeded into fresh flasks. The cells were put into growth medium to re-activate them from quiescence, and G-418 was added to eliminate cells that did not contain the vectors CeGFP-C1 or CpEPI-eGFP. These groups of cells will herein be referred to as 'Reserve CeGFP-C1' and 'Reserve CpEPI-eGFP'. Selection was maintained for 10 days. Fresh media containing G-418 was added to each culture every 48 hours, and the cells were passaged when culture confluency reached ~70%. On the final day of selection cell samples were collected again from each flask for analysis. The remaining cells were re-seeded into fresh flasks, and the cells were passaged for a further 39 days. On the final day of the experiment, cell samples were collected again from each flask for analysis.

The other four flasks of the CeGFP-C1 population and the four flasks of the CpEPI-eGFP population were put under low serum (1% FCS) methionine-depleted medium 24 hours post transfection in order to arrest the cells at the G0/G1 phase of the cell cycle (Kitzman *et al*., 1998). Methonine is needed for cells to grow, but is not essential for cell viability (Nadal-Ginard, 1978). These groups of cells will herein be referred to as 'G0 CeGFP-C1' and 'G0 CpEPI-eGFP'. When kept in this medium for an extended period of time the cells remain in G0/G1 until the medium is changed back to growth medium. In order to investigate the percentage of cells in G and S phases in relation to the amount of time the cells were kept in this medium, samples were taken at several time points, cells' nuclei were isolated, genomic DNA was stained with DAPI, and the fluorescence signal of the DAPI was quantified for each nucleus by flow cytometry (as described in Chapter 2: Materials & Methods).

The signal obtained of DAPI fluorescence in the flow cytometer depends upon the amount of DNA present within a cell. In the G1 phase of the cell cycle cells have not replicated their DNA yet. In the S phase, the cells begin to replicate their DNA and a quantifiably greater signal, which is variable according to the stage of S phase the cells are in, is emitted. By G2 phase all DNA in a cell has been replicated, and therefore doubled, meaning the stained nuclei would give a signal intensity that is double that obtained from nuclei of cells in the G1 phase. This is how each phase was determined. Cells held in G0 are cells that have been arrested in the G1 phase of the cell cycle.

Vermeulen *et al* have confirmed that cells that pass the checkpoint at the end of G1 phase cannot go back into G1 and must complete the cell cycle (Vermeulen *et al*., 2003). Those which do not and are serum-starved may enter directly from G1 into G0. If the cells are starved after they have passed G1 into S phase, they have been found to be required to complete the cell cycle and pass through mitosis before entering the quiescent state (Pardee 1974).

The cells were cultured in low serum methionine-depleted medium and arrested for 7 days, after which the medium was changed back to growth medium for 24 hours to allow the cells to recover and re-enter the cell cycle. The cells were then put under selection for 10 days by the addition of Geneticin (G-418) to the cultures' media, fresh media containing G-418 was added to each culture every 48 hours, and the cells were passaged when culture confluency reached ~70%. On the final day of selection cell samples were collected from each flask for analysis. The remaining cells were re-seeded into fresh flasks, and the cells were passaged for a further 39 days. On the final day of the experiment, cell samples were collected again from each flask for analysis.

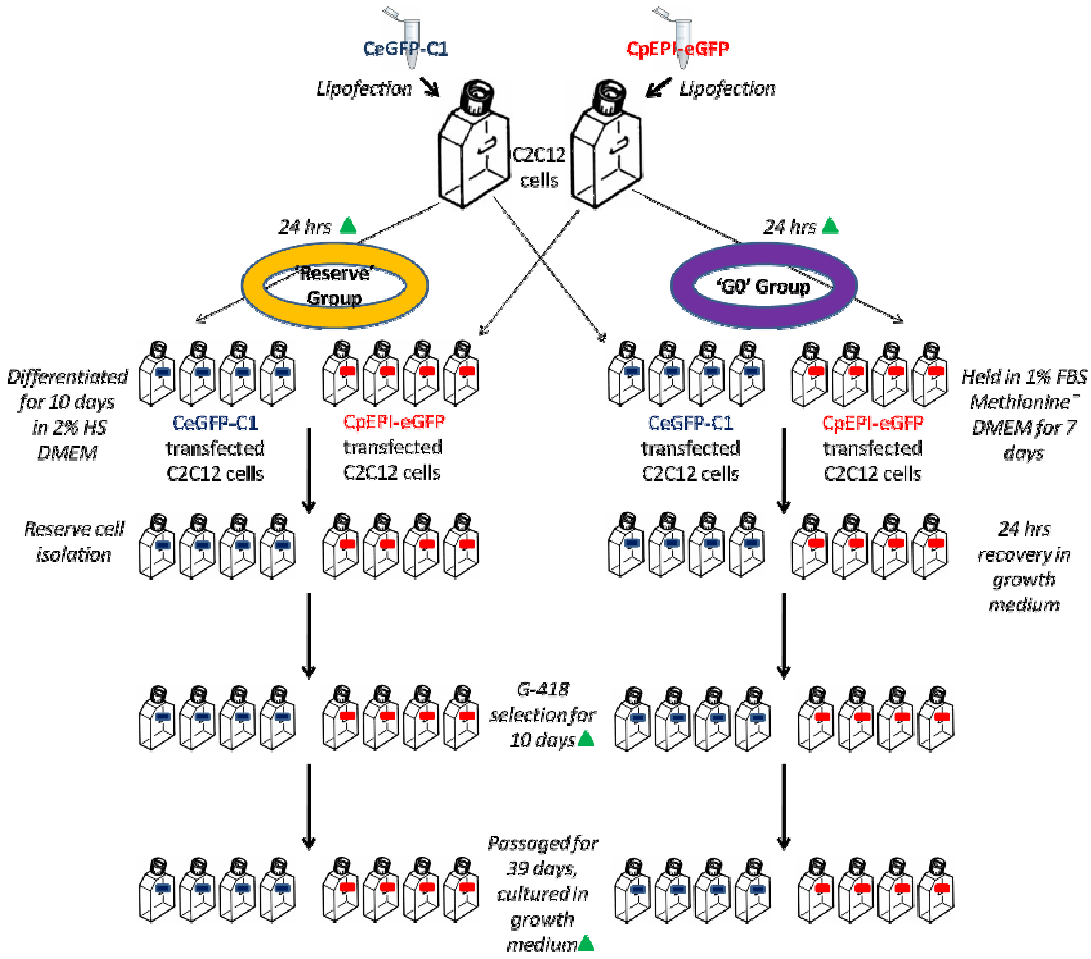


Figure 6.1 Flow diagram depicting experimental sequence. C2C12 cells transfected with CeGFP-C1 (blue) or CpEPI-eGFP (red), separated each into 8 flasks, 4 of each in the 'Reserve' group and 4 in the 'G0' group. The 'Reserve' group refers to the cells that were differentiated, then had the myotubes separated from the reserve cells. The 'G0' group refers to the cells that were held in G0 phase of the cell cycle using methionine depleted media. The cells in both groups were then put under selection prior to proliferation in growth medium for 39 days. The green triangle marks the time points at which samples were taken from each flask for further analysis.

6.2.2 Differentiation Of eGFP Positive C2C12 Myoblasts Transfected With CeGFP-C1 and CpEPI-eGFP Into Myotubes

In order to induce cells within a cycling population of C2C12 cells to enter quiescence it was necessary to first differentiate them into myotubes by switching the cells from growth medium to differentiation medium (as described in Chapter 2: Materials & Methods). The myoblasts then commit to either a path of differentiation into myotubes, or enter quiescence and become reserve cells. C2C12 cells were transfected with CeGFP-C1 or CpEPI-eGFP, then allowed to differentiate for 10 days. Fluorescence and phase contrast images were taken of eGFP positive myotubes in the differentiating cultures to show the progression from myoblasts to myotubes. An observed difference was noted between both the number of cells and the expression levels, as judged by the brightness of eGFP fluorescence, of cells transfected with CeGFP-C1 (Figure 6.2, row A, Day 1) and of those transfected with CpEPI-eGFP (Figure 6.2, row C, Day 1), which were greater than those of the CpEPI-eGFP population at Day 1. By Day 2 the cells had been in low serum medium for 24 hours and could be seen to have changed morphology by lengthening and thinning. By Day 4 many myoblasts had fused together to form long myotubes (Figure 6.2, rows B and D, Day 4). The intensity of the eGFP appeared to increase as differentiation proceeded in the CpEPI-eGFP transfected population, either due to an accumulation of eGFP protein or an increase in expression as a result of differentiation. Interestingly, this was not observed in the CeGFP-C1 transfected population. By Day 7 longer myotubes were observed in both populations (Figure 6.2, rows B and D, Day 7), and eGFP positive myotubes were present in both transfected populations at this time point (Figure 6.2, rows A and C, Day 7).

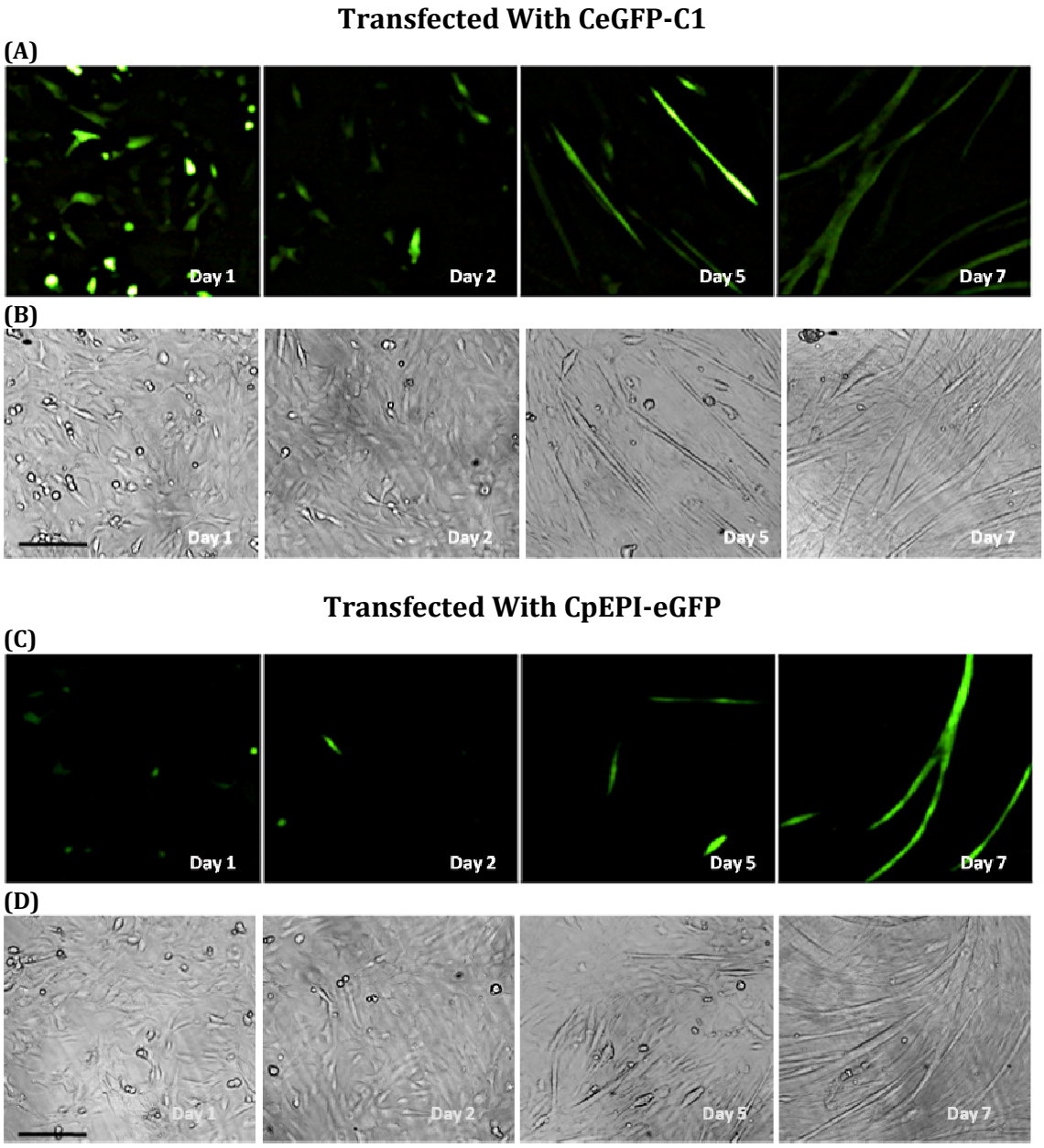


Figure 6.2 Fluorescence images of C2C12 cells transfected with CeGFP-C1 plasmid (row A), and phase contrast images of the same cells (row B), and of cells transfected with CpEPI-eGFP plasmid (row C), and their phase contrast images (row D), differentiating after being put under low serum medium 24hrs post-transfection. An observed difference was found of the amount of positive cells and the intensity of the eGFP between the CeGFP-C1 and CpEPI-eGFP transfected cells on Day 1, where they were greater in the CeGFP-C1 population. By Day 2 the cells had begun to change morphology and elongate. By Day 5 many cells had fused together to from myotubes, and by Day 7 longer and larger myotubes were observed. It appeared that the intensity of eGFP fluorescence increased in the CpEPI-eGFP transfected population with time, but not in the CeGFP-C1 transfected population. Bar, 90µm.

6.2.3 Confirmation Of eGFP Positive Reserve Cell Isolation By Lysenin/eGFP Immunostaining

Reserve cells were isolated after the myoblast culture had been differentiated for ten days by a series of washes with a low concentration of trypsin (as described in Chapter 2: Materials & Methods). This caused the myotubes to detach first from the flask, leaving the reserve cells still attached to the substrate. The reserve cells were then detached using a higher concentration of trypsin and plated into fresh flasks.

C2C12 cells were also seeded into chamber slide wells, differentiated, and the myotubes separated from the reserve cells. These cells were not trypsinised but were rinsed with PBS and immediately fixed in 4% PFA/PBS prior to staining. Rabbit anti-lysenin and mouse anti-eGFP primary antibodies were used, and swine anti-rabbit and goat anti-mouse secondary antibodies labelled with red fluorescent Alexa Fluor 488 dye and green fluorescent Alexa Fluor 488 dye, respectively. The staining was conducted in duplicates and images were taken under fluorescence microscopy. From the images taken it was evident that both the CeGFP-C1 and CpEPI-eGFP transfected populations had contained eGFP positive reserve cells (Figure 6.3). It was also shown that not all the cells that had remained in the chamber wells after trypsinisation were reserve cells, as many did not stain positive for lysenin.

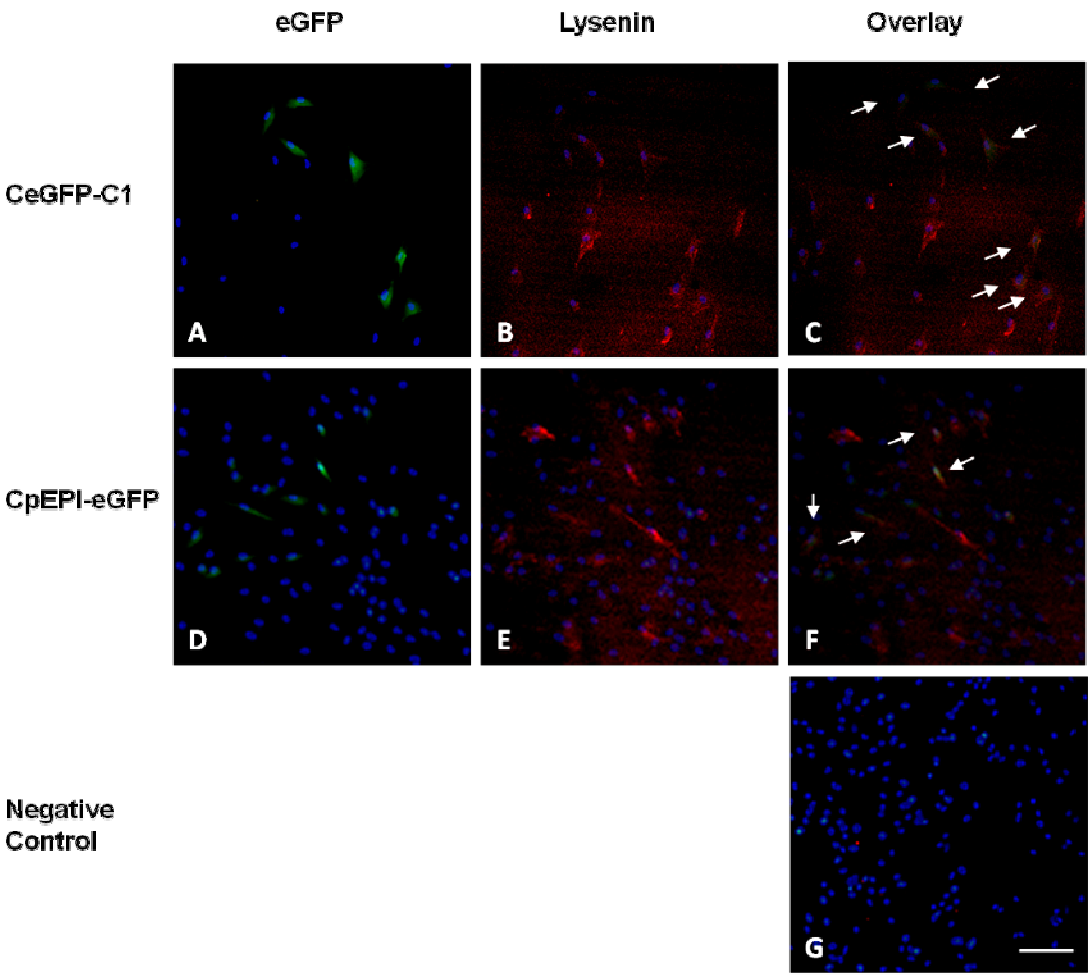


Figure 6.3 Immunostaining of C2C12 cell samples transfected with CeGFP-C1 or CpEPI-eGFP and stained for eGFP (images A and D, respectively) or lysenin (images B and E, respectively), and overlayed images of the eGFP and lysenin staining (images C and F, respectively) are shown. Cells were fixed immediately after the isolation of myotubes from reserve cells and stained with anti-Lysenin in order to identify reserve cells (red), and anti-eGFP in order to indentify eGFP-positive cells (green), then mounted with Dapi mounting medium to stain the nuclei (blue). White arrows point out eGFP positive lysenin positive cells. The images were taken under a fluorescence microscope and were overlayed in order to determine the presence of eGFP positive reserve cells within the two populations. The images show that eGFP positive reserve cells were identified in both transfected populations and thus confirmed that the plasmids transfected were present in the nuclei of the quiescent C2C12 reserve cells in the differentiated cultures. The lysenin staining also showed that some differentiated cells that had not fused into myotubes, but were not reserve cells either as they did not stain positive for lysenin, still remained within the culture. The negative control was a sample of proliferating untransfected C2C12 cells stained for eGFP and lysenin, images overlayed (image G). Bar, 90µm.

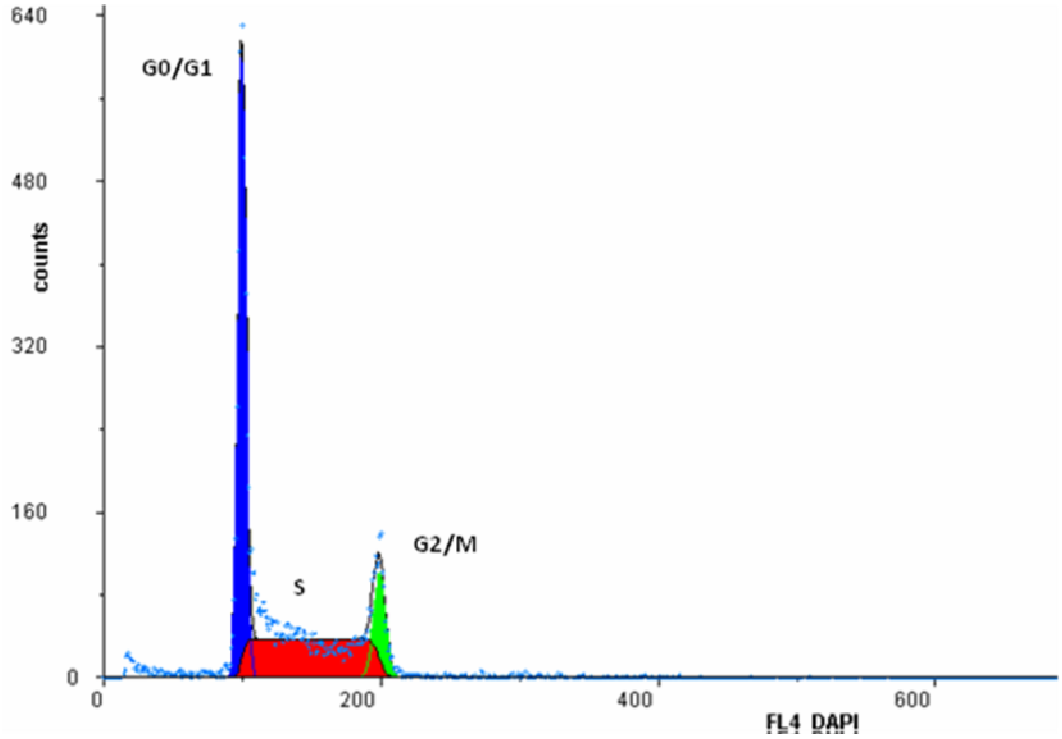
6.2.4 Analysis of Cell Cycle Arrest Of C2C12 Cells Held In G0 By Flow Cytometry

Nuclei of untreated proliferating C2C12 cells and of C2C12 cells put under low serum methionine depleted medium, where samples were taken at days 2, 5, and 7, were isolated (as described in Chapter 2: Materials & Methods). The genomic DNA was stained with DAPI and the number of nuclei was plotted against the fluorescence intensity of the DAPI from each nucleus in order to determine the percentage of cells in each phase of the cell cycle. The percentages of cells in each phase were plotted (Figure 6.4) where the blue peak represents cells in the G0/G1 phase, the red represented cells in the S phase, and the green peak represented cells in the G2/M phase. Figure 5.4 shows representative distributions obtained at Days 0 and Day 7, and indicate the decrease in the number of cells in S phase after cells were put in low serum methionine depleted medium for 7 days, as cells in the G1 phase entered and were arrested in the G0 phase without progression into the S phase.

The analysis was conducted in triplicate (n = 3), the SEM values were calculated, and the values plotted (see Figure 6.5). Untreated proliferating C2C12 cells (Day 0 sample) contained 42.1% ± 1.9% in the G0/G1 phase, 44.7% ± 1.5% in the S phase, and 13.2% ± 0.4% in the G2/M phase, thus indicating a large amount of cells were actively proliferating and were in the S phase. At Day 2 (two days after the addition of the low serum methionine depleted medium) the number of cells in the G0/G1 phase had already increased to 61.5% ± 0.4%, the number of cells in the S phase had declined to 20.1% ± 0.2%, and the number of cells in the G2/M phase had increased to 18.4% ± 0.6%. This showed that many cells that had previously been in the S phase had passed through the G2/M phase and entered the G0/G1 phase. At Day 5 the number of cells in each phase remained similar to those at Day 2, where those in the G0/G1 phase made up 62.8% ± 2.3 of the population, those in S phase made up 18.2 % ± 0.5 of the population, and those in the G2/M phase made up 18.9% ± 1.9 of the population. At Day 7 the percentage of cells in G0/G1 phase was 74.1% ± 4.7, in S phase 5% ± 1.6, and in G2/M phase 20.9% ± 4. It is clear that by Day 7 the number of cells arrested in G0/G1 phase had increased and those in S phase had further declined and passed into the G2/M phase. The percentages in the G2/M phase had remained similar

throughout, possibly indicating that the number of cells passing the checkpoint from S phase into G2/M was similar to the number of cells passing from G2/M phase into G0/G1 phase. It was also observed that a large amount of cell death had occurred by Day 7, and fewer cells were present and still viable for flow cytometry analysis.

(A) Percentage of C2C12 cells in G0/G1, S, and G2/M phase proliferating under growth medium



(B)Percentage of C2C12 cells in G0/G1, S, and G2/M phase after 7 days in methionine depleted medium

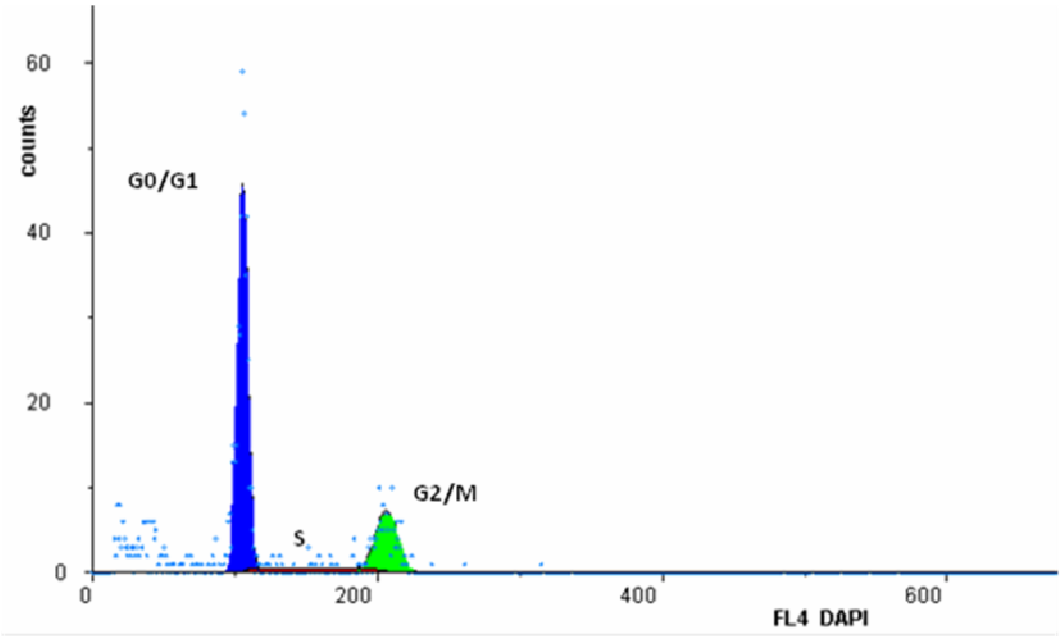


Figure 5.4 Flow cytometry analysis of cell cycle status of untreated C1C12 population at Day 0 (image A) and of C2C12 cells after 7 days in low serum methionine depleted medium (image B). Nuclei were isolated from each sample and their genomic DNA was stained with DAPI. The number of nuclei were

counted (counts) and plotted against their fluorescence intensity (FL4 DAPI). The G2/M peak showed cells in this phase emitted double the intensity of those cells in the G0/G1 phase due to the completion of the doubling of genomic DNA once the cells have entered G2 phase. The blue peak in the charts represents cells in the G0/G1 phase which, in the untreated cells comprised 40.04% of the population, and at Day 7 comprised 68.81%. The red represents cells in S phase and comprised 46.52% of the untreated cell population, and 8.11% at Day 7. The green peak represents cells in the G2/M phase and comprised 13.44% of the total population of untreated cells, and 23.08% at Day 7. It can be seen that the number of cells in G0/G1 phase had increased after 7 days in low serum methionine depleted medium, and that the number of cells in S phase had declined, suggesting that most cells that had exited the S phase continued in the cell cycle until they reached the G1 phase, where they were able to arrest in G0.

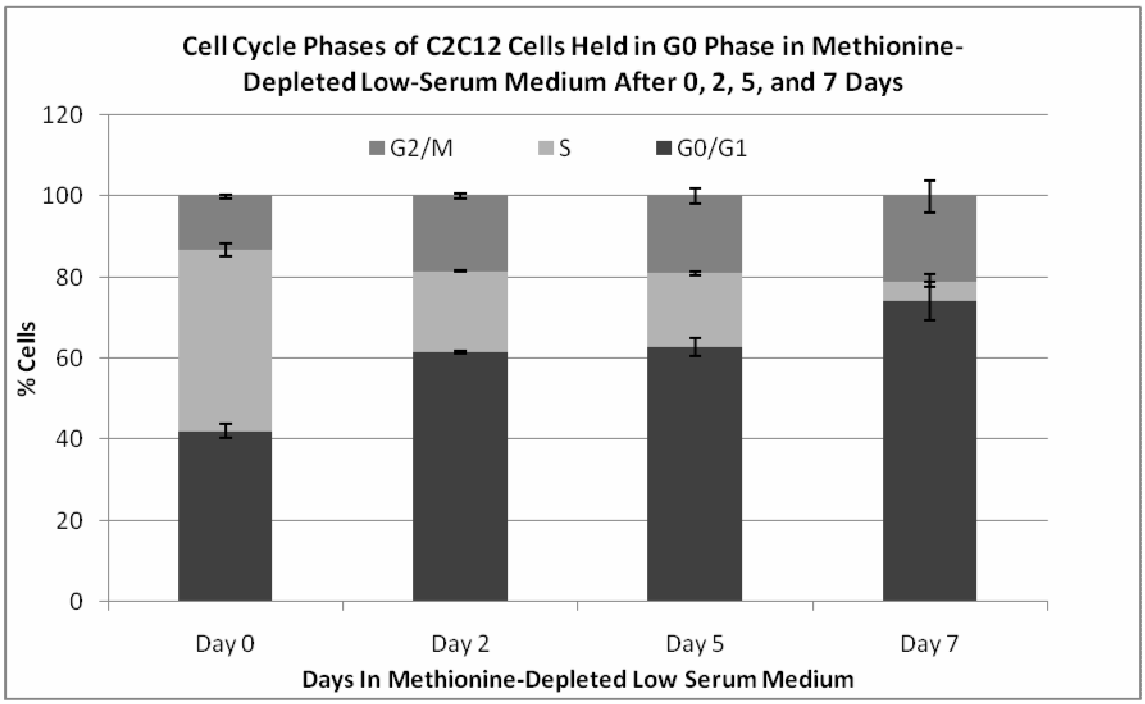


Figure 6.5 Chart of mean percentages ± SEM (error bars) (*n* = 3) of C2C12 cells in G0/G1, S, and G2/M phases after 0, 2, 5, and 7 days under low serum methionine depleted medium. From Day 0 to Day 2 the number of cells in the G0/G1 phase increased, the number of cells in the S phase declined, and the number of cells in the G2/M phase had increased, suggesting that many cells that had previously been in the S phase had passed through the G2/M phase and entered the G0/G1 phase. At Day 5 the number of cells in each phase remained similar to those at Day 2. The number of cells arrested in G0/G1 phase had increased by Day 7 and those in S phase had further declined and passed into the G2/M phase.

6.2.5 Identification Of eGFP Positive Cells Within CeGFP-C1 and CpEPI-eGFP Transfected Cultures Held In G0 For 7 Days By Fluorescence Microscopy

24 hours after C2C12 cells were transfected with CeGFP-C1 or CpEPI-eGFP plasmid the medium was changed from growth medium to methionine-depleted low-serum medium. Cells were kept in this medium for 7 days. Fluorescent images of the cells were taken at Days 1, 3, 5, and 7 and showed that although a large number of cells perished when kept under this condition for an extended period of time, eGFP positive cells were still present within both transfected populations by Day 7 (Figure 6.6).

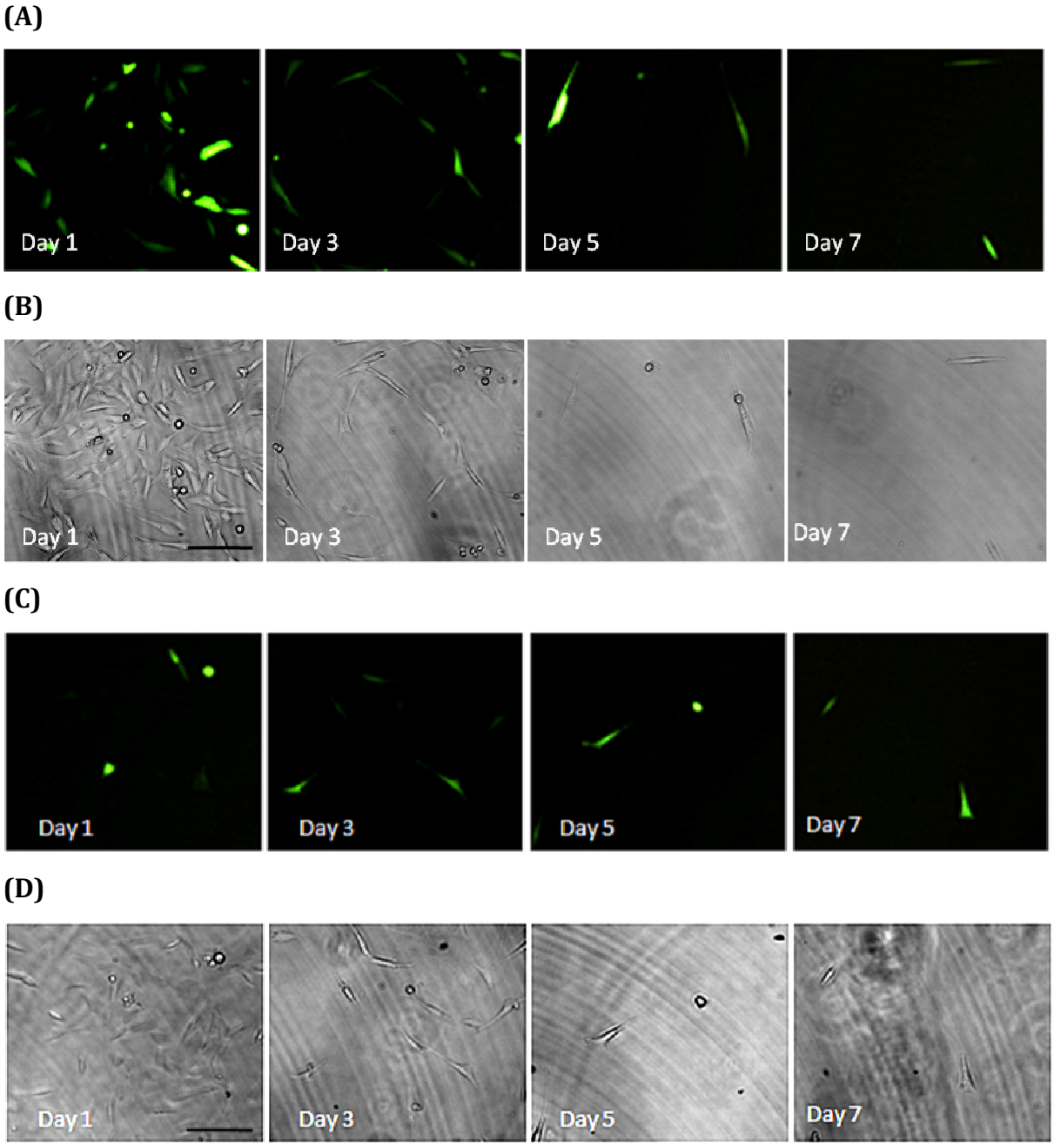


Figure 6.6 Fluorescence images of eGFP positive C2C12 cells transfected with CeGFP-C1 (row A) or CpEPI-eGFP (row C). Day 1 is 24hrs post transfection and the first day of the addition of the low serum methionine depleted medium. The difference can be noted between the number of cells transfected by CeGFP-C1 and the expression levels, both of which were subjectively observed to be greater than those of the CpEPI-eGFP transfected population at Day 1. It is clear that eGFP positive cells were present in both transfected cultures after being put in low serum methionine depleted medium for 7 days. Phase contrast images of cells transfected with CeGFP-C1 (row B) and cells transfected with CpEPI-eGFP (row D) show a decline in the amount of cells present due to cell death following deprivation of methionine in the low serum medium. Bar, 90 µm.

6.2.6 Investigation Of The Presence Of Vector Sequences Within CeGFP-C1 or CpEPI-eGFP Transfected Populations By PCR

PCR analysis was conducted on samples taken on Day 1, Final Day Selection, and Final Day from G0 CeGFP-C1, Reserve CeGFP-C1, G0 CpEPI-eGFP, and Reserve CpEPI-eGFP populations. 50ng of total DNA extracted from each sample was used per PCR reaction, and it was run for 35 cycles. The primers used were designed by Primer3 and lie within the Kanamycin resistance gene present in the transfected vectors. The PCR was run for 35 cycles and the expected product size was 503bp. Products were run on 0.8% agarose gels counterstained with ethidium bromide and visualised under UV light. PCR on all the samples confirmed the presence of CeGFP-C1 or CpEPI-eGFP vectors in their respective transfected G0 and Reserve groups at each time point tested (Figure 6.7)

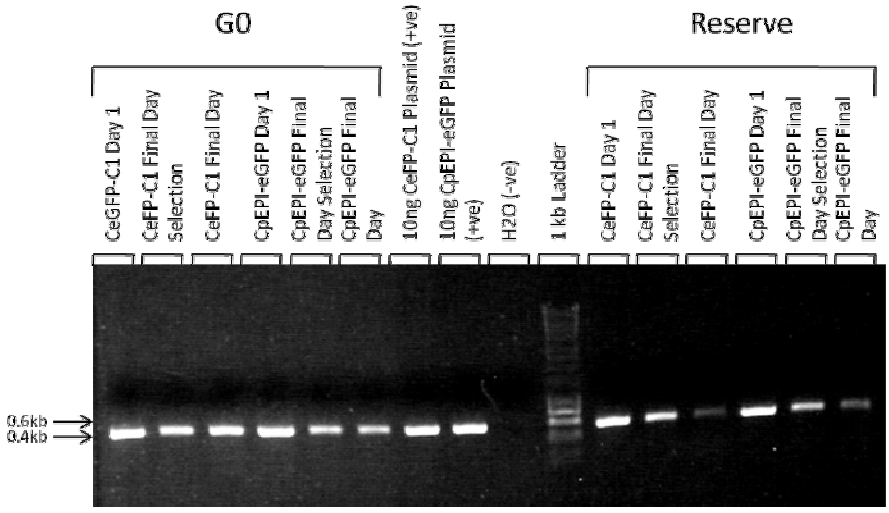


Figure 6.7 PCR analysis conducted on 50ng of total C2C12 DNA transfected with CeGFP-C1 or CpEPI-eGFP plasmid in the G0 and Reserve groups, run on samples taken at Day 1, Final Day Selection, and Final Day. The positive controls were 10ng of CeGFP-C1 and 10ng of CpEPI-eGFP, and the negative controls was H_20. The primers used were within the Kanamycin resistance gene, and product size expected was 503bp. The PCR was run for 35 cycles and the products were run on a 0.8% agarose gel. Copies of the vectors were present within cells of the respective transfected populations in both the G0 and Reserve groups at each time point tested.

6.2.7 Transgene Expression Analysis Of Reserve CeGFP-C1, Reserve CpEPI-eGFP, G0 CeGFP-C1, and G0 CpEPI-eGFP Myoblast Populations By Flow Cytometry

eGFP expression of $1x10^5$ C2C12 myoblasts transfected with CeGFP-C1 or CpEPI-eGFP in the Reserve and G0 groups was analysed by flow cytometry at Day 1 (24 hours post-transfection), final day of selection (after a 10 day selection period post holding in G0 or quiescence in reserve cells), and Final Day (final day of experiment after 39 days of cell proliferation without selection pressure after the period of selection) to determine the mean percentage of positive cells and the mean fluorescence intensity of each population. The percentages, mean fluorescence intensity, and total eGFP ± the standard error of the means (SEM) (*n* = 4) were charted (Figures 6.8-6.10). The total eGFP fluorescence values were converted to log as the data was spread over a wide range. Statistical tests were applied using NCSS statistics program to determine significance. The test used to compare the transfected populations to each other at each time point, and to compare all of the groups at of the time points was One-Way ANOVA. The P-value was set to 0.05 and indicated a significant difference between groups if $P<0.05$. The post-hoc Tukey-Kramer test was used to determine specifically which groups differed significantly from other groups.

From the flow cytometry results obtained of the Day 1 samples it was found that the transfection efficiency and the expression of the CeGFP-C1 plasmid were significantly higher than those of the CpEPI-eGFP plasmid, as the percentage of cells transfected (means of 47.9 ± 0.2% and 7.1 ± 0.2% respectively) (Figure 6.8), the mean expression (mean of 12266.5 ± 310.9 and 2744.3 ± 122.4 respectively) (Figure 6.9), and the total expression (means of 586793.8 ± 12764 and 19490.6 ± 996.3 respectively) (Figure 5.10) of the CeGFP-C1 plasmid were significantly higher (P value = 0.001).

By the final day of selection, where all the surviving cells contained a copy of the plasmid sequence and expressed the Neomycin resistance gene, the percentage of cells that were eGFP positive had decreased significantly in the CeGFP-C1 transfected populations, but not significantly in the CpEPI-eGFP populations (Figure 6.8). An ANOVA test concluded that there was no significant difference between all four groups at the final day of selection in the percentage of cells found to be eGFP positive (means

of 6.2 ± 2.0% for the CeGFP-C1 Reserve population, 3.6 ± 0.3 for the CpEPI-eGFP Reserve population, 8.9 ± 0.4% for the CeGFP-C1 G0 population, and 5.6 ± 1.8% for the CpEPI-eGFP G0 population; P = 0.09).

The mean fluorescence intensities of the transfected populations declined from Day 1 to the final day of selection (Figure 6.9). There was a difference in the mean expressions between the transfected populations at the final day of selection, where the CeGFP-C1 Reserve and G0 populations had similar mean fluorescence intensities but were significantly higher than the mean intensities of the CpEPI-eGFP Reserve and G0 populations, whose intensities were similar to one another (means of 1851 ± 245.1 for the CeGFP-C1 Reserve population, 998 ± 53.1 for the CpEPI-eGFP Reserve population, 2082 ± 200.3 for the CeGFP-C1 G0 population, and 903 ± 93.2 for the CpEPI-eGFP G0 population; P = 0.0001).

After calculating the total expression of each group at the final day of selection, it was found that the total expression of the CeGFP-C1 Reserve and G0 groups did not differ from each other, and the total expression of the CpEPI-eGFP Reserve and G0 groups did not from each other (Figure 6.10) as indicated by the Tukey-Kramer test. However, the CeGFP-C1 G0 group was found to have a total expression significantly greater than both the CpEPI-eGFP Reserve and G0 groups, but not the CeGFP-C1 Reserve group, whose total expression did not differ significantly from any of the groups on the final day of selection (means of 11235.8 ± 3152.7 for the CeGFP-C1 Reserve population, 3557.6 ± 280.2 for the CpEPI-eGFP Reserve population, 18394.2 ± 1444.2 for the CeGFP-C1 G0 population, and 5282.4 ± 2196 for the CpEPI-eGFP G0 population; P value = 0.001).

By the final day of the experiment, the percentage of eGFP positive cells and the mean fluorescence intensities did not decline significantly in the Reserve CeGFP-C1, the Reserve CpEPI-eGFP, or the G0 CpEPI-eGFP populations from Day 21, but did decline significantly in the G0 CeGFP-C1 populations from Day 21. Interestingly, the percentage of eGFP positive cells by the final day of the experiment did not differ significantly between the groups (means of 2.5 ± 2% for the CeGFP-C1 Reserve population, 0.5 ± 0.09% for the CpEPI-eGFP Reserve population, 3 ± 1.5% for the CeGFP-C1 G0 population, and 0.9 ± 0.5% for the CpEPI-eGFP G0 population; P = 0.46), nor did the mean expression (means of 934.3 ± 103.8 for the CeGFP-C1 Reserve population, 977.5 ± 461.9 for the CpEPI-eGFP Reserve population, 911.3 ± 209.8 for

the CeGFP-C1 G0 population, and 729.8 ± 231.6 for the CpEPI-eGFP G0 population; P value = 0.9) or, hence, the total expression (means of 2319.2 ± 1806.4 for the CeGFP-C1 Reserve population, 594.8 ± 354.8 for the CpEPI-eGFP Reserve population, 2734.1 ± 1199.3 for the CeGFP-C1 G0 population, and 933.2 ± 755.9 for the CpEPI-eGFP G0 population; P = 0.5).

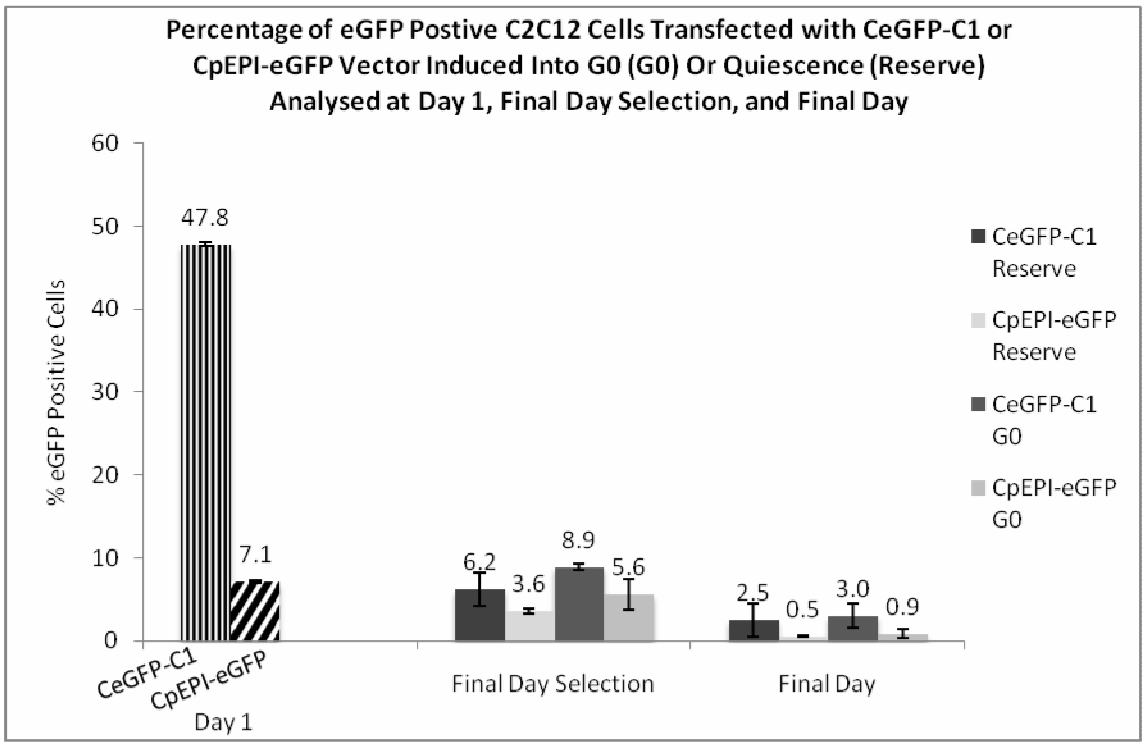


Figure 6.8 Flow cytometry analysis of the mean percentage ± SEM(error bars) (*n* = 4) of eGFP positive C2C12 cells transfected with CeGFP-C1 or CpEPI-eGFP vector. The figures above the bars indicate the values of the mean percentages of positive cells. The Day 1 CeGFP-C1 and CpEPI-eGFP transfected populations were split into two groups each: 'Reserve' and 'G0', and each group was composed of four replicates. 'Reserve' denotes transfected myoblasts that were put under differentiation medium 24 hours post-transfection for 10 days yielding eGFP positive quiescent reserve cells and myotubes. 'G0' denotes transfected myoblasts that were induced into the G0 phase of the cell cycle 24 hours post transfection and were held arrested in G0 for 7 days. After holding the cells in quiescence/G0 the cells were put under G-418 selection for 10 days. Samples were taken and analysed at 24 hours post transfection (Day 1), the final day of selection (Final Day Selection), and the final day of the experiment after proliferation without any selective pressure post-selection (Final Day), where $1x10^5$ cells were analysed per sample. At Day 1 the percentage of CeGFP-C1 transfected eGFP positive cells was markedly higher than that of the CpEPI-eGFP group. By the final day of selection there was no significant difference between any of the four groups. By Day 60 there was a decline in the percentage of eGFP positive cells for the four groups, however none of them differed significantly from one another.

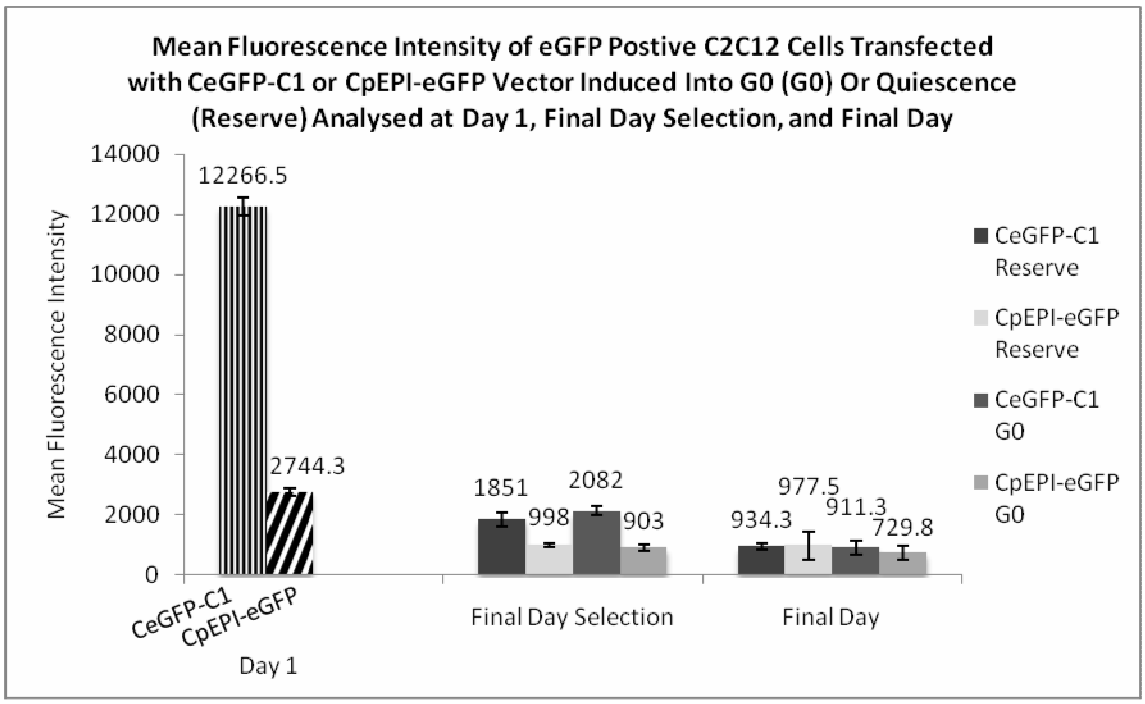


Figure 6.9 Flow cytometry analysis of the mean fluorescence intesity ± SEM(error bars) (*n* = 4) of eGFP positive C2C12 cells transfected with CeGFP-C1 or CpEPI-eGFP vector. The figures above the bars indicate the mean fluorescence intensity values. The Day 1 CeGFP-C1 and CpEPI-eGFP transfected populations were split into two groups each: 'Reserve' and 'G0', and each group was composed of four replicates. 'Reserve' denotes transfected myoblasts that were put under differentiation medium 24 hours post-transfection for 10 days yielding eGFP positive quiescent reserve cells and myotubes. 'G0' denotes transfected myoblasts that were induced into the G0 phase of the cell cycle 24 hours post transfection and were held arrested in G0 for 7 days. After holding the cells in quiescence/G0 the cells were put under G-418 selection for 10 days. Samples were taken and analysed at 24 hours post transfection (Day 1), the final day of selection (Final Day Selection), and the final day of the experiment after proliferation without any selective pressure post-selection (Final Day), where $1x10^5$ cells were analysed per sample. At Day 1 the mean expression of CeGFP-C1 transfected eGFP positive cells was markedly higher than that of the CpEPI-eGFP group. By the final day of selection, the CeGFP-C1 Reserve and G0 populations had similar mean fluorescence intensities but were significantly higher than the mean intensities of the CpEPI-eGFP Reserve and G0 populations, whose intensities were similar to one another. By Day 60 there was no significant difference between any of the groups in terms of mean expression.

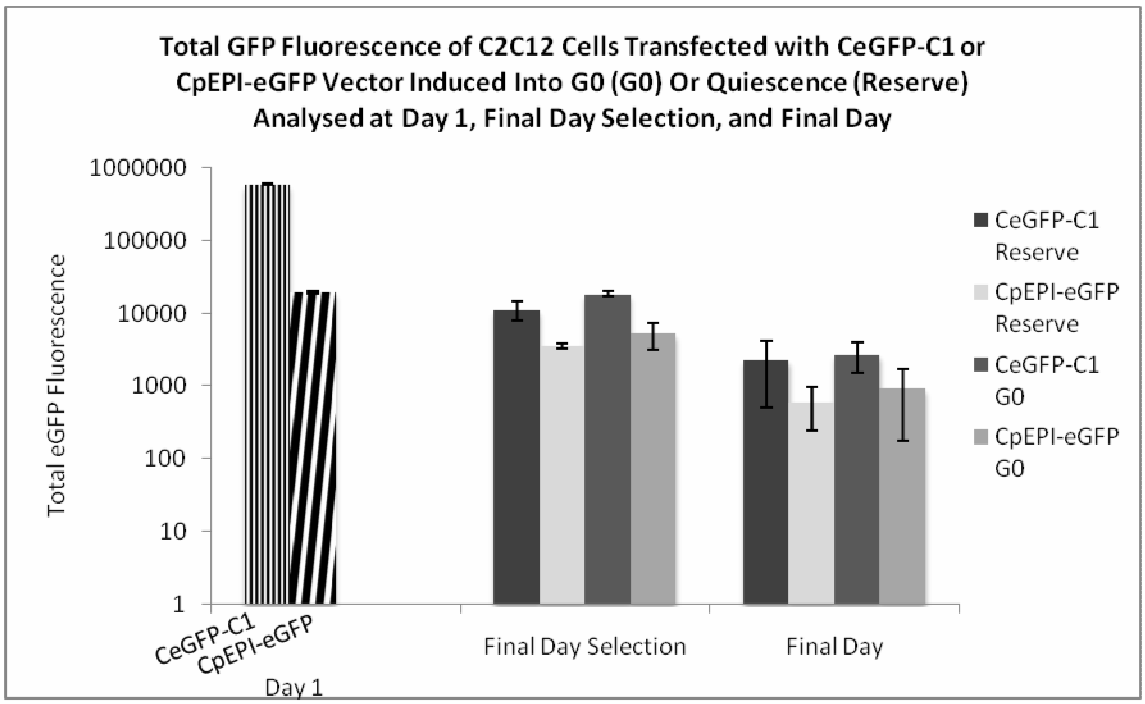


Figure 6.10 The total eGFP fluorescence ± SEM (error bars) (*n* = 4) of eGFP positive C2C12 cells transfected with CeGFP-C1 or CpEPI-eGFP vector. The figures above the bars indicate the mean total eGFP fluorescence. The Day 1 CeGFP-C1 and CpEPI-eGFP transfected populations were split into two groups each: 'Reserve' and 'G0', and each group was composed of four replicates. 'Reserve' denotes transfected myoblasts that were put under differentiation medium 24 hours post-transfection for 10 days yielding eGFP positive quiescent reserve cells and myotubes. 'G0' denotes transfected myoblasts that were induced into the G0 phase of the cell cycle 24 hours post transfection and were held arrested in G0 for 7 days. After holding the cells in quiescence/G0 the cells were put under G-418 selection for 10 days. Samples were taken and analysed at 24 hours post transfection (Day 1), the final day of selection (Final Day Selection), and the final day of the experiment after proliferation without any selective pressure post-selection (Final Day), where $1x10^5$ cells were analysed per sample. The total eGFP fluorescence values were converted to log as the data was spread over a wide range. At Day 1 the total expression of CeGFP-C1 eGFP positive cells was markedly higher than that of the CpEPI-eGFP group. By the final day of selection the groups of cells transfected with CeGFP-C1 (the Reserve group and the G0 group) did not differ from each other in their mean expressions, nor did the groups of cells transfected with CpEPI-eGFP (the Reserve group and the G0 group) differ from one other. However, the CeGFP-C1 G0 group expressed a significantly larger amount of eGFP than the CpEPI-eGFP Reserve and G0 populations. The CeGFP-C1 Reserve population had a mean intensity that did not differ significantly from any of the groups. By Day 60 none of the groups differed significantly in terms of total expression.

6.2.8 Expression Analysis of Reserve CeGFP-C1, Reserve CpEPI-eGFP, G0 CeGFP-C1, and G0 CpEPI-eGFP Myotube Populations By Immunostaining

The differentiation of transfected myoblasts into myotubes may have an effect upon transgene expression from the CeGFP-C1 and CpEPI-eGFP plasmids. $5x10^4$ cells were isolated from each transfected population at Day 1 (24 hours post-transfection), Final Day Selection (final day of a 10 day selection period post-holding in G0/quiescence in reserve cells), and Final Day (final day of experiment after 39 days of cell proliferation without selection pressure after the 10 day period of selection) and seeded into $1cm^2$ chamber wells. The cells were differentiated by changing the medium to differentiation medium for 5 days. The cells were then fixed in 4% PFA/PBS and immunostained for eGFP.

The results of the staining showed that on the final day of selection and at Final Day, eGFP was detected in the G0 CeGFP-C1, G0 CpEPI-eGFP, and Reserve CpEPI-eGFP transfected populations (Figure 6.11), but none was found in the Reserve CeGFP-C1 transfected population. The images suggest that the amount of eGFP positive cells in the populations which contained positive cells, and their fluorescence intensities, increased by the final day of selection in comparison to Day 1. The images also suggest that the amount of eGFP positive myotubes and their brightness were similar in the G0 CeGFP-C1 and G0 CpEPI-eGFP populations, and both were greater in these two populations than in the Reserve CpEPI-eGFP population. These observations may indicate that holding C2C12 cells in G0 for 7 days post transfection may lead to improved and long-term transgene expression over holding the cells in quiescence in reserve cells.

It can also be seen from the images that the CeGFP-C1 and CpEPI-eGFP transfected cells in the G0 group did not fuse together during differentiation. The G0 CpEPI-eGFP population contained some myotubes at the final day of selection time point, however the G0 CeGFP-C1 population at the same time point did not, nor did either population at the Final Day time point.

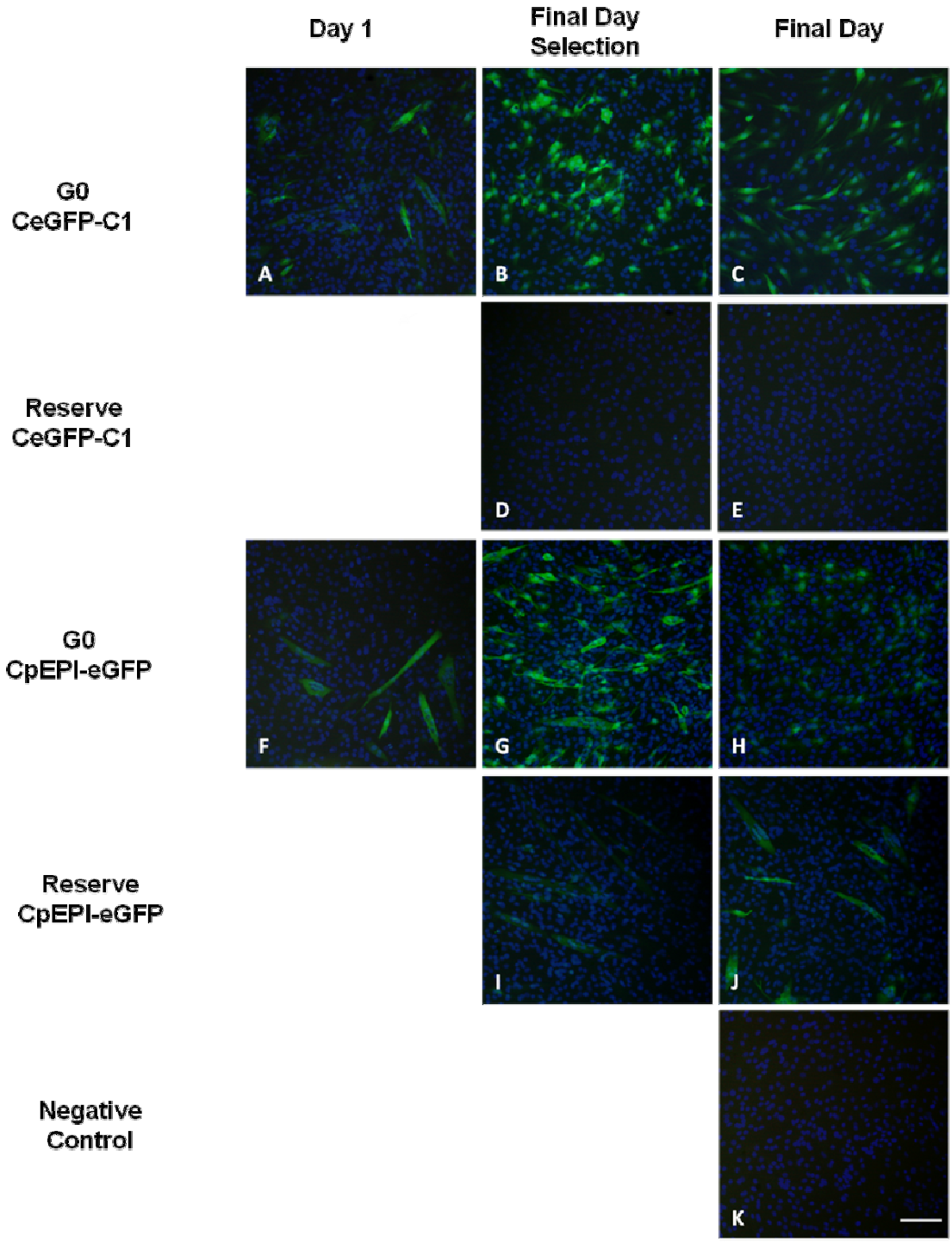


Figure 6.11 eGFP immunostaining of fixed C2C12 myoblasts transfected with CeGFP-C1 or CpEPI-eGFP vector and differentiated into myotubes under differentiation medium for 5 days. Samples from the transfected populations were differentiated at three time points: Days 1, Final Day Selection, and Final Day. Day 1 indicates the CeGFP-C1 and CpEPI-eGFP transfected populations that were put into differentiation medium 24 hours post transfection (images A and F respectively). Final Day Selection indicates G0 CeGFP-C1, Reserve eGFP-C1, G0 pEPI-eGFP, and Reserve pEPI-eGFP populations put under differentiation medium on the final day of selection after a 10 day selection period post-holding in G0/quiescence in Reserve cells (images B, D, G, and I, respectively). Final Day indicates G0 CeGFP-C1,

Reserve eGFP-C1, G0 pEPI-eGFP, and Reserve pEPI-eGFP populations put into differentiation medium on the 39th day of proliferation without any selective pressure, post-selection (images C, E, H, and J, respectively). The negative control was untransfected C2C12 myotubes (image K). Green indicates eGFP positive cells and blue is the DAPI staining of the nuclei. The images show that on the final day of selection and at Final Day, eGFP positive myoblasts/myotubes were detected in the G0 CeGFP-C1, G0 CpEPI-eGFP, and Reserve CpEPI-eGFP transfected populations, but none were found in the Reserve CeGFP-C1 transfected population. A subjective observation is that the amount of eGFP positive cells in the populations that contained them, and their fluorescence intensities, increased in final day of selection in comparison to Day 1. Another observation is that the amount of eGFP positive myotubes and their brightness were similar in the G0 CeGFP-C1 and G0 CpEPI-eGFP populations, and both were greater in these two populations than in the Reserve CpEPI-eGFP population. It can also be seen from the images that the CeGFP-C1 and CpEPI-eGFP transfected cells in the G0 group did not fuse together during differentiation. The G0 CpEPI-eGFP population contained some myotubes at the final day of selection time point, however the G0 CeGFP-C1 population at the same time point did not, nor did either population at the Final Day time point. Bar, 90µm.

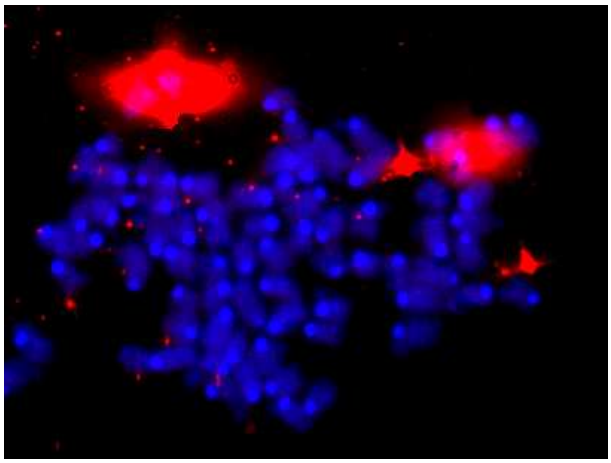
6.2.9 Investigation Of Episomal/Integrant Status Of CpEPI-eGFP and CeGFP-C1 Vectors In Transfected C2C12 Cells Of The Reserve And G0 Groups By Fluorescent In-Situ Hybridisation (FISH) Analysis On Final Day Samples

FISH was used to determine the episomal and/or integrant status of the transfected vectors (as conducted in Chapters 3 & 4). The positive controls used in this experiment were C2C12 cells 24 hours post transfection with CpEPI-eGFP or CeGFP-C1 (Figure 6.12 (A) and (B), respectively). The negative controls in this experiment were untransfected C2C12 cells (Figure 6.12 (C)). From the images taken under the fluorescence microscope it can be seen that a large amount of plasmid was present within the positive control cells, as the fluorescence signal emitted was very strong. It was also clear that much of the plasmid observed was clustered together.

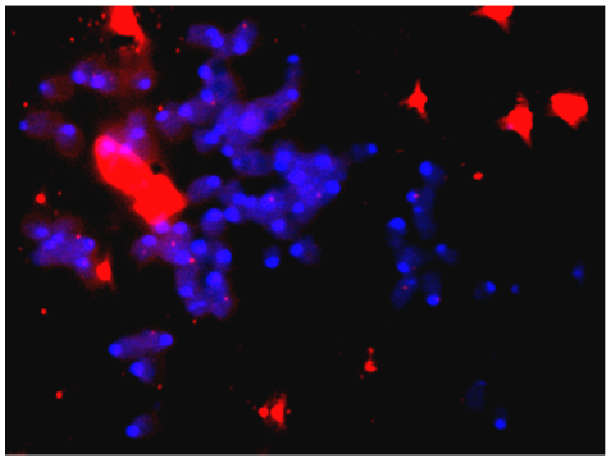
The FISH images indicated that the CpEPI-eGFP plasmid in the Reserve group may have integrated into the host genome (4 integrant spreads of 9 positive spreads) (Figure 6.13 (A)), as was the evidence shown for the CeGFP-C1 plasmid of the Reserve group (3 out of 6 integrant spreads) (Figure 6.13 (B)), in addition to episomal copies still being present within both populations. It is interesting to note that evidence of integration of an S/MAR-containing plasmid has been found, suggesting that holding the transfected reserve cells in quiescence and then putting them under selection for 10 days lead to its potential integration.

The G0 CpEPI-eGFP and CeGFP-C1 samples analysed showed that CpEPI-eGFP and CeGFP-C1 were retained as episomes without any evidence of integration into the host genome (0/6 and 0/5 integrant spreads, respectively) (Figure 6.14 (A) and (B)).

(A) C2C12 cells 24 hr after transfection with CpEPI-eGFP



(B) C2C12 cells 24 hr after transfection with CeGFP-C1



(C)C2C12 cells untransfected

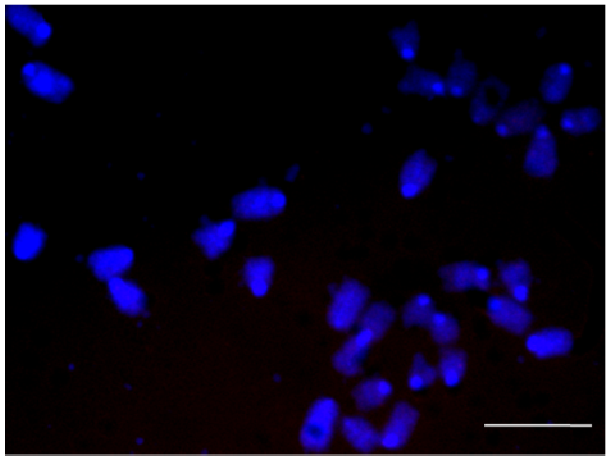
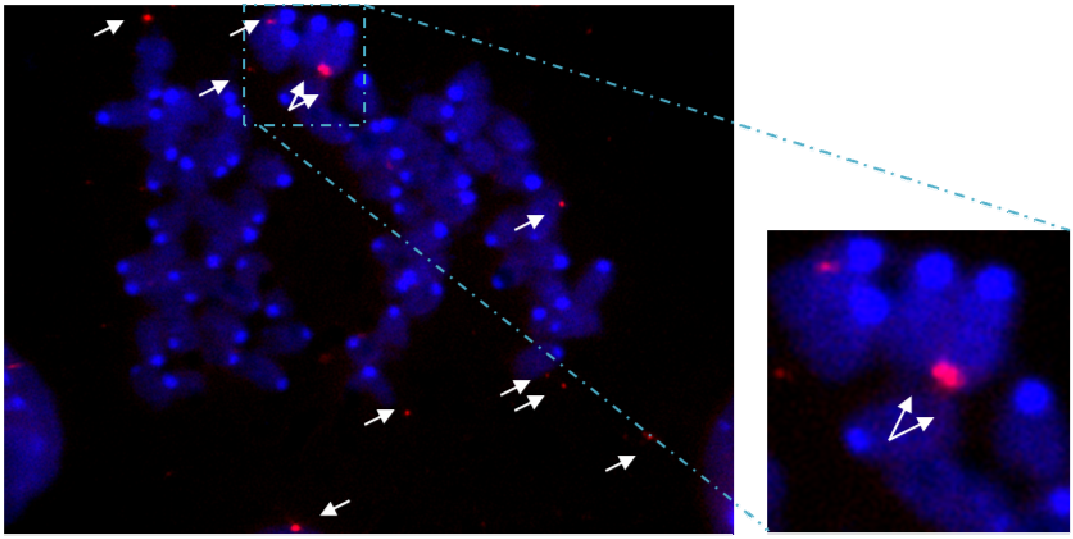


Figure 6.12 Fluorescent In-Situ Hybridisation positive control of C2C12 cells 24 hours post transfection with the CpEPI-eGFP plasmid (A) and CeGFP-C1 plasmid (B). The plasmid can be seen as the bright red fluorescing red spots. Genomic DNA was counterstained with Dapi (blue). As can be seen there was a large amount of vector present within the transfected cells, and the plasmid was found mostly clustered together. The negative control was untransfected C2C12 cells (C) indicating no background. Bar, 5µm.

Transfected C2C12 cells Final Day of experiment

(A)Reserve CpEPI-eGFP



(B)Reserve CeGFP-C1

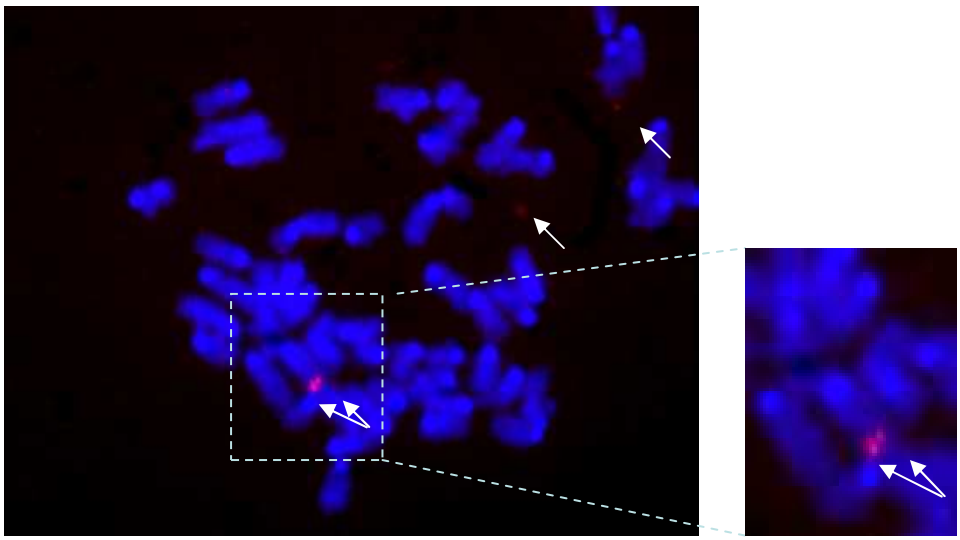
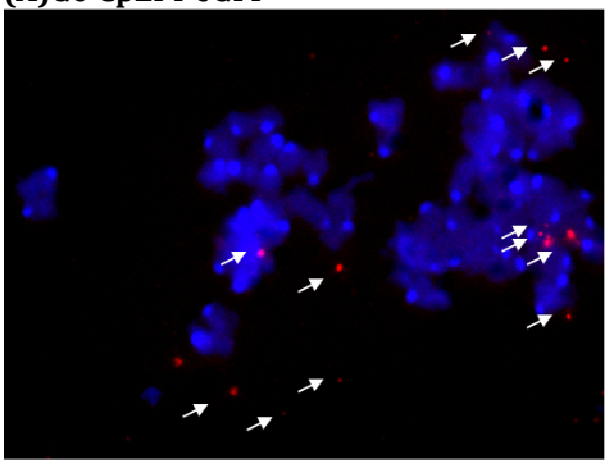


Figure 6.13 Fluorescent In-Situ Hybridisation of C2C12 Reserve CpEPI-eGFP sample (A) and Reserve CeGFP-C1 (B) sample taken on the final day of the experiment. Single white arrows indicate red fluorescence spots where the probe has hybridised to plasmid sequence and denotes the episomal status of the plasmid. The double arrows mark spots of potential integration. Genomic DNA was counterstained with Dapi (blue). From these images it can be seen that CpEPI-eGFP and CeGFP-C1 could have integrated into the C2C12 host genome (4/9 and 3/6 integrant spreads, respectively) in addition to being retained as an episome by Day 60.

(A)G0 CpEPI-eGFP



(B)G0 CeGFP-C1

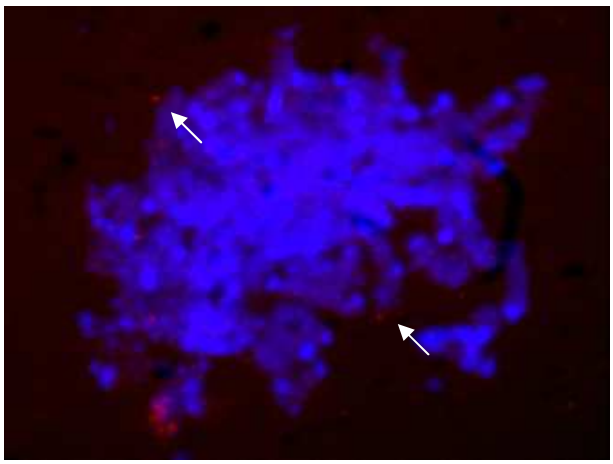


Figure 6.14 Fluorescent In-Situ Hybridisation of C2C12 G0 CpEPI-eGFP sample (A) and G0 CeGFP-C1 (B) sample taken on the final day of the experiment. Single white arrows indicate red fluorescence spots where the probe has hybridised to plasmid sequence and denotes the episomal status of the plasmid. Genomic DNA was counterstained with Dapi (blue). No evidence of integration was found of CpEPI-eGFP and CeGFP-C1 into the C2C12 host genome (0/6 and 0/5 integrant spreads, respectively), and vector copies in both populations were found present in episomal form.

6.2.10 Summary of Results:

		PCR	Flow Cytometry			Immuno-staining of Myotubes	FISH	
			% positive cells	MFI	Total		Integra-tion	Epi-somal
Vector	**Day**							
CeGFP-C1 Reserve	**1**	√	47.9 ± 0.2	12266.5 ± 310.9	586793.8 ± 12764	++	3/6	√
	Final Day Selection	√	6.2 ± 2.0	1851 ± 245.1	11235.8 ± 3152.7	-		
	Final Day Exp	√	2.5 ± 2%	934.3 ± 103.8	2319.2 ± 1806.4	-		
CpEPI-eGFP Reserve	**1**	√	7.1 ± 0.2	2744.3 ± 122.4	19490.6 ± 996.3	+	4/9	√
	Final Day Selection	√	3.6 ± 0.3	998 ± 53.1	3557.6 ± 280.2	+		
	Final Day Exp	√	0.5 ± 0.09	977.5 ± 461.9	594.8 ± 354.8	++		
CeGFP-C1 G0	**1**	√	47.9 ± 0.2	12266.5 ± 310.9	586793.8 ± 12764	++	0/5	√
	Final Day Selection	√	8.9 ± 0.4	2082 ± 200.3	18394.2 ± 1444.2	++		
	Final Day Exp	√	3 ± 1.5	911.3 ± 209.8	2734.1 ± 1199.3	++		
CpEPI-eGFP G0	**1**	√	7.1 ± 0.2	2744.3 ± 122.4	19490.6 ± 996.3	+	0/6	√
	Final Day Selection	√	5.6 ± 1.8	903 ± 93.2	5282.4 ± 2196	++		
	Final Day Exp	√	0.9 ± 0.5	729.8 ± 231.6	933.2 ± 755.9	++		

Table 6.2 Table summarising the results obtained in this chapter for transfected C2C12 cells, indicating each vector that was transfected and the time point at which samples were extracted and analysed. PCR results are presented where a 'v' denotes that the Kanamycin sequence was amplified and a band was present; Flow Cytometry results presented include the percentage of positive cells per transfected population, the mean fluorescence intensity (MFI) of the positive cells, and the total amount of expression which was the product of expression and number of positive cells; the Immunostaining data is presented by '-' indicating the absence of positive myotubes, '+' indicating the presence of less than 10 positive myotubes, and '++' indicating the presence of more that 10 positive myotubes; the FISH data is presented as the number of slides which suggested an integration event having occurred out of the total number of slides analysed on Day 60 only, a 'v' in the episomal column indicated the presence of episomal copies of the vector observed on the slides, and a '-' indicates none was observed.

6.3. Discussion:

6.3.1 Transfection efficiency of CeGFP-C1 was superior to that of CpEPI-eGFP in C2C12 cells

As was demonstrated by the percentage of positive cells present in each transfected population on Day 1 by flow cytometry, the ability of CeGFP-C1 to transfect C2C12 cells was greater than that of CpEPI-eGFP. This was also observed in the previous results chapter (Chapter 4) and was discussed in the Chapter 4 Discussion (Section 4.3.2).

6.3.2 Transgene expression resulting from CeGFP-C1 was superior to that from CpEPIeGFP at Days 1 and Final Day Selection

The mean fluorescence intensities of the CeGFP-C1 transfected populations were found to be significantly greater at Days 1 and Final Day Selection to those of the CpEPI-eGFP transfected populations, as shown by flow cytometry. This difference was also found in the previous results chapter (Chapter 4), where several explanations for this finding included the difference in size between the two vectors, the degrees of positive/negative supercoiling of each vector due to the presence/absence of the S/MAR element, and the potential effects, due to the existence of several polyadenylation signals within the S/MAR element, on mRNA transcript stability of the transgene as expressed from CpEPI-eGFP. For a full discussion of these issues please refer to Chapter 4: Discussion (Section 4.3.2).

6.3.3 Percentage of positive cells and transgene expression was comparable between all groups of transfected myoblasts by the final day of the experiment

Despite the fact that the vectors were held in transfected cells using two different methods, and the fact that one vector contained an S/MAR element and the other did not, it was found by flow cytometry that there was no significant difference in the percentage of positive cells or mean fluorescence intensity between the transfected

populations by the final day of the experiment. This result had also occurred despite the fact that CeGFP-C1 and CpEPI-eGFP were shown to potentially have integrated into the host genome in the Reserve group, whereas no evidence of integration had been apparent in the G0 group. This shows two things. First, that holding the cells in G0 may have had a positive effect on vector retention and in the prevention of vector integration. This is especially significant considering that the percentage of positive cells and the levels of expression of both CeGFP-C1 and CpEPI-eGFP were not significantly different to those in the populations where CeGFP-C1 and CpEPI-eGFP appeared to have integrated into the host genome. And second, that there was a difference between holding cells in quiescence in reserve cells to holding cells in G0 using methionine depleted low serum media, as in one evidence that the vectors had integrated was present whereas in the other there was not.

6.3.4. Differences Exist In Vector Status Between Cells Held In G0 And Those Held In Quiescence In Reserve Cells Post Transfection With CeGFP-C1 and CpEPI-eGFP

One of the aims of this chapter was to determine whether expression and vector status resulting from holding transfected cells in G0 by the use of low serum methionine-depleted medium was different to that obtained by holding the cells in quiescence in reserve cells. The flow cytometry data and the statistical analyses suggested that there was no significant difference between any of the four groups in the mean percentage of positive cells or the mean fluorescence intensity by the final day of the experiment. However, the FISH data suggest that differences may exist, when considering the finding that potential integration of both the CeGFP-C1 and CpEPI-eGFP vectors into the host genome was found when the transfected cells were held in quiescence in reserve cells, whereas no evidence of integration of either vector was observed in the transfected populations held in G0. It is difficult to consider any reasons why both vectors would integrate in the Reserve group but not in the G0 group, as there is no experimental evidence in this chapter or in the literature to confirm that C2C12 myoblasts induced into G0 are different to quiescent reserve cells, besides the experimental manner in which they were induced (differentiation of a culture in low serum media as opposed to culturing in low serum methionine depleted media).

6.3.5 Evidence of a vector containing the β-IFN S/MAR element integrated in the C2C12 host genome

A result which had not been noted in previous experiments in this thesis, nor in many S/MAR studies (as reviewed in Chapter1: Introduction, section 1.10), was the evidence pointing towards the integration of a vector containing a β-IFN S/MAR element within the open reading frame of a transgene, CpEPI-eGFP in this case, into the host genome. The cause of the possible integration is unclear, however it shows that holding transfected cells in quiescence as reserve cells by differentiation of a culture, followed by 10 days of selection, could have lead to the integration of the vector into the host genome. If the manipulation of primary myoblasts *ex vivo* prior to transplantation into a patient were being considered, this method of manipulation may not be considered suitable due to the safety implications associated with integration.

6.3.6 A Proportion Of Cells Within The 'Reserve' And 'G0' Populations Had Not Been Arrested

It was found by lysenin staining that not all mononucleated cells that were isolated after differentiating the CeGFP-C1 and CpEPI-eGFP transfected C2C12 cells were reserve cells, as not all of them had stained positive for lysenin. This indicates that there was a possibility that within the isolated reserve cell population, a proportion of eGFP positive cells had not been induced into quiescence to become reserve cells, which are all lysenin positive/CD34 positive/MyoD negative (Nagata *et al*., 2006; Beauchamp *et al*., 2000).

Nagata *et al* showed that in a C2C12 differentiating culture, besides the presence of quiescent reserve cells and myotubes, there were also some mononucleated differentiated cells, as indicated by negative staining for lysenin and positive staining for myogenin, which is expressed in differentiated cells, and not in myoblasts or reserve cells (Nagata *et al*., 2006). These cells lose their ability to proliferate once they have differentiated, and are diluted out of the population by continuous passaging. Therefore, any cells of this type that was isolated with the reserve cells within the transfected culture would have been lost and would not be considered to have influenced the outcome of the experiment.

However, there also existed a subset of cells which were still cycling, and were not quiescent. In the study by Nagata *et al* it was found that a proportion of cells which were isolated with the reserve cells incorporated BrdU even after 4 days in differentiation medium and did not stain positive for lysenin (Nagata *et al*., 2006). These cells would be undifferentiated myoblasts which would stain negative for lysenin and CD34, where some would stain positive for MyoD and others would be MyoD negative (Beauchamp *et al*, 2000). These cells, had they been transfected with CeGFP-C1 or pEPI-eGFP and isolated with the reserve cell population, would have survived selection and comprised a proportion of the populations that were being passaged and investigated in this study. This indicates that the results obtained do not represent those derived of a pure culture of cells that were arrested in quiescence in reserve cells post-transfection.

It was also found that not all the cells of the CeGFP-C1 and CpEPI-eGFP transfected populations that were put in methionine-depleted medium were arrested in G0, as ~5% were still in S phase and ~25% in G2/M phase of the cell cycle at Day 7. Even though it was shown that a majority of the transfected cells had been arrested for several days in G0/G1, this indicates that eGFP positive cells that had not been arrested in G0/G1 may have also survived selection and had been present in the G0 CeGFP-C1 and G0 CpEPI-eGFP populations. This indicates that the results obtained, as in the reserve group, also do not represent those derived of a pure culture of cells that were arrested in G0 in myoblasts post-transfection.

6.3.7. C2C12 myoblasts can lose their abilty to fuse into myotubes

It can be noted from the eGFP immunostaining images of the cells that were put under low serum medium in order to differentiate (Section 5.2.8: Figure 5.11) that some of the transfected populations did not form myotubes, as in the samples taken at the final day of selection from the G0 CpEPI-eGFP and G0 CeGFP-C1 populations and the Day 60 sample from the Reserve CpEPI-eGFP population, and others only formed a few, as in the sample taken at the final day of selection from the G0 CpEPI-eGFP population. A differentiating culture contains several types of cells such as reserve cells, myotubes, cycling myoblasts, and mononucleated differentiated cells (Nagata *et al*., 2006). When a culture is induced to differentiate but does not form myotubes it is

possible that some cells in the culture had initiated differentiation, if the culture had become confluent at any point, which can occur in situations where cells are passaged for an extended period of time. This confluency can occur as a result of overgrowth before passaging, or due to cell clumping and/or uneven spreading of the cells in a flask where one region is more populated and thus more confluent than another region of the same flask. When this occurs and differentiation is initiated in some cells, these cells lose their ability to proliferate or to form myotubes. These cells are considered to be differentiated myoblasts, identified by negative staining for lysenin and positive staining for myogenin (Nagata *et al*., 2006).

Another subset of these cells retain the ability to proliferate but lose the ability to differentiate. These cells remain slowly cycling in a differentiated culture and identified as undifferentiated myoblasts by staining lysenin negative/CD34 negative/MyoD negative or positive (Beauchamp *et al*, 2000). Therefore, a culture that is put under low serum medium and contains a few myotubes but many mononucleated cells may contain reserve cells, mononucleated differentiated cells, and undifferentiated cells. This can explain the variability in the differentiation capacities of the transfected cultures observed.

In this case it was difficult to assess the expression of the transgene of the transfected vectors in myotubes, as the different populations had differentiated in varying, unquantified, and undetermined amounts, and it was uncertain whether the eGFP positive cells were positive due to an accumulation of the protein in cells that were cycling more slowly due to the confluency of the culture, or due to differentiation. A method useful for the determination of total eGFP expression in this case would be Quantitative Reverse-Transcriptase Polymerase Chain Reaction (Q-RT-PCR), where cDNA of eGFP mRNA extracted from each sample could be made, which would then be quantified against a standard curve, hence determining the total amount of eGFP mRNA transcripts per sample. Other methods include Western Blotting or Enzyme-Linked Immunosorbent Assay (ELISA) using anti-eGFP antibodies, which allow the quantification of total eGFP protein in each sample.

6.3.8 Populations of myoblasts that were differentiated showed eGFP positive myoblasts/myotubes were present from Day 1 to Final Day in G0 CeGFP-C1, G0 CpEPI-eGFP, and Reserve CpEPI-eGFP, but not in Reserve CeGFP-C1, populations

In differentiating the transfected cells of each group, it was found that positive cells/myotubes were found up to the final day of the experiment in all groups except the CeGFP-C1 transfected population of the Reserve group, as indicated by immunostaining (Section 5.2.8: Figure 5.11). In this group, the cells/myotubes were eGFP-negative even at the final day of selection. It was not possible to deduce the reason for the absence of eGFP within this population from the experiments conducted in this study. However, considering the fact that the flow cytometry data indicated that the transfected myoblasts of all groups contained eGFP positive cells, it is possible that this lack of expression after differentiation of the culture may be related to differentiation. It is possible that CeGFP-C1 in the Reserve group had preferentially integrated into regions of the C2C12 genome which stopped being expressed upon differentiation, thus leading to a decrease in eGFP transgene expression resulting in eGFP positive cells/myotubes not being detected by immunostaining. This may be a plausible explanation considering the study by Deato and Tijan (Deato and Tijan, 2007) where a decrease was found in the amount of the transcription factor TFIID upon differentiation of C2C12 myoblasts, leading to the shutting off of genes responsible for proliferation, and an increase in a different core promoter complex, resulting in different genes being switched on and transcribed for differentiation to occur (Deato and Tijan, 2007). Therefore it is possible that this change in transcription factor complexes was responsible for the silencing of the transgene upon differentiation. Although episomal copies were found to be present within this population by Final Day, it may be that the level of transgene expression resulting from these copies was not at a sufficient level to be detected by immunostaining.

Nonetheless, the finding of eGFP positive differentiated myoblasts/myotubes within the other three populations was considerable in light of the fact that this had not been observed in the previous experiments conducted in this thesis. It also indicated that gene expression had not been silenced, and that the CAGG promoter was able to drive long-term transgene expression within C2C12 cells.

6.3.9 Holding C2C12 cells in G0/G1 for up to 7 day prior to selection may prevent vector integration into the host genome

It was interesting to note that in the CeGFP-C1 and CpEPI-eGFP transfected populations of the G0/G1 group no evidence of integration had been detected, indicating that holding the cells in G0/G1 may prevent vector integration into the host genome. This is reminiscent of the lack of vector integration into the HepG2 cells observed in Chapter 3 of this thesis. Further investigation would be required to investigate the mechanisms by which this may have occurred in future studies.

6.3.10 The presence of the S/MAR element within CpEPI-eGFP led to superior episomal vector retention and passing on to daughter cells within C2C12 myoblasts than CeGFP-C1

The transfection efficiency of CeGFP-C1 in the experiment was shown to have been ~6 fold greater than that of CpEPI-eGFP. It was also shown that the decline in the percentage of positive cells with time in the CpEPI-eGFP population was not significant, whereas it was in the CeGFP-C1 population. This indicated that, although the percentage of positive cells in the CeGFP-C1 and CpEPI-eGFP transfected populations were similar by the final day, that the loss of CpEPI-eGFP episomal copies with cell division was less than that of CeGFP-C1, leading to the conclusion that CpEPI-eGFP was retained and passed on to daughter cells more effectively by C2C12 cells than CeGFP-C1. This can be attributed to the the presence of the S/MAR element within the vector, considering that the only difference between the sequences of the two vectors was the presence S/MAR element within CpEPI-eGFP. This can only be said for the transfected cells in the G0/G1 group where no evidence of integration into the host genome had been found to occur and transgene expression was, if not completely then predominantly, from episomal vector copies. As possible integration had been observed by both vectors in the Reserve group, where transgene expression may have been a result of a combination of episomal and integrant vector copies of unknown proportions of each, no conclusions could be drawn on the aspect of efficiency of vector retention within the populations of this group without further investigations.

Chapter 7: The Assessment and Comparison of The Requirement/Efficacy of Several Vector Establishment Methods In The C2C12 *in vitro* Model of Muscle Cells In Leading to Long Term Episomal Transgene Expression by CeGFP-C1, CpEPI-eGFP, and Novel S/MAR vector CMYC-pEPI

7.1 Introduction

In the previous chapter C2C12 cells were held in the G0G1 phase of the cell cycle for up to 7 days post transfection with the CeGFP-C1 and CpEPI-eGFP plasmids, and long-term expression, with no evidence of integration of the vectors, was achieved. It then became imperative to determine whether this was attributed to holding the cells in G0/G1 post transfection, selection, or the combination of the two, as safe, long-term transgene expression from episomal plasmid vectors resulting without the requirement for selection would make ideal gene therapy vectors. Therefore, this was tested in this chapter.

The C2C12 cell line was a suitable *in vitro* model for this investigation, as the cycling myoblasts represented activated satellite cells and cycling primary myoblasts (Kitzmann *et al*., 1998; Yoshida *et al*., 1998), where holding the cells in G0/G1 represented the proportion of cells that had re-entered quiescence, as is likely to occur *in vivo*. Arresting cycling C2C12 cells after transfection in quiescence or G0/G1 for a period of time before allowing them to proliferate for many generations would also be an *in vitro* representation of the sequence of events that can occur after plasmid delivery into muscle *in vivo*. *In vivo*, if the plasmid delivered was to transfect cycling satellite cells, induced into activation due to muscle damage to replenish the muscle cell and fibre populations, then a proportion of these cells would return to resting in G0, or quiescence, for a period of time before re-activation and replication, again triggered by muscle injury. Considering that muscle damage occurs frequently in patients with Muscular Dystrophy, the ability to show that a potential gene therapy vector carrying a transgene can divide and maintain an adequate copy number per cell to ensure sustained, long-term, and sufficient levels of transgene expression within dividing muscle cells is essential.

A second aim of this chapter was to assess the efficiency with which cycling myoblasts transfected with CeGFP-C1 or CpEPI-eGFP achieved long term transgene expression without being held in G0 post-transfection or the use of selection, and to determine whether a difference would be found between CpEPI-eGFP and the control vector CeGFP-C1. This would be an *in vitro* representation of the proportion of cycling myoblasts transfected with vector which would not re-eneter quiescence and divide many times before terminal differentiation and the formation of myofibres in order to replenish the population in response to damaged fibres where, *in vivo*, selection is not applied.

As a control in the experiment, and in order to determine whether any significant improvement was achieved by the other methods tested in comparison to this more conventional method of establishing a vector within a cell line, CeGFP-C1 and CpEPI-eGFP transfected cells were placed under selection for ten days post-transfection, before allowing proliferation without selection.

One further aim of this study was to test a novel vector containing the murine c-MYC S/MAR. This S/MAR element was identified by the studies conducted by Girard-Reydet and Cai *et al* (Girard-Reydet *et al*., 2004; Cai *et al*., 2003). The *c-myc* oncogene has the ability to amplify and exist as a double minute (as described in Chapter 1: Introduction section 1.11.1). This may be related to the c-MYC's S/MAR element that allows the propagation of this amplified circular piece of DNA that needs neither a centromere nor telomeres to replicate within tumour cells. As it has been seen that S/MAR elements may act as origins of replication themselves, perhaps it is possible that the S/MAR is the element that is maintaining the over-amplified *c-myc* region as a double minute and allowing expression. These characteristics made the c-MYC S/MAR an interesting alternative S/MAR to test in replicating cells. Thus this novel S/MAR plasmid was created, identical to CpEPI-eGFP in sequence, except for the S/MAR elements' sequences (plasmid map in Chapter 2: Materials & Methods), and will herein be referred to as 'CMYC-pEPI'. In this study, CMYC-pEPI was transfected into C2C12 cells, and expression and vector status of the transfected populations were compared to those of CeGFP-C1 and CpEPI-eGFP transfected populations.

The analyses conducted on the cell samples derived included PCR to confirm the presence of the vector sequences within each population. Flow cytometry allowed the assessment of the expression of the eGFP reporter gene over time in each population,

and the trends observed in each population in comparison to the others. Finally, FISH was used to investigate the status of the vectors within the cells as episomal vectors and/or as integrants.

Table of plasmid vectors tested in this chapter:

Vector Name	Vector Feature	Promoter-Transgene	Antibiotic Resistance	Vector Length (kb)
CeGFP-C1	No S/MAR element	CAGG-eGFP	Kan/Neo	5.0
CpEPI-eGFP	Full-length 2kb β-IFN S/MAR element	CAGG-eGFP	Kan/Neo	7.0
CMYC-pEPI	Full-length c-myc 1.7kb S/MAR element	CAGG-eGFP	Kan/Neo	6.7

Table 7.1 Vectors used in this chapter, indicating the name, the presence of an S/MAR element, the transgene promoter, the transgene used for assessment studies, the antibiotic resistance gene, and the length (kb) of each construct.

7.2 Results

7.2.1 Experimantal Design

The experiment was designed as follows (flow chart Figure 7.1). Three 185cm^2 flasks were seeded with 18.5x10^5 C2C12 cells. After 24 hours the cells had reached ~70% confluency. One flask was transfected with CeGFP-C1, the second with CpEPI-eGFP, and the third with CMYC-pEPI by lipofection (as described in Chapter 2: Materials & Methods). 24 hours after transfection each population of cells was analysed under a fluorescent microscope to confirm the presence of eGFP positive cells indicating that the cells had been successfully transfected with plasmid. Cell samples were collected from each population for analysis. Each transfected population was then divided into 4 different groups with designations (A), (B), (C), and (D):

For group (A) 1x10^5 cells from each population were sorted by FACs (as described in Chapter 2: Materials & Methods) for eGFP positive cells 24 hours post transfection, and the positive cells were seeded into three separate 10cm^2 flasks. The cells were allowed to proliferate for 35 days in growth medium and were passaged each time culture confluency reached ~70%. Cell samples from each flask were collected on the final day of the experiment for further analysis.

For group (B) cells were taken from each of the three populations and seeded at a density of 1x10^5 cells into three separate 25cm^2 flasks each. These cells were held in the G0/G1 phase of the cell cycle in low serum methionine depleted medium for 7 days and then put through a FACs sorter and sorted for GFP positive cells. They were then allowed to proliferate for 35 days in growth medium in 10cm^2 flasks and were passaged each time culture confluency reached ~70%. Cell samples from each flask were collected on the final day of the experiment for further analysis. This condition was used to show whether holding the cells in G0/G1 was sufficient to establish the plasmid within the cells and was the closest *in vitro* model to represent what may occur *in vivo* considering selection cannot be used *in vivo*.

For group (C) cells were taken from each of the three populations and seeded at a density of 1x10^5 cells into four separate 25cm^2 flasks each. These cells were held in

G0/G1 phase of the cell cycle in low serum methionine depleted medium for 7 days, then put under growth medium for 24 hours. Next, they were put under G-418 selection for 10 days, fresh media containing G-418 was added to each culture every 48 hours, and the cells were passaged when culture confluency reached ~70%. On the final day of selection cell samples were collected from each population for analysis. The remaining cells were seeded again into fresh $10cm^2$ flasks, and the cells were passaged for a further 35 days. Cell samples from each flask were again collected on the final day of the experiment for further analysis. This condition was used in Chapter 5 in order to determine whether the combination of both holding the cells in G0/G1 followed by selection was essential for the observed long term transgene expression.

For group (D) cells were taken from each of the three populations and seeded at a density of $1x10^5$ cells into four separate $25cm^2$ flasks each. The cells were then put under G-418 selection for 10 days. On the final day of selection cell samples were collected from each population for analysis. The remaining cells were seeded again into fresh $10cm^2$ flasks, and the cells were passaged for a further 35 days. Cell samples from each flask were again collected on the final day of the experiment for further analysis. This condition was used in order to determine whether holding the cells in G0/G1 was unnecessary where only selection was essential for vector establishment, and acted as a control for the experiment.

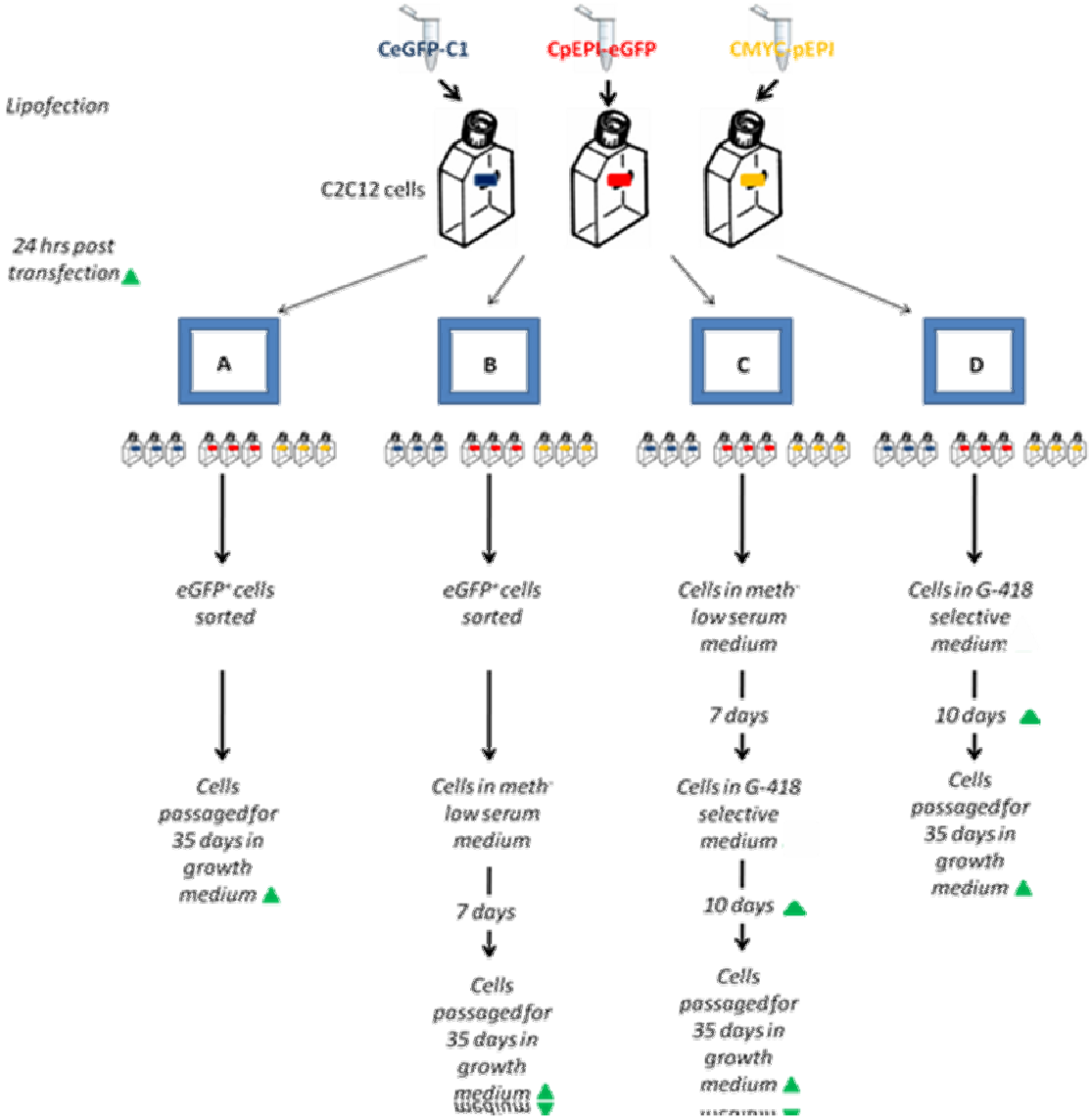
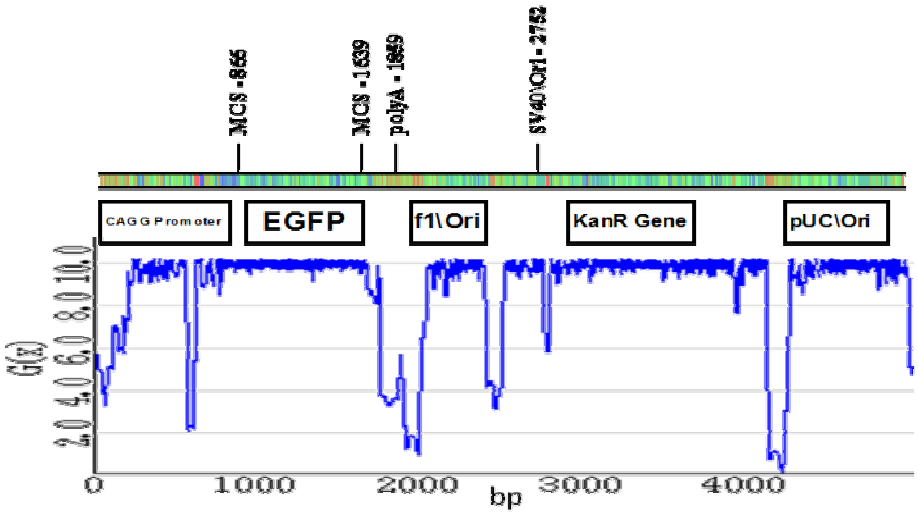


Figure 7.1 Flow chart diagram of experimental sequence. Three flasks of C2C12 cells were seeded, each transfected with a different vector (CeGFP-C1 (blue), CpEPI-eGFP (red), and CMYC-pEPI (yellow)) by lipofection. The cells were split into 4 groups (A-D), with three flasks of each transfected population per group. Groups A and B were sorted for eGFP positive cells 24 hours post transfection. Group A was then passaged for 35 days in growth medium. Group B was held in G0/G1 for 7 days before switching to growth medium for 35 days. Group C cells were held in G0/G1 for 7 days, selected with G-418 for 10 days before switching to growth medium for 35 days. Group D was selected for 10 days then switched to growth medium for 35 days. The green triangles mark time points at which cell samples were taken for further analysis.

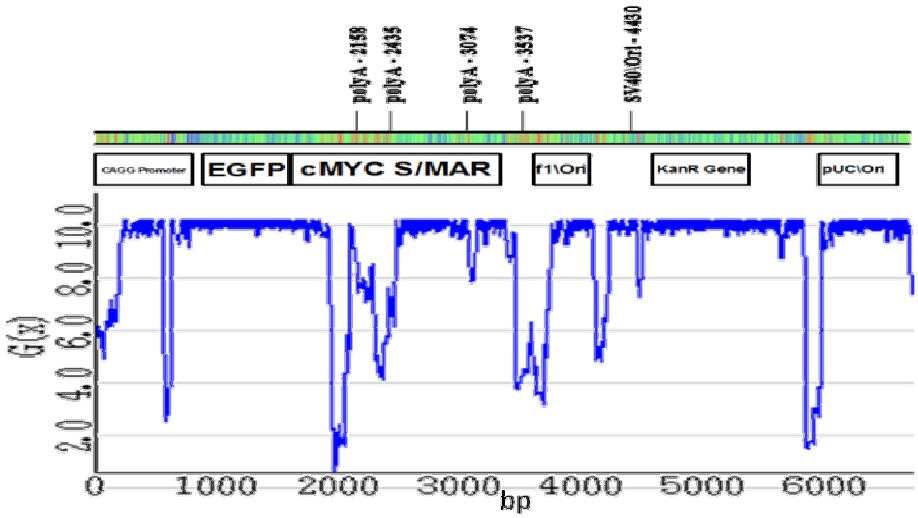
7.2.2 Stress-Induced Duplex Destabilisation (SIDD) Profile Generated For CMYC-pEPI Vector And Compared To CeGFP-C1 And CpEPI-eGFP

The Stress-Induced Duplex Destabilisation (SIDD) profile of the CAGG-driven CMYC-pEPI vector construct was calculated by the WebSIDD program designed by Dr. Craig Benham of the UC Davis Genome Centre (http://genomics.ucdavis.edu/benham/sidd/) and compared to those of CeGFP-C1 and CpEPI-eGFP previously calculated in Chapter 4 (Figure 7.2). As was observed in the S/MAR region of CpEPI-eGFP, the cMYC S/MAR region is one that contains extensive strand separation potential, requiring relatively minimal energy to do so. However the unpairing region does not span the entire length of the S/MAR, which may result in a lower degree of negative supercoiling in the regions surrounding the element within this plasmid. The benefit of this is that less energy is required, in comparison to the CpEPI-eGFP plasmid, for the CAGG promoter region to denature, and the same applies to the region surrounding the origins of replication. The profile also resembles that of the CeGFP-C1 and CpEPI-eGFP plasmid in its regions where the DNA is most likely to denature, such as promoter and origins or replication regions (Figure 7.2), and the energy required for strand separation at the start of the CAGG promoter region is in between that of the two vectors, where CAGG in CeGFP-C1 requires approximately 3.0 kcal/mol (Figure 7.2(A)); CAGG in CMYC-pEPI requires approximately 5.0 kcal/mol, and CAGG in CpEPI-eGFP requires approximately 8.0 kcal/mol (Figure 7.2 (B) and (C), respectively). From this profile it was expected that expression of the reporter gene by CMYC-pEPI would be lower than that by CeGFP-C1, but higher than that by CpEPI-eGFP.

(A) CeGFP-C1 Vector



(B) CMYC-pEPI Vector



(C) CpEPI-eGFP

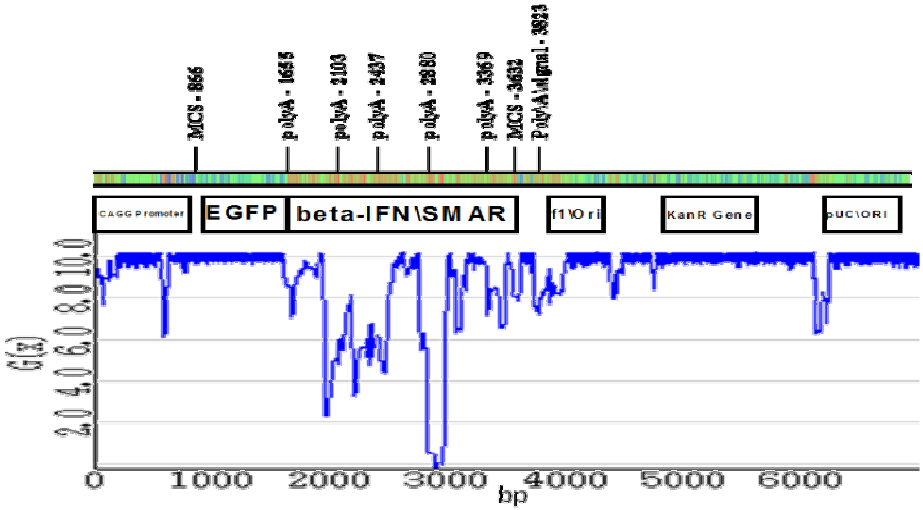


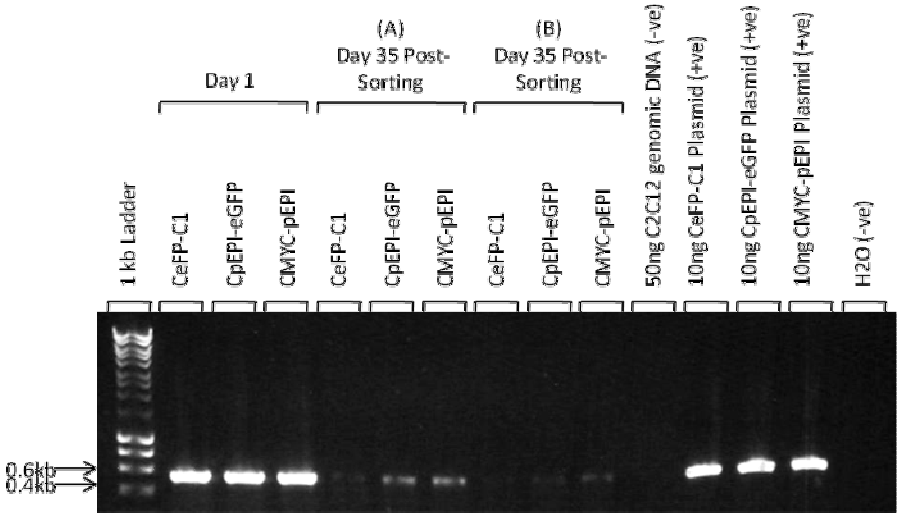
Figure 7.2 SIDD profile of CeGFP-C1 (A), CMYC-pEPI (B), and CpEPI-eGFP (C) calculated using WebSIDD program. Regions requiring lower amounts of energy ($G_{(x)}$[kcal/mol]) to base unpair are centred around

the origins of replication (F1 Ori and pUC Ori), and the SV40 promoter region drivng Neomycin resistance gene. The start of the promoter region of CeGFP-C1 requires just over 3.0 kcal/mol for denaturation, that of CMY-pEPI requires approximately 5.0 kcal/mol, and that of CpEPI-eGFP requires approximately 8.0 kcal/mol. This suggests transgene expression of CMYC-pEPI may be lower than that of CeGFP-C1 but greater than that of CpEPI-eGFP.

7.2.3 Investigation of The Presence Of CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI Vector Sequences Within C2C12 Transfected Populations In Groups (A)-(D) By PCR

PCR analysis was conducted on samples taken on Day 1 (24 hours post transfection), Final Day of Selection (after 10 days under G-418 selective pressure for groups C (post holding in G0) and D; groups A and B were sorted and not selected), and Day 35 Post Sorting/Selection, from C2C12 cells transfected with CeGFP-C1, CpEPI-eGFP, or CMYC-pEPI plasmids. 50ng of total DNA extracted from each sample was used per PCR reaction, and it was run for 35 cycles. The primers used were designed by Primer3 and lie within the Neomycin resistance gene present in the transfected vectors. The PCR was run for 35 cycles and the expected product size was 503bp. Products were run on 0.8% agarose gels counterstained with ethidium bromide and visualised under UV light. PCR on all the samples confirmed the presence of CeGFP-C1, CpEPI-eGFP, or CMYC-pEPI vector in each respective transfected population, in samples of each time point tested under each of the different conditions (A)-(D) (Figure 7.3).

(I)



(II)

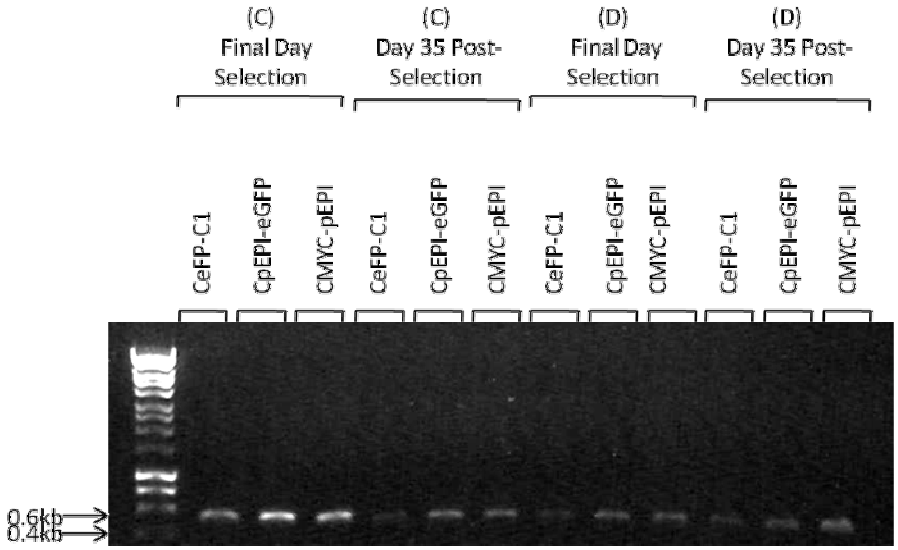


Figure 7.3 PCR analysis conducted on 50ng of total C2C12 DNA transfected with CeGFP-C1, CpEPI-eGFP, or CMYC-pEPI plasmid and put under conditions (A) and (B) (image I), and under conditions (C) and (D) (image II). Samples were taken on Day 1 and Day 35 post-sorting. The positive controls were 10ng of CeGFP-C1, 10ng of CpEPI-eGFP, and 10ng of CMYC-pEPI, and the negative controls were 50ng untransfected C2C12 genomic DNA, and H_20. The image indicates that there were copies of plasmid sequence within each population at each time point.

7.2.4 Quantitative Analysis of Long-Term Transgene Expression Of CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI Transfected Myoblast Populations Put Under Conditions (A)-(D) By Flow Cytometry

Samples from each group (A)-(D) were analysed by flow cytometry at Day 1 (24 hours post-transfection) and at Day 35 post sorting/selection (final day of experiment after 35 days of cell proliferation without selective pressure, after the period of selection) to determine the mean percentage of positive cells and the mean fluorescence intensity of each population. The mean percentages, mean fluorescence intensity, and total eGFP ± the standard error of the means (SEM) (n = 3) were plotted (Figure 7.4-7.6). Statistical tests were applied using NCSS statistics program to determine significance. The test used to compare the transfected populations to each other at each time point, and to compare all of the groups at of the time points was One-Way ANOVA. The P-value was set to 0.05 and indicated a significant difference between groups if $P<0.05$. The post-hoc Tukey-Kramer test was used to determine specifically which groups differed significantly from other groups.

The flow cytometry data showed that at Day 1 the percentage of cells transfected by each of the three plasmids CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI were significantly different from each other (means of 75.2 ± 3.3%, 43.3 ± 3.3%, and 26.2 ± 1.0% respectively; P = 0.007) (Figure 7.4), as were the mean fluorescence intensities (means of 22543 ± 501.4, 1844.3 ± 40.6,and 3701.3 ± 80.9 respectively; P = 0.02) (Figure 7.5), and hence the total eGFP expressions (means of 1620738 ± 51237, 73801.4 ± 1068, and 93424.6 ± 834.9 respectively; P = 0.01) (Figure 7.6). The plasmid which transfected the highest percentage of cells and led to the greatest mean expression, and therefore the largest amount of total eGFP expression, was the CeGFP-C1 plasmid. The CpEPI-eGFP plasmid transfected fewer cells than CeGFP-C1 but more than CMYC-pEPI. However, mean expression of the CpEPI-eGFP transfected population was lower than the other two populations, leading to its total eGFP expression to be the lowest of the 3 populations. The CMYC-pEPI plasmid transfected the lowest amount of cells but due to its mean expression, which was higher than that from CpEPI-eGFP, the CMYC-pEPI transfected population's total expression was higher than that of the CpEPI-eGFP transfected population.

By the final day of the experiment at Day 35 post selection/sorting, in comparing all the groups together, many of the groups differed significantly from one another in

terms of the percentage of cells found to be eGFP positive (P = 0.01). The populations that had the highest percentages of eGFP positive cells and were similar to each other yet significantly different to the rest of the populations were the CpEPI-eGFP population of group (C) (mean of 2.5 ± 0.3%) and the eGFP-C1 population of group (D) (mean of 2.4 ± 0.2%). The populations that were all similar to each other in their percentages of eGFP positive cells with the lowest percentages of eGFP positive cells were all the populations (CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI) in group (A) (means of 0.3 ± 0.1%, 0.4 ± 0.1%, and 0.7 ± 0.03% respectively) and all the populations in group (B) (means of 0.3 ± 0.03%, 0.4 ± 0.03%, and 0.9 ± 0.1% respectively), in addition to the CMYC-pEPI population of group (D) (mean of 0.7 ± 0.1%) and the CeGFP-C1 population of group (C) (0.8 ± 0.2%). The CMYC-pEPI population of group (C) and the CpEPI-eGFP population of group (D) (means of 1.1 ± 0.1% and 1.2 ± 0.1% respectively) both had significantly higher percentages of positive cells than the CeGFP-C1 and CpEPI-eGFP populations of both groups (A) and (B) (P = 0.01). In comparing all the CeGFP-C1 populations to one another, the ones of groups (A), (B), and (C) were all comparable, whereas that of group (D) had a significantly higher percentage of positive cells (P = 0.01). In comparing all the CpEPI-eGFP populations to one another, the populations of groups (A) and (B) were comparable and had the lowest percentage, followed by population of group (D) which was significantly higher, and finally that of group (C) which had the highest proportion of positive cells. In comparing the CMYC-pEPI populations to one another, those of all the groups (A), (B), (C), and (D) were similar in the number of eGFP positive cells with no significant differences.

In terms of the mean expressions compared amongst the populations of the groups at Day 35 post selection/sorting, the only population that was significantly different and had the highest mean expression was the CeGFP-C1 population of group (D) (mean of 417 ± 28.5; P = 0.0004). The only population that did not differ from any of the populations was the CeGFP-C1 one of group (A) (mean of 358 ± 57.5). The mean fluorescence intensities of all the remaining populations were similar, which included the CpEPI-eGFP and CMYC-pEPI populations of group (A) (means of 278.7 ± 9.8 and 285.7 ± 5.7 respectively), the CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI transfected populations of group (B) (means of 274.7 ± 6.9, 284.3 ± 14.6, and 272.7 ± 7.1 respectively) and of group (C) (means of 269.3 ± 4.2, 290.7 ± 1.2, and 270.7 ± 3.5

respectively), and the CpEPI-eGFP and CMYC-pEPI transfected populations of group (D) (means of 292.7 ± 4.8 and 275.7 ± 7.8 respectively).

After calculating the total expression of eGFP of each population and statistically analysing the means, it was found that there were significant differences between the populations of the different groups (P value = 0.002). The population with the highest total expression that was significantly different to all the other groups was the CeGFP-C1 population of group (D) (mean of 1007.4 ± 53.8). The next significantly highest total expression that was not similar to any other group but was significantly lower than the CeGFP-C1 population of group (D) was the CpEPI-eGFP population of group (C) (mean of 736.7 ±80.9). This was followed by the CpEPI-eGFP population of group (D) (mean of 340.7 ± 21.7) whose expression was significantly higher than that of the CeGFP-C1 and CpEPI-eGFP populations of group (A) (means of 127.4 ± 53.2 and 110.3 ± 12.2 respectively) and of group (B) (73.6 ± 10.6 and 124.2 ± 16.2 respectively), but was comparable to that of the CMYC-pEPI population of group (C) (mean of 288.8 ± 25.0). The expression of the CMYC-pEPI population of group (C) was only significantly higher than that of the CeGFP-C1 population of group (B). The rest of the populations all had amounts of total expression that did not significantly differ from one another and were the ones with the lowest total expressions. These populations included all populations (CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI) of groups (A) (means of 127.4 ± 53.2, 110.3 ± 12.2, and 209.3 ± 8.1 respectively) and (B) (means of 73.6 ± 10.6, 124.2 ± 16.2, and 245.6 ± 19.0 respectively), as well as the CeGFP-C1 population of group (C) (mean of 214.8 ± 67.5) and the CMYC-pEPI population of group (D) (mean of 184.3 ± 21.5). In comparing all the CeGFP-C1 populations to one another the populations of groups (A), (B), and (C) were not significantly different from each other, and the population that had the highest significant total expression was that of group (D). In comparing all the CpEPI-eGFP populations to one another, those of groups (A) and (B) had the lowest total expression and were similar to each other. The population with significantly higher total expression than groups (A) and (B) was that of group (D). The one with the highest was that of group (C). In comparing all the CMYC-pEPI populations to one another it was found that none of them differ significantly from each other.

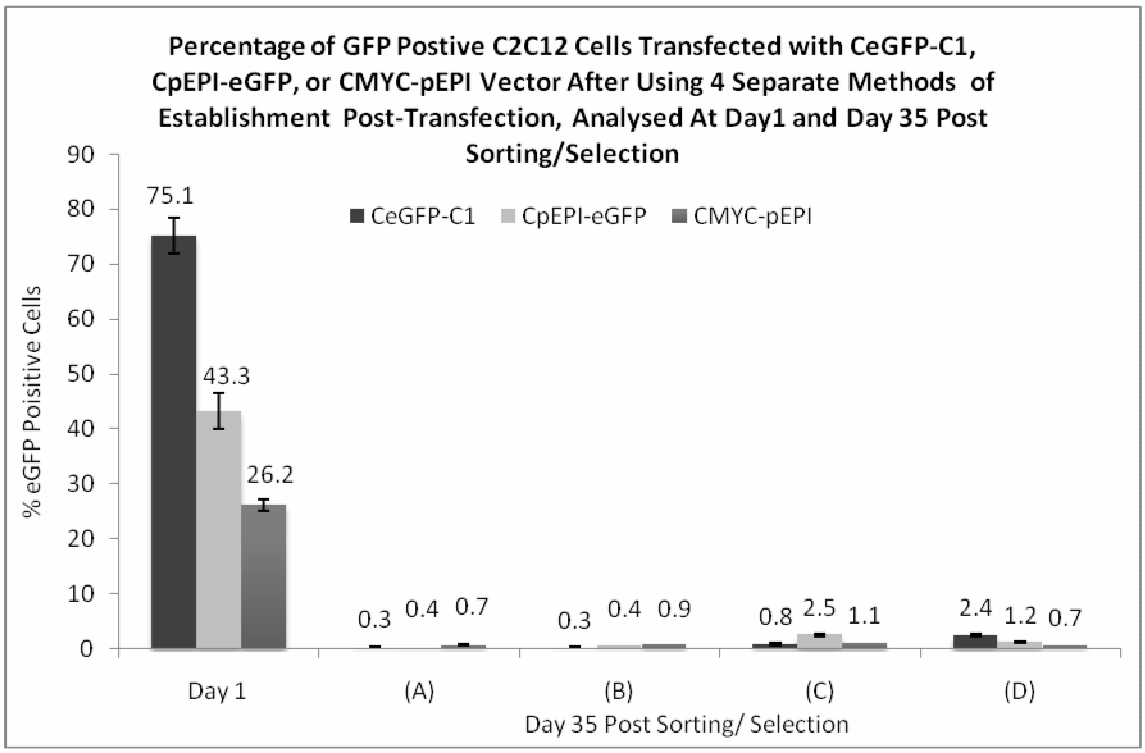


Figure 7.4 Flow cytometry analysis of the mean percentage ± SEM (error bars) of eGFP positive C2C12 cells transfected with CeGFP-C1, CpEPI-eGFP, or CMYC-pEPI plasmid then separated into 4 different groups each (A)-(D), with each group composed of three replicates (*n* = 3). The values above the bars represent the mean percentage of eGFP positive cells. Group (A) cells were sorted for eGFP positive cells 24 hours post transfection, and the positive cells were allowed to proliferate for 35 days. Group (B) cells were held in G0 phase of the cell cycle in low serum methionine depleted medium for 7 days, sorted for eGFP positive cells, then allowed to proliferate for 35 days. Group (C) cells were held in G0 phase of the cell cycle in low serum methionine depleted medium for 7 days, put under growth medium for 24 hours, put under G-418 selection for 10 days, then allowed to proliferate for 35 days. Group (D) cells were put under G-418 selection for 10 days then were allowed to proliferate for 35 days. At Day 1 the CeGFP-C1 plasmid transfected the largest number of cells, followed by the CpEPI-eGFP plasmid, and finally the CMYC-pEPI plasmid. By Day 35 post sorting/selection the populations that had the highest percentage of eGFP positive cells and were similar to each other yet significantly different to the rest of the populations were the CpEPI-eGFP population of group (C) and the CeGFP-C1 population of group (D). The populations which contained the lowest percentages of positive cells and were all similar to each other in their percentages were the CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI populations in group (A) and in group (B), in addition to the CeGFP-C1 population of group (C), and the CMYC-pEPI population of group (D). The CMYC-pEPI population of group (C) and the CpEPI-eGFP population of group (D) both contained a significantly higher proportion of positive cells than the CeGFP-C1 and CpEPI-eGFP populations of both groups (A) and (B).

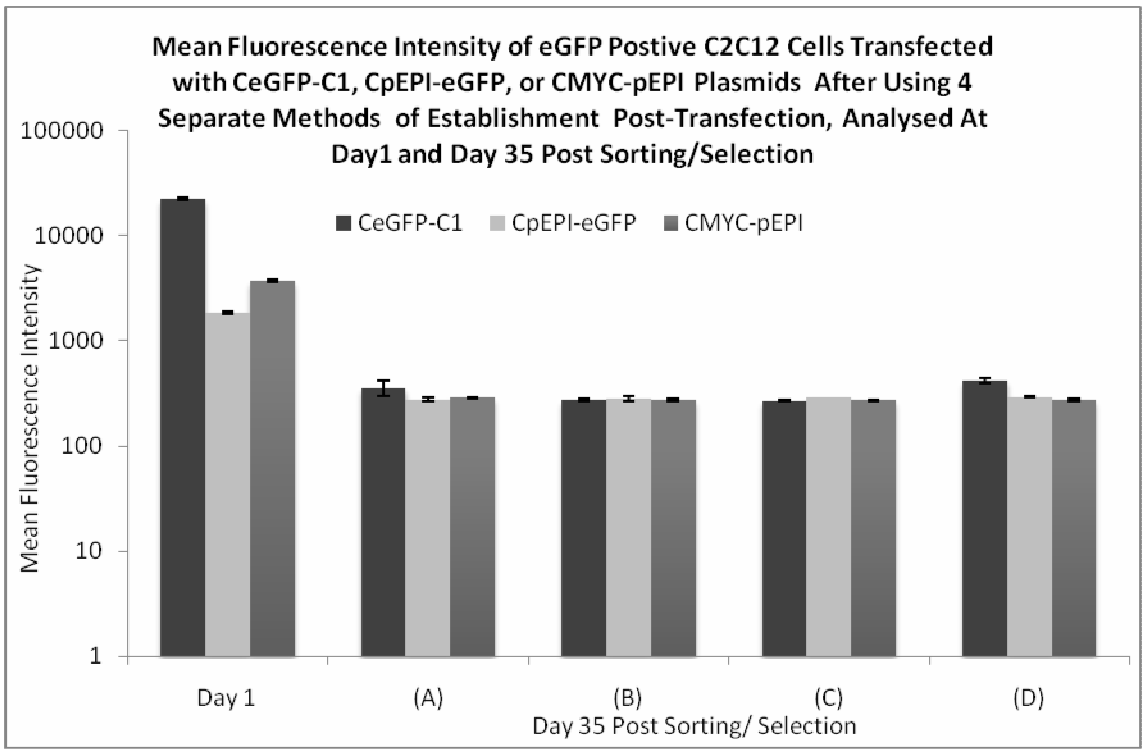


Figure 7.5 Flow cytometry analysis of the mean fluorescence intesity ± SEM (error bars) of eGFP positive C2C12 cells transfected with CeGFP-C1, CpEPI-eGFP, or CMYC-pEPI plasmid then separated into 4 different groups each (A)-(D), with each group composed of three replicates (*n* = 3). The were converted to log as the data was spread over a wide range. At Day 1 CeGFP-C1 cells showed the highest mean expression, followed by the CMYC-pEPI cells, and finally the CpEPI-eGFP cells. At Day 35 post sorting/selection where the mean expressions were compared amongst the populations of the groups, all of them were similar to each other. The only sample that was significantly different and had the highest mean expression was the CeGFP-C1 population of group (D).

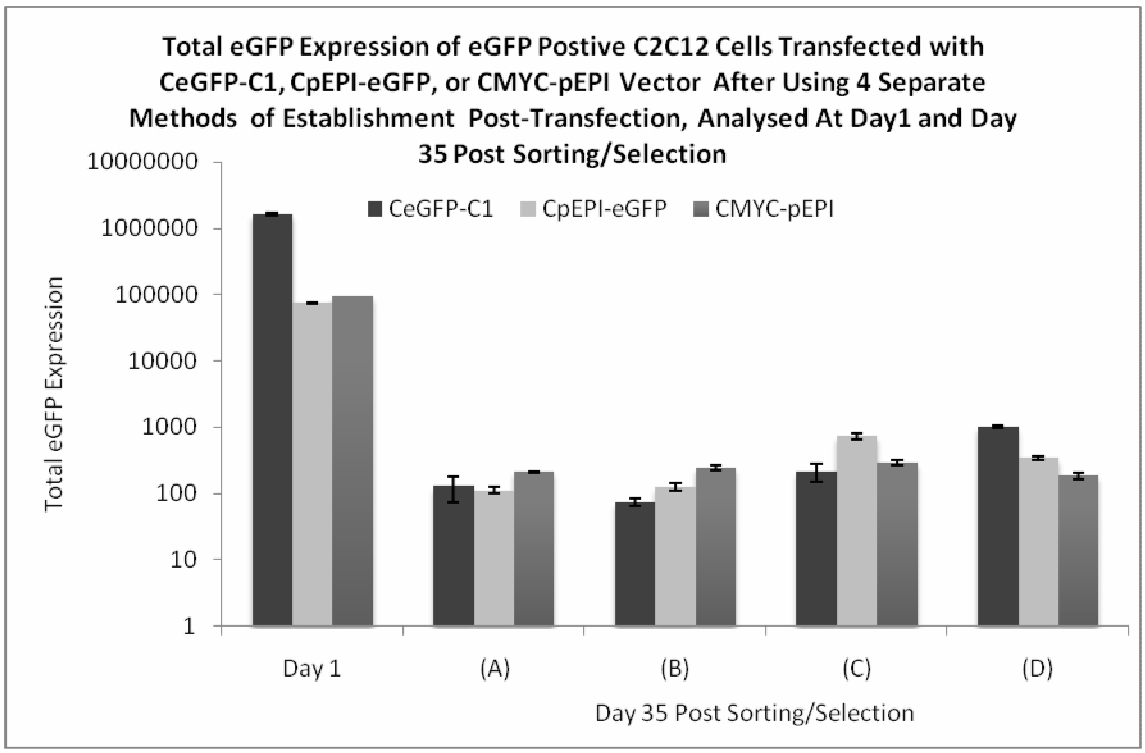


Figure 7.6 Total expression ± SEM (error bars) of eGFP positive C2C12 cells transfected with CeGFP-C1, CpEPI-eGFP, or CMYC-pEPI plasmid then separated into 4 different groups each (A)-(D), with each group composed of three replicates (*n* = 3). The total eGFP fluorescence values were converted to log as the data was spread over a wide range. At Day 1 CeGFP-C1 cells showed the highest total expression, followed by the CMYC-pEPI cells, and finally the CpEPI-eGFP cells. By Day 35 post sorting/selection it can be seen that the population with the highest total expression was the CeGFP-C1 population of group (D). The next significantly highest total expression was achieved by the CpEPI-eGFP population of group (C). This was followed by the CpEPI-eGFP population of group (D) whose expression was significantly higher than that of the CeGFP-C1 and CpEPI-EGFP populations of group (A) and of group (B), but was comparable to that of the CMYC-pEPI population of group (C). The expression of the CMYC-pEPI population of group (C) was only significantly higher than that of the CeGFP-C1 population of group (B). The rest of the populations all had amounts of total expression that did not significantly differ to one another and were the ones with the lowest total expression. These populations included all populations (CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI) of groups (A) and (B), as well as the CeGFP-C1 population of group (C) and the CMYC-pEPI population of group (D).

7.2.5 Investigation Of Episomal/Integrant Status Of CpEPI-eGFP, CeGFP-C1, and CMYC-pEPI Vectors In Transfected C2C12 Cell Populations Of Groups (A)-(D) By Fluorescent In-Situ Hybridisation (FISH) Analysis On Final Day Samples

FISH was used to determine the episomal and/or integrant status of the transfected vectors, as conducted in the previous chapters (as described in Chapter 2: Materials & Methods). The positive controls used in this experiment were C2C12 cells 24 hours post transfection with CpEPI-eGFP, CeGFP-C1, or CMYC-pEPI plasmid (Figure 7.7 (i), (ii), and (iii), respectively). From the images taken under the fluorescence microscope it can be seen that a large amount of plasmid was present within the positive control cells, as the fluorescence signal emitted was very strong. It was also clear that much of the plasmid observed was clustered together. The negative control in this experiment was untransfected C2C12 cells (Figure 7.7 (iv)).

The images shown for the populations in group (A) indicate that there was no evidence of integration of CpEPI-eGFP (0/7 positive spreads) or CeGFP-C1 (0/6 positive spreads) into the host genome, and the plasmids were retained as episomes (Figure 7.8 (i) and (ii), respectively). Evidence of integration of the CMYC-pEPI into the host genome was found (1 integrant spread out of 7 positive spreads), indicating that this plasmid may have been able to integrate into the C2C12 genome without selection pressure. CMYC-pEPI was also found in episomal status in addition to evidence being found for its integration (Figure 7.8 (iii)).

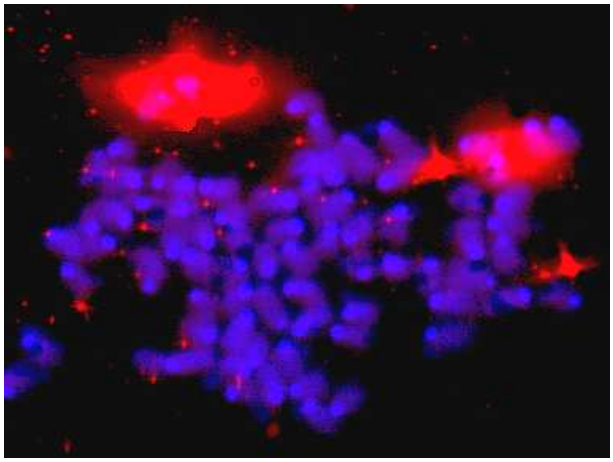
Interestingly, in group (B) none of the vectors were found to have integrated into the host genome (0/6 CeGFP-C1, 0/7 CpEPI-eGFP, and 0/6 CMYC-pEPI positive spreads) (Figure 7.9 (i)-(iii)).

Group (C) images of the CpEPI-eGFP and CeGFP-C1 populations show that the vectors remained episomal and there was no evidence of integration (0/7 and 0/6 positive spreads, respectively) (Figure 7.10 (i) and (ii), respectively). CMYC-pEPI had again potentially integrated into the host genome (2 integrant spreads out of 7 positive spreads), in addition to being retained episomally (Figure 7.10 (iii)).

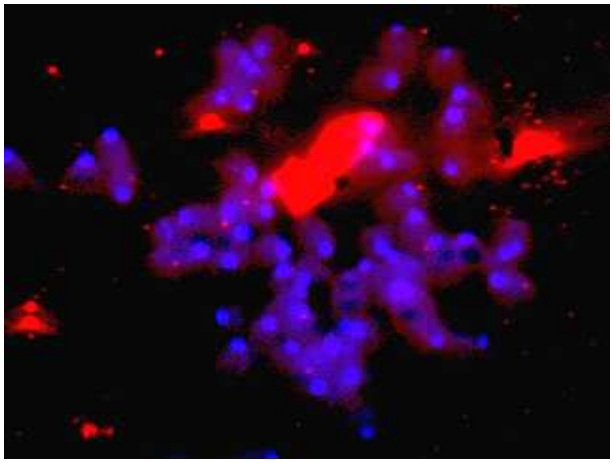
Group (D) images show that once again no evidence of CpEPI-eGFP integration was found and the vector remained episomal (0 integrant spreads out of 6 positive spreads) (Figure 7.11 (i)), whereas evidence for both CeGFP-C1 and CMYC-pEPI

integration into the genome was observed (2/7 and 1/5 integrant spreads, respectively) as well as remaining episomal (Figure 7.11 (ii) and (iii)).

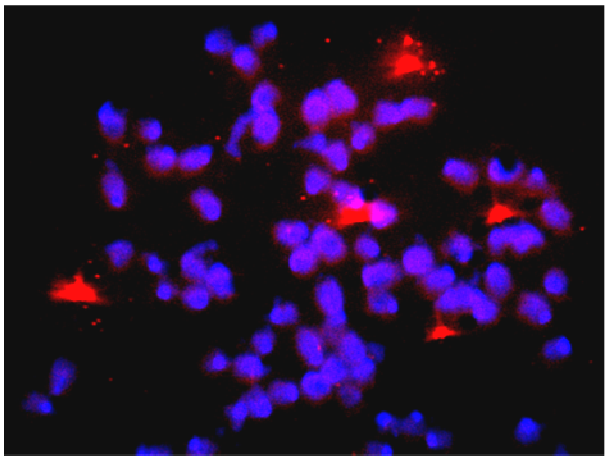
(i) C2C12 cells 24 hr after transfection with CpEPI-eGFP



(ii) C2C12 cells 24 hr after transfection with CeGFP-C1



(iii) C2C12 cells 24 hr after transfection with CMYC-pEPI



(iv) C2C12 cells untransfected

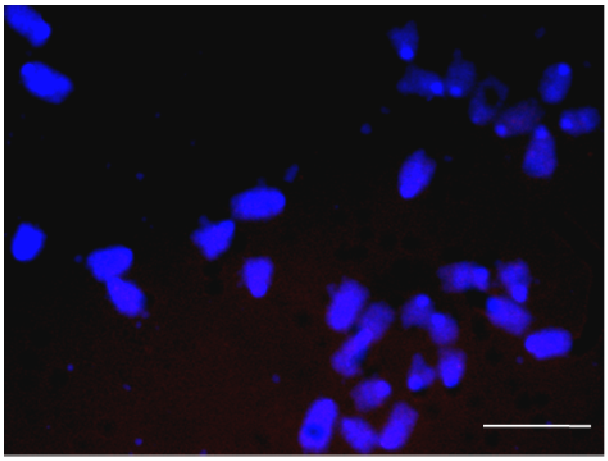
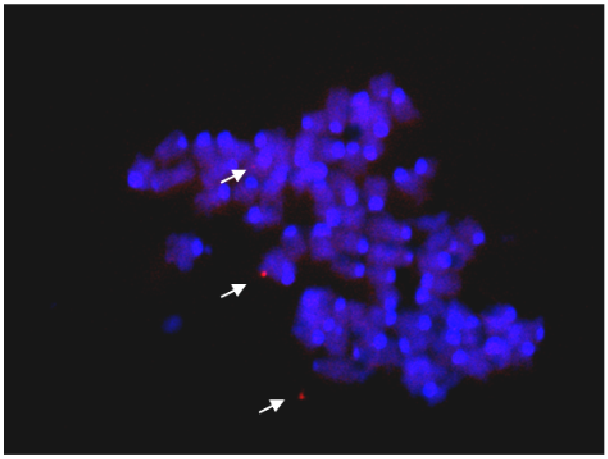


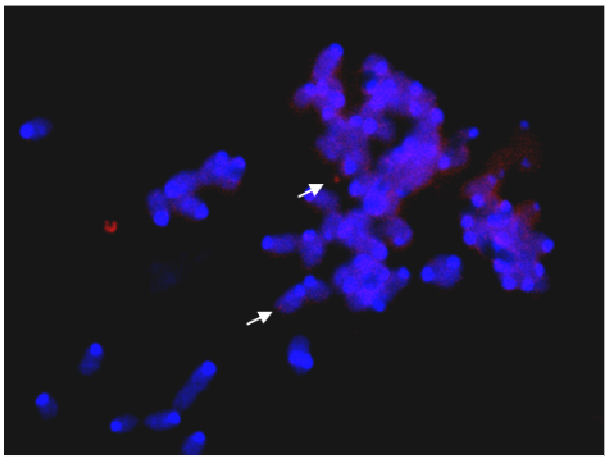
Figure 7.7 Fluorescent In-Situ Hybridisation positive control of C2C12 cells 24 hours post transfection with the CpEPI-eGFP (A), CeGFP-C1 (B), and CMYC-pEPI (C). The plasmid can be seen as the bright red fluorescing spots. Genomic DNA was counterstained with Dapi (blue). As can be seen there was a large amount of vector present within the transfected cells, and the plasmid was found mostly clustered together. The negative control was untransfected C2C12 cells (D) indicating no background. Bar, 5µm.

Group (A) transfected C2C12 cells on Final Day

(i) CpEPI-eGFP



(ii) CeGFP-C1



(iii) CMYC-pEPI

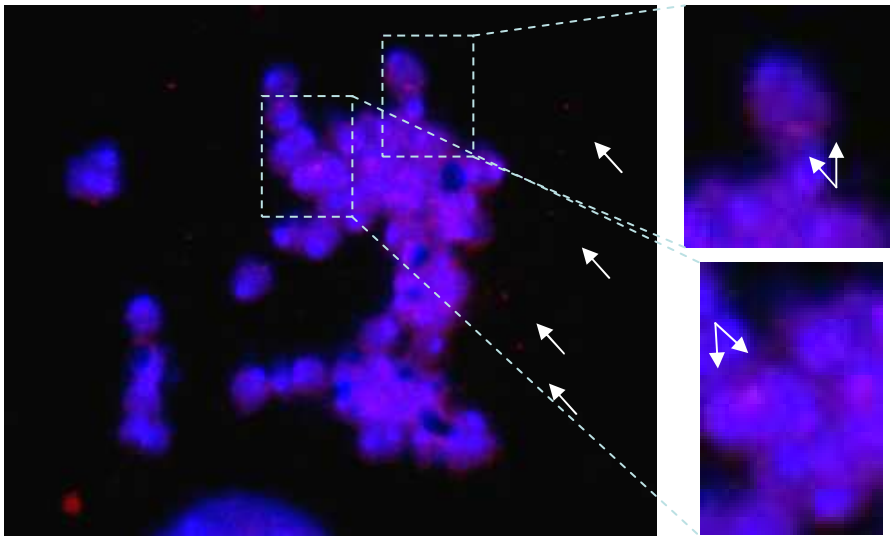
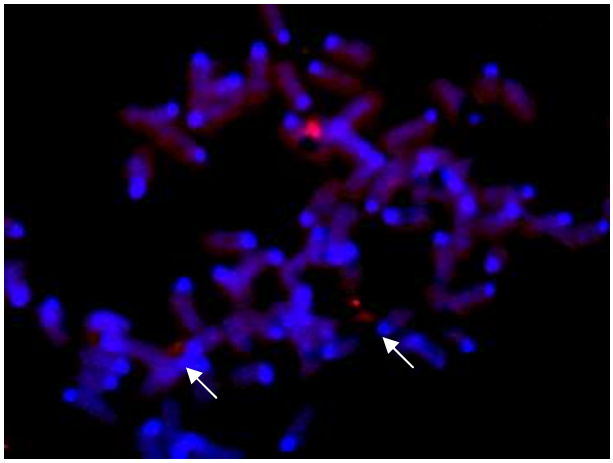


Figure 7.8 Fluorescent In-Situ Hybridisation of C2C12 cells transfected with CpEPI-eGFP plasmid (i), CeGFP-C1 plasmid (ii), CMYC-pEPI (iii), Day 35 post sorting, group (A). Single white arrows indicate red

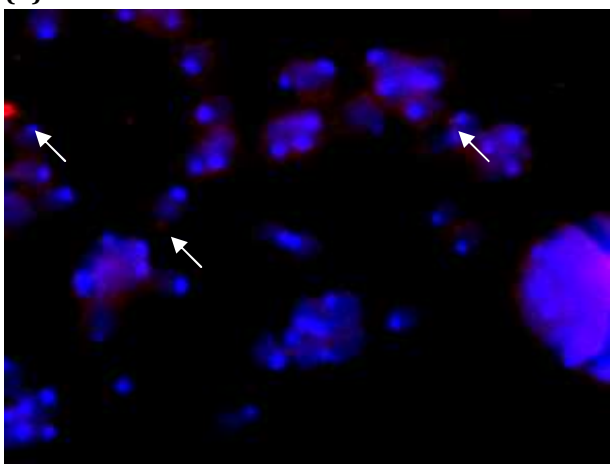
fluorescence spots where the probe has hybridised to plasmid sequence and denotes the episomal status of the plasmid. The double arrows mark spots of potential integration. Genomic DNA was counterstained with Dapi (blue). Images (i) and (ii) show CpEPI-eGFP and CeGFP-C1 copies, were retained as episomes with no evidence of integration (0/7 and 0/6 positive spreads, respectively). Image (iii) suggests CMYC-pEPI integrated into the C2C12 host genome (1 integrant spread out of 7 positive spreads) in addition to being retained as an episome.

Group (B) transfected C2C12 cells on Final Day

(i) CpEPI-eGFP



(ii) CeGFP-C1



(iii) CMYC-pEPI

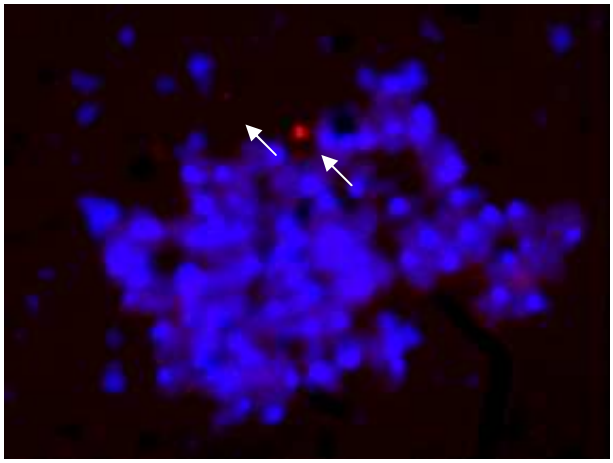
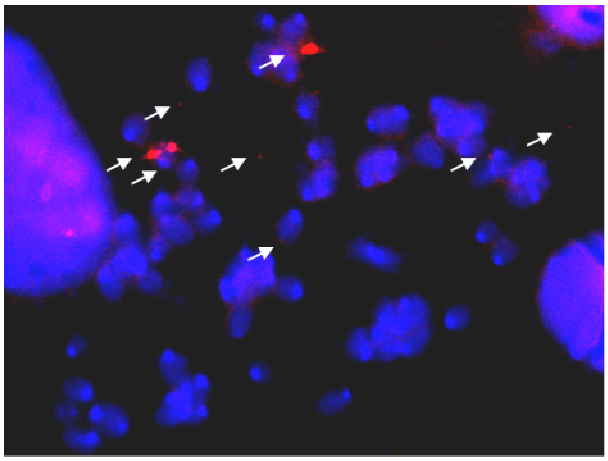


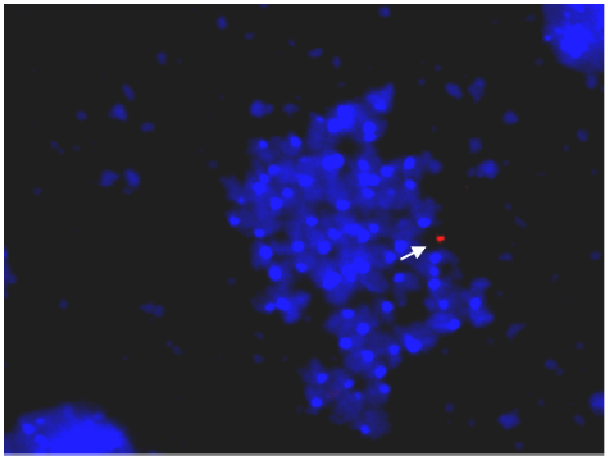
Figure 7.9 Fluorescent In-Situ Hybridisation of C2C12 cells transfected with CpEPI-eGFP (i), CeGFP-C1 (ii), or CMYC-pEPI (iii), Day 35 post sorting, group (B). Single white arrows indicate red fluorescence spots where the probe has hybridised to plasmid sequence and denotes the episomal status of the plasmid. Genomic DNA is counterstained with Dapi (blue). The images show no evidence of integration of the vectors in any of the transfected populations (0/6 CeGFP-C1, 0/7 CpEPI-eGFP, and 0/6 CMYC-pEPI positive spreads), and the vectors were present as episomes in these cells.

Group (C) transfected C2C12 cells on Final Day

(i) CpEPI-eGFP



(ii) CeGFP-C1



(iii) CMYC-pEPI

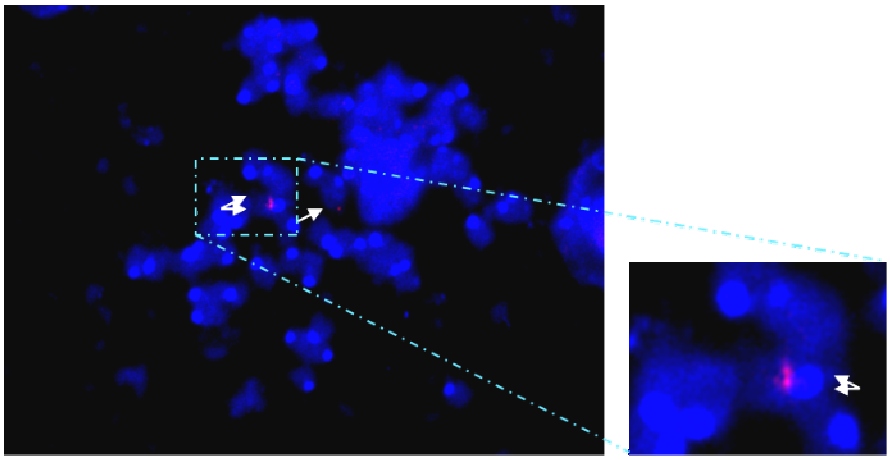
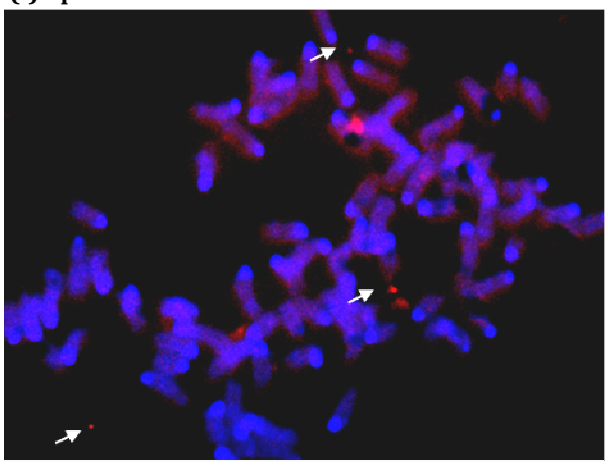


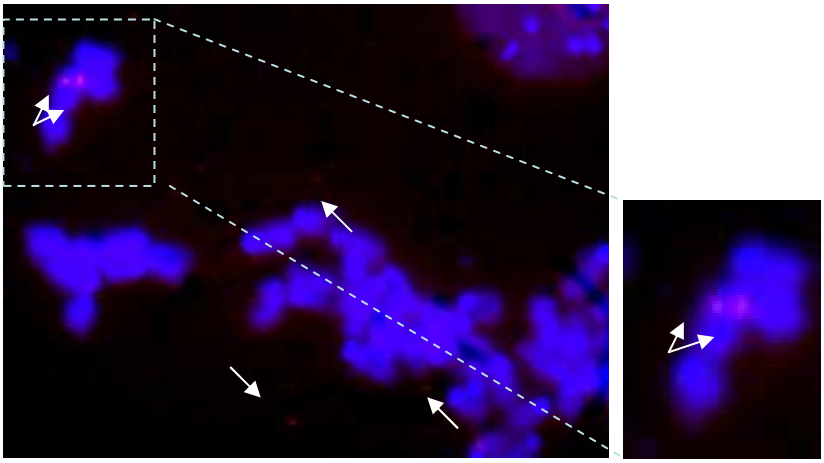
Figure 7.10 Fluorescent In-Situ Hybridisation of C2C12 cells transfected with CpEPI-eGFP plasmid (i), CeGFP-C1 plasmid (ii), or CMYC-pEPI (iii), Day 35 post selection, group (C). Single white arrows indicate red fluorescence spots where the probe has hybridised to plasmid sequence and denotes the episomal status of the plasmid. The double arrows mark spots of potential integration. Genomic DNA was counterstained with Dapi (blue). Images (i) and (ii) show the CpEPI-eGFP and CeGFP-C1 plasmids were retained as episomes with no evidence of integration. Image (iii) shows that CMYC-pEPI, was found as episomes in addition to appearing to have integrated into the C2C12 host genome (2 integrant spreads out of 7 positive spreads, respectively).

Group (D) transfected C2C12 cells on Final Day

(i) CpEPI-eGFP



(ii) CeGFP-C1



(iii) CMYC-pEPI

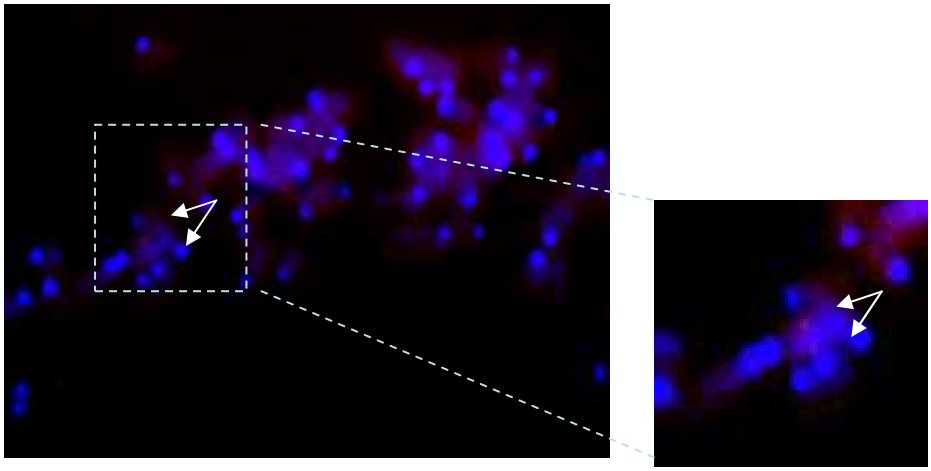


Figure 7.11 Fluorescent In-Situ Hybridisation of C2C12 cells transfected with CpEPI-eGFP plasmid (i), CeGFP-C1 plasmid (ii), CMYC-pEPI (iii), Day 35 post selection, group (D). Single white arrows indicate red fluorescence spots where the probe has hybridised to plasmid sequence and denotes the episomal

status of the plasmid. The double arrows mark spots of potential integration. Genomic DNA was counterstained with Dapi (blue). Image (i) shows CpEPI-eGFP was retained as an episome with no evidence of integration (0/6 spreads). Images (ii) and (iii) show that CeGFP-C1 and CMYC-pEPI were present as episomes in addition to appearing to have integrated into the C2C12 host genome (2/7 and 1/5 integrant spreads, respectively).

7.2.6 Summary of Results:

Vector	Day	PCR	Flow Cytometry			FISH	
			% Positive Cells	MFI	Total	Integra-tion	Epi-somal
CeGFP-C1 A	1	√	75.2 ± 3.3	22543 ± 501.4	1620738 ± 51237	0/6	√
	Final Day	√	0.3 ± 0.1	358 ± 57.5	127.4 ± 53.2		√
CeGFP-C1 B	1	√	75.2 ± 3.3	22543 ± 501.4	1620738 ± 51237	0/6	√
	Final Day	√	0.3 ± 0.03	274.7 ± 6.9	73.6 ± 10.6		√
CeGFP-C1 C	1	√	75.2 ± 3.3	22543 ± 501.4	1620738 ± 51237	0/6	√
	Final Day	√	0.8 ± 0.2	269.3 ± 4.2	214.8 ± 67.5		√
CeGFP-C1 D	1	√	75.2 ± 3.3	22543 ± 501.4	1620738 ± 51237	2/7	√
	Final Day	√	2.4 ± 0.2	417 ± 28.5	1007.4 ± 53.8		√
CpEPI-eGFP A	1	√	43.3 ± 3.3	1844.3 ± 40.6	73801.4 ± 1068	0/7	√
	Final Day	√	0.4 ± 0.1	278.7 ± 9.8	110.3 ± 12.2		√
CpEPI-eGFP B	1	√	43.3 ± 3.3	1844.3 ± 40.6	73801.4 ± 1068	0/7	√
	Final Day	√	0.4 ± 0.03	284.3 ± 14.6	124.2 ± 16.2		√
CpEPI-eGFP C	1	√	43.3 ± 3.3	1844.3 ± 40.6	73801.4 ± 1068	0/7	√
	Final Day	√	2.5 ± 0.3	290.7 ± 1.2	736.7 ±80.9		√
CpEPI-eGFP D	1	√	43.3 ± 3.3	1844.3 ± 40.6	73801.4 ± 1068	0/6	√
	Final Day	√	1.2 ± 0.1	292.7 ± 4.8	340.7 ± 21.7		√
CMYC-pEPI A	1	√	26.2 ± 1.0	3701.3 ± 80.9	93424.6 ± 834.9	1/7	√
	Final Day	√	0.7 ± 0.03	285.7 ± 5.7	209.3 ± 8.1		√
CMYC-pEPI B	1	√	26.2 ± 1.0	3701.3 ± 80.9	93424.6 ± 834.9	0/6	√
	Final Day	√	0.9 ± 0.1	272.7 ± 7.1	245.6 ± 19.0		√
CMYC-pEPI C	1	√	26.2 ± 1.0	3701.3 ± 80.9	93424.6 ± 834.9	2/7	√
	Final Day	√	1.1 ± 0.1	270.7 ± 3.5	288.8 ± 25.0		√
CMYC-pEPI D	1	√	26.2 ± 1.0	3701.3 ± 80.9	93424.6 ± 834.9	1/5	√
	Final Day	√	0.7 ± 0.1	275.7 ± 7.8	184.3 ± 21.5		√

Table 7.2 Table summarising the results obtained in this chapter for transfected C2C12 cells, indicating each vector that was transfected and the time point at which samples were extracted and analysed. PCR results are presented where a 'v' denotes that the Kanamycin sequence was amplified and a band was present; Flow Cytometry results presented include the percentage of positive cells per transfected population, the mean fluorescence intensity (MFI) of the positive cells, and the total amount of expression which was the product of expression and number of positive cells; the FISH data is presented as the number of slides which suggested an integration event having occurred out of the total number of slides analysed on Day 60 only, a 'v' in the episomal column indicated the presence of episomal copies of the vector observed on the slides, and a '-' indicates none was observed.

7.3 Discussion:

7.3.1 CMYC-pEPI transfected cells had greater transgene expression than CpEPI-eGFP at Day 1 but had a lower transfection efficiency than both CeFP-C1 and CpEPI-eGFP.

A general comparison of the SIDD profiles calculated of the three vectors CeFP-C1, CpEPI-eGFP, and CMY-pEPI showed similaries in their destabilised regions, such as at the CAGG and SV40 promoter regions and the F1 and pUC origins of replication. The profiles also indicated that the energy required for strand separation at the start of the CAGG promoter region in CMYC-pEPI was greater than that of CeGFP-C1 but lower than that of CpEPI-eGFP, which suggested that the levels of transgene expression that would result from CMYC-pEPI would be lower than CeGFP-C1 but greater than CpEPI-eGFP. This was confirmed by flow cytometry, where the mean fluorescence intensity and total expression resulting from the CMYC-pEPI vector were significantly higher than those resulting from CpEPI-eGFP, and were significantly lower than those resulting from CeGFP-C1.

Considering that the transfection efficiency of CpEPI-eGFP was lower than that of CeGFP-C1, as was shown in this chapter by flow cytometry, and as was previously found in Chapter 4, and was attributed to the presence of the S/MAR element, whose potential for destabilisation may have caused its vector to be in a more open-circular conformation as opposed to more a supercoiled one, the transfection efficiency of CMYC-pEPI was also expected to be lower than that of CeGFP-C1 due the presence of the S/MAR element within its sequence (as discussed in Chapter 4 Discussion: Section 4.3.2). This was confirmed by flow cytometry, where the percentage of positive cells transfected by CeGFP-C1 was significantly greater to that transfected by CMYC-pEPI. However, it was also found that CMYC-pEPI's transfection efficiency was significantly lower to that of CpEPI-eGFP, even though both vectors are the same size of 6.7kb. One possible explanation is that the most significant and major region of destabilisation in CpEPI-eGFP was the S/MAR region, whereas CMYC-pEPI had several regions of considerable propensity for strand separation, as was shown by the SIDD profiles, leading to a more open vector conformation than CpEPI-eGFP, and thus a lower transfection efficiency (Cherng *et al*., 1999).

7.3.2 Holding CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI transfected cells in G0/G1 alone did not improve long-term transgene expression over unselected transfected populations

Condition (B) under which the cells were put involved holding the transfected cells in the G0/G1 phase of the cell cycle by using methionine-depleted low-serum medium for 7 days and then sorting the eGFP positive cells, before allowing the cells to proliferate for 35 days. This condition was tested in order to determine whether holding the cells in G0/G1 would suffice to establish a plasmid as an episome, without the requirement for selection. The flow cytometry data for CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI transfected populations in group (B) indicated that the percentage of positive cells and total expression were similar to those seen in the populations of group (A), where the cells were transfected, sorted, and then allowed to proliferate for 35 days without any selection pressure imposed at any point. Expression was extremely low in every transfected population in both groups (A) and (B), and eGFP positive cells made up a very small percentage of the transfected populations, none of which differed significantly from the others by the final day of the experiment. This finding is important in that it has shown that holding transfected cells in quiescence is not enough to establish an episome where, upon re-activation of the cells and re-entry into the cell cycle, the vector would not be effectively retained by the cells and would be diluted out of the population as a result of cell proliferation. The implication of this finding in relation to gene therapy is that that targeted quiescent satellite cells/primary myoblasts would not effectively pass CeGFP-C1, CpEPI-eGFP, or CMYC-pEPI vectors to their progeny upon re-activation and throughout proliferation, leading to the loss of the vectors and inadequate expression of the therapeutic transgene.

Thus it has become evident holding cells in G0/G1 alone is not sufficient for increasing the number of positive cells and expression using the CpEPI-eGFP vector, and that selection plays an important role in the establishment of this vector. This could indicate that certain events within the cells, such as further DNA markings, were made to the CpEPI-eGFP vector DNA under selection, which lead to the improved expression and possibly plasmid retention noted in the CpEPI-eGFP transfected population of group (C).

7.3.3 Not all S/MAR elements prevent vector integration into the host genome

Evidence of CMYC-pEPI plasmid integration into the C2C12 host genome was found in the final day samples obtained from CMYC-pEPI transfected populations of groups (A), (C), and (D). It was interesting to note that evidence towards the integration of the vector had occurred in the group (A) population, even though no selection was imposed upon the transfected population in this group. Integration is normally expected when selection pressure is applied, unless the plasmid transfected has sequences which are also found within the host genome, in which case homologous recombination may occur, thus leading to integration. The cMYC S/MAR used in this experiment was of murine origin, as was the C2C12 cell line. Therefore it is possible that integration had resulted due to recombination between the cMYC S/MAR sequence on the plasmid transfected and the cMYC S/MAR sequence within the *c-myc* proto-oncogene locus in the C2C12 cells' genomic DNA.

It is known that S/MARs can be found at the 5' and 3' ends of genes and also within introns (Schubeler *et al*., 1996), and Boulikas theorised that there are different classes of S/MARs, and that the ones that are transcribed are regulated by the cells (as described in Chapter 1: Introduction, Section 1.9) (Boulikas, 1995). S/MARs found within introns must clearly be transcribed and are not likely to obstruct transcription. But not all S/MARs are transcribed, such as the non-intronic variety, and they do not function as enhancers as Bode *et al* had demonstrated in transient expression assays (Bode *et al*., 1995). Bode *et al* stated that some S/MAR constructs tend to integrate into the genome at a higher rate than even their non-S/MAR controls as a result of the 'high recombinogenic potential' of their sequences (Bode *et al*., 1995). It is possible that due to the fact that the cMYC S/MAR is of murine origin, and is a non-intronic element naturally located upstream of the *c-myc* proto-oncogene promoter, that integration, if occurring, was most likely as a result of the high recombinogenic potential of the sequences. It is interesting to note that it was shown in Chapter 3 that pEPI-eGFP, the vector containing the intronic β-IFN S/MAR of human origin (Bode *et al*., 1995), was tested in the two human derived cell lines HepG2 and HeLa, the genomes of which no integration was observed.

7.3.4 Potential CMYC-pEPI integration into the C2C12 host genome did not lead to significant or improved long-term transgene expression as was observed for long-term transgene expression by potentially integrated CeGFP-C1

It is interesting to note that the integration that was observed of CMYC-pEPI into the C2C12 host genome in groups (A), (C), and (D) did nothing to improve the expression of the CMYC-pEPI population, where the percentage of positive cells and the total expression were all comparable to one another. However, CeGFP-C1 was found to potentially have interated into the host genome under condition (D), where the transfected cells were put under selection for 10 days, and this led to a significant improvement in the resulting total expression and percentage of positive cells of the population at the final day of the experiment, total expression was significantly higher than any of the populations of any of the groups (A)-(D), and the percentage of positive cells was significantly higher than all the populations except the CpEPI-eGFP population of group (C), where the percentages were similar and not significantly different to each other.

One explanation for this difference in expression, despite the fact that both vectors may have integrated into the host genome, is that integration of the CMYC-pEPI vector may have occurred in similar regions within each cell, which is plausible considering integration in the CMYC-pEPI transfected populations occurred without the requirement for selection pressure, whereas CeGFP-C1 integration only occurred due to selective pressure, as the evidence from the FISH data suggests. These regions of CMYC-pEPI integration may have not been favourable for the expression of the vector, thus leading to the same low level pattern of expression observed in the CMYC-pEPI populations of groups (A), (C), and (D). In a study by Heng *et al* plasmids containing reporter transgenes and S/MARs were integrated into the host genome of mouse and human cells in order to study positional effects of S/MARs on transgene expression within genomic DNA (Heng, *et al*., 2004). When gene expression was analysed it was found that not all the transgenes were being expressed but that the level of expression correlated with the number of transgenes associated with the matrix via an S/MAR. The expression of a transgene and the number of copies integrated do not show a linear relationship or a direct correlation. They also showed that the correlation that was observed between gene expression and the transgene's proximity to the matrix

highlights the significance of DNA's proximity and contact with the matrix for transcription, and it appears to be vital for transcription to take place. In this study they also found that when integrating 20-40 copies of the transgene into the genome, the expression profile was comparable to that where 1-2 copies were integrated. This is because cells are selective about which S/MARs will anchor to the matrix and so not every insert will be expressed (Heng *et al*., 2004). Thus, the position of the integrant is crucial, where the integration of multiple copies of a gene does not necessarily lead to an increase in gene expression.

7.3.5 Transgene expression was similar in CMYC-pEPI transfected populations in groups (A) – (D)

It is interesting to note that integration was not observed in the CMYC-pEPI transfected population under condition (B) and yet the expression profile of this population was comparable to that of the CMYC-pEPI populations under conditions (A), (C), and (D) where the vector had appeared to have integrated into the host genome. There are two possibilities which could explain this. The first is that expression of the integrants had been silenced and, considering this vector had also remained episomal within all the populations, the low and comparable level of expression exhibited in each population was due to expression resulting from these episomal copies. This would lead to the conclusion that none of conditions (A)-(D) affected expression and similar amounts of CMYC-pEPI vector remained episomally within C2C12 cells without the requirement, nor potential to improve expression/retention, with the use of selection, or holding in G0/G1 prior to selection.
The second possibility is that holding the CMYC-pEPI population in G0/G1 without selection had a more profound effect on expression and number of positive cells expressing the transgene, where the eGFP product was produced simply from the episomes, with no contribution from integrants, whereas the other populations produced eGFP from expression of both their episomal and their integrant copies. In this case, the conclusion would be that holding the CMYC-pEPI vector within cells in G0 for 7 days could be said to have a marked effect upon improving expression of the episomal vector. However, neither possibility can be determined from the data obtained from these experiments, and further analysis would be required to come to a

solid conclusion. Such further investigations may include the use of LAM PCR to identify the regions into which CMYC-pEPI had integrated into the genome in each population to identify whether the integrants were positioned in heterochromatin, leading to silencing of expression.

7.3.6 Holding CeGFP-C1, CpEPI-eGFP, and CMYC-pEPI transfected cells in G0/G1 prevented or decreased the incidence of vector integration into the C2C12 host genome

The group (B) transfected cells were held in G0/G1 phase of the cell cycle by using methionine-depleted low-serum medium for 7 days and the eGFP positive cells were then sorted, before allowing the cells to proliferate for 35 days. Interestingly, the FISH results suggested that no integration had occurred in the CeGFP-C1, CpEPI-eGFP, or CMYC-pEPI transfected populations, but that episomal copies were detected of each of the three vectors within their respective populations. This finding was even more surprising in that the CMYC-pEPI vector had shown evidence of integration even when no selection pressure had been applied, as in condition (A). It is unclear as to how holding the transfected cells in G0/G1 for 7 days before allowing the cells to proliferate for 35 days may lead to the prevention or decrease in the frequency of vector integration within the host genome. However, considering that the idea of holding transfected cells in G0/G1 was conducted in order to give C2C12 cells more time before cell replication to epigenetically mark the transfected vectors, it is possible that this marking was linked to the lack of integration observed. Even in group (C) where the transfected cells were put under selection for 10 days after being held in G0/G1 for 7 days, the CeGFP-C1 and CpEPI-eGFP vectors were found not to integrate into the host genome. This finding had also been observed in the previous chapter (Chapter 5). Although CMYC-pEPI had been evidenced to integrate into the host genome in the CMYC-pEPI transfected population of group (C), it is possible, as previously discussed, that this had occurred as a result of recombination due to a combination of the cMYC S/MAR's high recombinogenic potential and the imposition of selection, and may not occur with vectors that do not contain sequences with high recombinogenic potential in C2C12 cells, such as CeGFP-C1 and pEPI-eGFP. This highlights the importance of using sequences in vectors that do not have the potential to integrate into the host

genome if the vectors are to be safe for use in gene therapeutic applications. It also highlights the importance of further investigations to compare the epigenetic markers imposed on the transfected vectors, which may be different in transfected populations of different groups, and which may also differ between the vectors, as this may give an indication of the protective effect of holding transfected cells in G0/G1 prior to proliferation.

7.3.7 Transgene expression declined more significantly with time in populations where CeGFP-C1 integration was not detected

The CeGFP-C1 population of group (D), where the cells were put under selection for 10 days post transfection and CeGFP-C1 had potentially integrated into the host genome, had a significantly higher percentage of positive cells and total expression than the CeGFP-C1 populations of groups (A)-(C). As was shown from the PCR data, CeGFP-C1 vector copies were still present within the cell populations in the CeGFP-C1 populations of groups (A)-(C), and the FISH analysis indicated that the vector was present within cells of these populations as an episome with no evidence of integration. Therefore, it is possible that without CeGFP-C1 integration into the host genome, expression had declined much more significantly than where it had been found to have integrated, regardless of the condition the cells were put under post transfection, due to a loss of episome with cell division. However, since the possibility that the vector had integrated into the genome in all the populations tested cannot be excluded, it is possible then that the vector may have integrated into each of the populations, but had done so in regions which did not lead to transgene expression levels as high as those that were observed in the CeGFP-C1 population of group (D).

7.3.8 Holding CpEPI-eGFP transfected C2C12 Cells in G0/G1 for 7 days followed by selection for 10 days significantly improved long-term transgene expression and episomal vector retention over selection alone

The populations within group (C) were transfected then held in G0 for 7 days, followed by selection for 10 days, before allowing the cells to proliferate for 35 days. The flow cytometry data showed that a statistically significant larger percentage of positive cells

and total expression were observed in the Day 35 sample of the CpEPI-eGFP population of group (C) than the CpEPI-eGFP populations of groups (A), (B), and even (D), where the cells were transfected and put under selection for 10 days before allowing them to proliferate for 35 days. This indicates that putting a population of cells transfected with CpEPI-eGFP under the conditions of group (C) clearly improves the vector's expression and retention/maintenance in a higher percentage of the cells. The PCR data showed that CpEPI-eGFP vector copies were still present within the population by the final day of the experiment, and the FISH analysis of this population showed plasmid was present within the cells as episomes, with no evidence of integration. Without ruling out integration it can still be said that if integration did occur it was an infrequent event, considering none was detected.

In considering the fact that CpEPI-eGFP transfected significantly fewer cells than CeGFP-C1 again in this chapter (as was also demonstrated in Chapters 4 and 5), but it was demonstrated by flow cytometry that the percentage of positive cells within the CpEPI-eGFP transfected population of group (C) was significantly greater to that of the CeGFP-C1 population of group (C) by the final day of the experiment, and considering that the only difference between CpEPI-eGFP and CeGFP-C1 was the presence of the β-IFN S/MAR element in the former vector, it can be concluded that CpEPI-eGFP was being retained episomally more effectively in proliferating C2C12 cells than CeGFP-C1 was, and that the observed superior vector retention could be attributed to the combination of the use of the β-IFN S/MAR element in the vector, and the holding of this vector in cells in G0 post-transfection for 7 days, followed by a 10 day selection period, both of which may have allowed the vector to be marked effectively for episomal maintenance and retention.

Interestingly, it was seen in Chapter 5 that the percentage of positive cells and total expression of the CeGFP-C1 and CpEPI-eGFP populations held under such conditions as that of group (C) were not significantly different to one another, suggesting that holding both vectors under the conditions of group (C) had a similar effect on both vectors, whereas that was not the case in this chapter, as has been shown. This discrepancy between the results of Chapter 5 and this chapter regarding CeGFP-C1 may indicate that CeGFP-C1 may have not been maintained as an episome within C2C12 cells and had been diluted out of the population with cell proliferation. And, where a larger amount of CeGFP-C1 may have been initially transfected per cell in

Chapter 5 than may have been in the cells in this chapter, resulting in greater amounts of vector present within the cells by the final day of the experiment, leading to greater total expression which matched the expression and percentage of positive cells of the CpEPI-eGFP population, fewer copies may have initially been transfected in this chapter in comparison, thus leading to lower total expression and fewer numbers of positive cells. This would lead to the conclusion that holding cells transfected with CeGFP-C1 in G0/G1 for 7 days followed by selection for 10 days had no direct or significant effect on vector retention and/or transgene expression by CeGFP-C1.

7.3.9 CpEPI-eGFP population held in G0 for 7 days followed by selection for 10 days contained a significantly greater percentage of positive cells and expressed greater amounts of transgene than the same population put under selection alone post transfection.

A significant finding in this study was the confirmation that holding transfected cells in G0/G1 for up to 7 days prior to selection (CpEPI-eGFP population of group (C)) led to an improvement in vector retention and passing on to daughter cells in a dividing population of C2C12 myoblasts over using selection alone (CpEPI-eGFP population of group (D)). This may be an indication that slowing the cells in G0/G1 phase of a rapidly proliferating population could lead to an increased number of plasmid vectors being epigenetically marked, ultimately improving long-term transgene expression within the population.

7.3.10 CpEPI-eGFP population held in G0 for 7 days followed by selection for 10 days contained a similar percentage of eGFP positive cells by episomally retained CpEPI-eGFP as a population containing potentially integrated and episomal CeGFP-C1 vector

Holding cells transfected with CpEPI-eGFP in G0 for 7 days, followed by selection for 10 days before allowing the cells to proliferate significantly improved expression to the point where its population had a larger amount of positive cells than any of the other populations at the final day of the experiment, except for the CeGFP-C1 population of group (D). As it was shown that the CeGFP-C1 population of group (D) had had the CeGFP-C1 plasmid potentially integrated into the host genome in addition to being

episomal, whereas there was no evidence that the CpEPI-eGFP population of group (C) had had the CpEPI-eGFP plasmid integrate and that it had been found to be retained episomally, the positive effect of condition (C) on the CpEPIeGFP population was even more significant.

Chapter 8: Concluding Remarks

8.1

The results obtained in Chapter 3 indicate several things. First, pEPI-eGFP had not been found to integrate into any of the three cell lines: HeLa, C2C12, or HepG2. Second, it can be seen that the S/MAR plasmid is expressed and maintained differently according to cell type. Whereas plasmid expression was high in the HepG2 cells by the final day of the experiment, it was very low in the other two cell lines HeLa and C2C12. Variable expression of pEPI-eGFP between cell lines is in agreement with the study by Papapetrou *et al* (Papapetrou *et al*., 2006).

Expression levels appeared to improve in the C2C12 cells once the myoblasts were differentiated into myotubes, either as a result of eGFP accumulation within these non-dividing cells or due to different factors expressed as a result of differentiation that allow better expression of the plasmids, or perhaps even a combination of both these factors.

The mini-S/MAR plasmid tested in the C2C12 cells did not show sustained transgene expression and it was shown to be very low by the final day of the experiment.

In investigating the status of the vectors within the cells at the final day of the experiment evidence was found by FISH analysis for eGFP-C1 integration into the HeLa and C2C12 host genomes by the final day of the experiment. It was also found that the mini-S/MAR plasmid had potentially integrated into the C2C12 host genome. Therefore, these vectors pose a serious safety issue within these cell lines and would not be suitable for use as vectors for gene delivery. Interestingly, eGFP-C1 was found not to have integrated into the HepG2 host genome, indicating that selection does not always lead to the integration of a plasmid that may not contain an S/MAR element.

8.2

Results Chapter 5 aimed to improve transgene expression of the S/MAR plasmid by changing the promoter from CMV to CAGG, and again comparing this vector to the

control plasmid in C2C12 cells. The SIDD profiles generated for the CAGG vectors indicated a more open conformation of the plasmids at the promoter region than that of the vectors driven by the CMV promoter. This showed that the sequences that make up a vector could have a direct effect upon the configuration of the vector which in turn could directly affect transcription and thus expression. Despite the more open conformation it was shown that expression declined with time, again to very low levels, by the final day of the experiment. As in Chapter 3, no evidence of integration was detected of the CpEPI-eGFP plasmid into the C2C12 host genome, however evidence for the CeGFP-C1 plasmid's integration was found again. The transfection efficiency of CeGFP-C1 was ~3 fold greater than that of CpEPI-eGFP. Considering this, and that by the the final day of selection and final day of the experiment expression was similar, and considering the lack of integration of CpEPI-eGFP but the potential integration of CeGFP-C1, the CpEPI-eGFP vector could be said to have been retained and passed on more effectively in this cell line than CeGFP-C1.

8.3

The next results chapter, Chapter 6, also aimed to improve transgene expression of the S/MAR vector, by arresting the cells in G0/G1 for up to 7 days or quiescence for up to 10 days post-transfection with CeGFP-C1 or CpEPI-eGFP, prior to 10 days of selection, and then allowing them to proliferate. When quantified, the expression levels of the CpEPI-eGFP and the CeGFP-C1 plasmids in both the G0/G1 and the Reserve groups were similar to each other by the final day. Significantly, however, expression was even detected in the differentiated myotubes on the final day of the experiment in the CpEPI-eGFP G0/G1 and Reserve populations on the final day of the experiment using immunostaining, which had not been observed in the first two results chapters. Surprisingly, evidence towards CpEPI-eGFP integration into the host genome was found in the Reserve group, as it had for CeGFP-C1. The reason for the CpEPI-eGFP potential integration into the host genome in cells held under such conditions are not clear, and further investigations are required to identify possible causative events. It does, however, indicate that cells held in quiescence as reserve cells, achieved by the differentiation of a culture of myoblasts, may exist in a different state to those held in

G0/G1 by switching the medium to low serum methionine depleted medium. These differences are also yet to be elucidated, and require further investigation.

Conversely, no evidence of CpEPI-eGFP integration in the G0/G1 group was found, leading to the conclusion that arresting the cells in G0/G1 may have led to a greater amount of epigenetically marked vectors within the cells which led to an improvement in long-term expression. CeGFP-C1 was also found not to have integrated into the host genome in the G0/G1 group, which led to the theory that holding C2C12 cells in G0/G1 may confer protection from plasmid vector integration into the host genome.

It was further concluded in this chapter that CpEPI-eGFP was retained and passed on to daughter cells significantly more effectively by C2C12 cells than CeGFP-C1 as an episome in the G0/G1 group, as a result of the presence of the S/MAR element within the vector.

8.4

Chapter 7 had several aims. One aim was to test a vector containing the *c-myc* proto-oncogene S/MAR in place of the β-IFN S/MAR in C2C12 cells. It was found that expression levels of this vector declined sharply and resulted in very low and comparable percentages of positive cells and total levels of transgene expression by the final day of the experiment, regardless of the condition under which the transfected cells were put under (conditions (A)-(D)). FISH analysis indicated that this vector was present as an episome within the cells, but it is possible that, as with the CeGFP-C1 vector, that they were only passively maintained and were being diluted out of the population with cell divisions. Evidence was also found for the vector integration into the host genome in the cell populations of three of the four conditions: (A), (C), and (D). The fact that the vector had potentially integrated under condition (A), where no selection was used so the cells were not required to retain the vector in some form in order to be able to survive, indicated that the vector may have integrated as a result of recombination due to high similarities between its sequences and some found within the host genome's. It was interesting to note that no evidence of integration of this vector was found in transfected cells held in G0/G1 in group (B), where no selection was subsequently imposed. This may be further evidence to indicate that

holding cells in G0/G1 protects from vector integration into the host genome. CMYC-pEPI did however appear to have integrated into the host genome in group (C) where selection was imposed after holding in G0/G1, which may indicate that selection may have induced integration of the vector, counteracting the effects of holding in G0/G1. There is no direct evidence to support this, however, and further investigations are required in order to deduce any solid conclusions. Still, the fact that expression was not significantly different between the populations where the vector had potentially integrated and was present as episomes, and where it had not integrated and was only present as episomes, suggests two things. One, that the observable expression was a result of transgene expression resulting from the episomal copies of the vector, and two, that this vector was possibly integrating into similar regions within the genome as a result of highly recombinogenic sequences which were leading to transgene silencing. This is plausible when comparing this vector to CeGFP-C1, where its lack of integration into the host genome led to low levels of transgene expression by episomal vector copies, yet once evidence had pointed towards its integration, the levels of expression were significantly greater, whereas in the case of CMYC-pEPI, low levels of expression were present regardless of integration. Considering that the cMYC S/MAR was of murine origin, as were C2C12 cells, the possibility of integration due to recombination could not be excluded. This showed that not all S/MAR elements can confer the unique properties to the plasmids into which they are inserted as the β-IFN S/MAR has generally been able to, as has been demonstrated in this thesis and other studies (as reviewed in Chapter 1: Introduction, Section 1.10). Hence, a vector containing the cMYC S/MAR element has been found to be an unsuitable candidate for use as a vector for gene therapy.

The experiments in Chapter 6 also aimed to determine several other things. First, it was important to determine whether holding C2C12 cells post-transfection in G0/G1 without the imposition of selection would suffice in establishing CpEPI-eGFP or eGFP-C1 as episomes. It was found that this was not the case, as vector was diluted out of the populations with cell division. In fact, the percentages of positive cells and total levels of transgene expression were very low in the populations of this group (B), and were comparable to those of group (A), where no method of vector establishment was imposed. Hence these results made it clear that if this *in vitro* model was an accurate representation of what may occur *in vivo*, then activated satellite cells transfected with

either CeGFP-C1 or CpEPI-eGFP which had committed to differentiation and re-population of myocytes and myofibres would not retain these vectors, and these vectors would eventually either become diluted out of the population or remain present in very few copy numbers. These results may also indicate that vector copy dilution would also occur in those activated satellite cells that had been transfected prior to becoming withdrawn from the cell cycle for a period of time and then re-activating in response to injury, as being held in G0/G1, or quiescence, was not sufficient to establish the vectors.

The results in this chapter also showed that selection was essential for CpEPI-eGFP vector establishment, as the percentage of positive cells and total level of expression of this population (group (C)) was significantly greater than those of the CpEPI-eGFP populations where no selection had been imposed (groups (A) and (B)). And, once again, CpEPI-eGFP was found not to have integrated into the host genome.

Furthermore, it was found that selection imposed post-arrest in G0/G1 resulted in significantly improved vector establishment in a population of CpEPI-eGFP transfected cells, as the percent of positive cells and total expression in this group were significantly greater than those of the CpEPI-eGFP population where only selection had been imposed post-transfection (group (C)). As was shown in Chapter 5, no evidence of CpEPI-eGFP integration into the host genome was found, meaning that expression was predominantly, if not completely, from vector copies present as episomes within the cells.

When selection was not employed in establishing the CeGFP-C1 vector, similar results were seen as the CpEPI-eGFP vector where minimal expression was observed by the final day of the experiment. No integration was indicated by FISH in these two conditions (A & B), however the presence of some episome was detected within some of the cells of these populations. Integration is normally expected to occur under selection, therefore without it the CeGFP-C1 vector remained episomal but was slowly being diluted out of the populations with each successive replication.

As in Chapter 5, holding the cells in G0/G1 prior to selection led to a lack of CeGFP-C1 integration into the host genome, as no evidence of integration had been found. This further highlighted the protective effect of holding transfected cells in G0/G1 before imposing selection in the prevention of integration into the host genome. However, this group of cells did not seem to benefit otherwise from this method of vector

establishment, as the percentage of positive cells and total transgene expression were at low levels by the final day of the experiment, comparable to those of groups (A) and (B) where no selection had been imposed. This further indicated that the significant improvement in vector retention of the CpEPI-eGFP population held under the same conditions could be attributed to the presence of the β-IFN S/MAR element within the vector, as without it the vector was diluted out of the population.

The only CeGPFP-C1 transfected population of cells which showed comparable percentages of positive cells by the final day to CpEPI-eGFP of group (C) was that of group (D), where CeGFP-C1 had been found to potentially integrate into the host genome. This was a significant finding, indicating that the efficiency of retention of episomal vector CpEPI-eGFP was equivalent in a population of proliferating C2C12 myoblasts to that of potentially stably integrated CeGFP-C1 vector copies within these cells, which also contained non-integrated vector copies. However, it was also found that this CeGFP-C1 population of group (D) expressed, significantly, the greatest amount of total transgene of any of the groups tested on the final day of the experiment, which may be due to positional effects of the integrants within the genome, the number of transcribed integrants/episomes per cell, and/or to superior transgene expression as a result of reduced negative supercoiling affecting the accessibility of the promoter due to the presence of an S/MAR element.

Although an improvement in episomal vector retention and passing on to daughter cells by a novel method of vector establishment has been achieved in this study, the feasibility of using this method of establishment with this vector *in vivo* must be ascertained. One possible method would be the manipulation of autologous extracted precursor cells such as mesangioblasts *ex vivo,* followed by their expansion, before re-transplantation into the patient. This *ex vivo* manipulation would be imperative considering the use of selection would be required to establish the vector within the cells. It remains to be investigated whether the percentages of positive cells and hence the total expression achieved by this method would be sufficient to correct the pathology of a patient with muscular dystrophy, however it seems unlikely as there are many factors which must be considered.

In a study by Galvez *et al* approximately 50% of mesangioblasts grafted back into patients reached their target muscle cells (Galvez *et al*., 2006). In considering that a large amount of cell death ensued from arresting the cells in G0/G1 for up to 7 days in

this thesis, this method may not be suitable for ex-vivo application considering that the number of cells being reimplanted back into the patient would be very low. Following this, in a study conducted by Chamberlain, it was found that in order for a significant correction to be made, the majority of muscle fibres must be able to accumulate a minimum or 20% of wild type levels of functional dystrophin (Chamberlain, 1997). As shown in this thesis, the CpEPI-eGFP transfected positive myoblasts which were then differentiated into myotubes were able to express the eGFP transgene without silencing, where with time the intensity of the eGFP appeared to be increasing, suggesting an increase in transgene product with time. However, after cells were put under selection where 100% of the cells contained vector copies and were hence expected to be eGFP positive, it was found that a majority of them were not positive. Furthermore, after 35 days of cell proliferation, only approximately 2.5% of the CpEPI-eGFP transfected cells held in G0/G1 and then put under selection were found to be eGFP positive. Hence, if such a small percentage of satellite cells were able to retain the vector, then the proportion of muscle fibres regenerated post damage in MD patients by these positive cells would not be a majority, which would indicate that the use of this method with this vector may still require further improvement in order to elicit a significant correction of muscular pathology, as is required for dystrophic patients.

A second method of utilising this method for the purposes of gene therapy would be the direct delivery of the vector into the muscle, by injection or electroporation or other methods previously discussed (Chapter 1: Introduction, Section 1.6.2). Electroporation resulting in the transfection of satellite cells *in vivo* was found possible to achieve in a study by Peng *et al* (Peng *et al*., 2005). The delivery of this vector to satellite cells could also be achieved with the use of self-inactivating lentivirus vectors (SIN), which have been found not to integrate into the host genome and are able to effectively transduce quiescent cells (reviewed in Wanisch and Yanez-Munoz, 2009). Any of these delivery methods would then be required to be in combination with an *in vivo* selection system such as that developed by Wong *et al* (Wong *et al*., 2011). This system involved the insertion of a Bcl-2 transgene expression cassette within an S/MAR vector, whose expression within liver cells protected them from the apoptotic effects of a bi-weekly administered antibody, Jo2, a Fas antibody which, upon engaging with its antigen, leads the antigen to induce cell death (Wong *et al*., 2011). However,

this method poses serious safetly implications, and a selective advantage to dystrophin-expressing S/MAR vector-containing cells would be required to be developed in order to take this novel method of establishment further.

Chapter 9: Future Work

In light of the results obtained in this study, there are many questions that have arisen, leading to a multitude of potential routes of investigations that are not only interesting, but essential to follow in order to obtain a deeper understanding of the mechanisms by which cells operate in maintaining and expressing transgenes on plasmid DNA introduced into host cells. These routes would include an array of multidisciplinary techniques such as chromatin immunoprecipitaion (ChIP) assays, inverse PCR (IPCR), quantitative PCR (Q-PCR), the testing of tissue-specific promoters, the use of therapeutic transgenes, long term *in vivo* studies in animal models such as the *mdx* mouse, and the assessment of the therapeutic index/amelioration of dystrophic symptoms as a result.

ChIP is a technique used to advance our understanding of genome stability, gene expression, and how it is regulated by reversible and dynamic protein-chromatin interactions (Wong and Wei, 2009). This technique has been used in order to investigate transcription factor binding sites (Iyer *et al*., 2001), the binding of histones and other nuclear structural proteins to chromatin (Glynn *et al*., 2004), and the identification of chromatin-bound histone modifications that can occur within cells (Barski *et al*., 2007).

ChIP is carried out by cross-linking the DNA to the proteins bound, and also covalently linking proteins bound to other proteins by the use of UV light or formaldehyde (Wong and Wei, 2009). The chromatin is then sonicated in order to break it up into smaller fragments and then immunoprecipitated by the use of antibodies specific for the target nuclear proteins. The cross-linking can then be reversed, and the DNA and protein purified separately. The DNA sequence can then be identified by several methods including cloning followed by sequencing (Grandori *et al*., 1996), or by hybridisation methods such as Southern blotting (Orlando *et al*., 1997) or the use of microarrays that represent the entire genome (Iyer *et al*., 2001).

The application of this technique was used in the study by Riu *et al* in order to determine whether plasmid DNA transfected into cells was subject to chromatinisation and successfully identified both euchromatin- and heterochromatin-associated histone markings (Riu *et al*., 2004). It would be interesting to investigate the presence of such

markers at different time points for all the vectors tested in this thesis. It would be also be interesting to compare the differences between the markings of the S/MAR plasmid from the HepG2 cell samples and those of the HeLa and C2C12 samples. Furthermore, this method could identify the epigenetic changes that may have occurred to the S/MAR plasmid when being held in G0/G1 or in quiescence in reserve cells, and the differences that may exist between them and the control plasmid. It would also be worth investigating whether epigenetic marking does, in fact, take place at all if selection is not initially imposed upon cells, and whether this is, in part, the reason behind why the S/MAR plasmid did not appear to be established when selection was not used.

It would also be worth using other methods in order to screen for integrants in a multiclonal population of transfected cells. One such technique that is extremely sensitive is inverse PCR (IPCR). This allows the identification of integrated sequences within a genome, and can even be coupled with sequencing if the site of integration is required to be identified. By being able to totally exclude the possibility of the integration of the S/MAR plasmid, the safety profile of this vector would lead it to be a highly desirable tool in gene therapeutics. Furthermore, if it does integrate at low levels, it would be interesting to investigate whether the sites of integration are quite specific or if they occur in more random locations. Also, the successful application of IPCR using a sample from a multiclonal population, as opposed to a clonal population, means that a greater number of events are simultaneously being screened, leading to a more solid conclusion about the nature of the vector's possible integration patterns, or lack of.

It is also imperative to test other promoters. In this study two ubiquitous promoters were tested, however both resulted in declined expression. Argyros *et al* found that when using an S/MAR vector and having the transgene expression driven by a tissue-specific promoter (liver-specific AAT in their case), that the expression levels remained higher than the other ubiquitous promoter tested, and that the S/MAR element actually conferred protection for the promoter from being methylated as it was seen that the control plasmid's promoter had, in fact, been methylated (Argyros *et al*., 2009).

Another important experiment that would be useful to conduct is quantitative PCR (Q-PCR) in order to determine the amount of S/MAR plasmid present as an episome

within cells at several different time points. This would be useful to undertake after more conclusive evidence is gathered against integration of the vector into the genome, that way it would be clear that it is the number of episomes being calculated and not integrants. The information derived from such an experiment could show whether the episome numbers decline with time, and if so, the rate at which they may do so.

Another interesting experiment would be to attempt to reduce or eliminate the CpG sequences present on the S/MAR plasmid prior to assessment of expression. This may lead to longer term expression and avoid silencing.

Otherwise, the S/MAR element could be inserted into a minicircle vector devoid of a bacterial backbone, which may further enhance expression and its duration. This construct could also potentially be used in actively replicating cells due to the presence of the S/MAR element.

In gathering more information about the way in which cells epigenetically mark S/MAR vectors, and how these vectors are retained and passed onto progeny cells, further advances in the improvement of this vector's expression and safety profiles can be achieved, thus leading to a clinically viable vector that can be used to deliver therapeutic genes in order to correct, or at least ameliorate, pathologies of genetic diseases such as Muscular Dystrophy.

Chapter 10: Bibliography

Abdurashidova, G., Danailov, M.B., Ochem, A., Triolo, G., Djeliova, V., Radulescu, S., Vindigni, A., Riva, S., and Falaschi, A. (2003). Localization of proteins bound to a replication origin of human DNA along the cell cycle. EMBO J. *22,* 4294-4303.

Acsadi, G., Dickson, G., Love, D.R., Jani, A., Walsh, F.S., Gurusinghe, A., Wolff, J.A., and Davies, K.E. (1991). Human dystrophin expression in mdx mice after intramuscular injection of DNA constructs. Nature *352,* 815-818.

Adachi, Y., Kas, E., and Laemmli, U.K. (1989). Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. EMBO J. *8,* 3997-4006.

Adam, E., Kerkhofs, P., Mammerickx, M., Burny, A., Kettmann, R., and Willems, L. (1996). The CREB, ATF-1, and ATF-2 transcription factors from bovine leukemia virus-infected B lymphocytes activate viral expression. J. Virol. *70,* 1990-1999.

Adams, M.E., Butler, M.H., Dwyer, T.M., Peters, M.F., Murnane, A.A., and Froehner, S.C. (1993). Two forms of mouse syntrophin, a 58 kd dystrophin-associated protein, differ in primary structure and tissue distribution. Neuron *11,* 531-540.

Aden, D.P., Fogel, A., Plotkin, S., Damjanov, I., and Knowles, B.B. (1979). Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. Nature *282,* 615-616.

Alexopoulou, A.N., Couchman, J.R., and Whiteford, J.R. (2008). The CMV early enhancer/chicken beta actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors. BMC Cell Biol. *9,* 2.

Alter, J., Lou, F., Rabinowitz, A., Yin, H., Rosenfeld, J., Wilton, S.D., Partridge, T.A., and Lu, Q.L. (2006). Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. Nat. Med. *12,* 175-177.

Anderson, J.T., Rogers, R.P., and Jarrett, H.W. (1996). Ca2+-calmodulin binds to the carboxyl-terminal domain of dystrophin. J. Biol. Chem. *271,* 6605-6610.

Andres, V., and Walsh, K. (1996). Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis. J. Cell Biol. *132,* 657-666.

Araishi, K., Sasaoka, T., Imamura, M., Noguchi, S., Hama, H., Wakabayashi, E., Yoshida, M., Hori, T., and Ozawa, E. (1999). Loss of the sarcoglycan complex and sarcospan leads to muscular dystrophy in beta-sarcoglycan-deficient mice. Hum. Mol. Genet. *8,* 1589-1598.

Arechavala-Gomeza, V., Graham, I.R., Popplewell, L.J., Adams, A.M., Aartsma-Rus, A., Kinali, M., Morgan, J.E., van Deutekom, J.C., Wilton, S.D., Dickson, G., and Muntoni, F. (2007). Comparative analysis of antisense oligonucleotide sequences for targeted skipping of exon 51 during dystrophin pre-mRNA splicing in human muscle. Hum. Gene Ther. *18,* 798-810.

Argyros, O., Wong, S.P., Niceta, M., Waddington, S.N., Howe, S.J., Coutelle, C., Miller, A.D., and Harbottle, R.P. (2008). Persistent episomal transgene expression in liver following delivery of a scaffold/matrix attachment region containing non-viral vector. Gene Ther. *15,* 1593-1605.

Argyros, O., Wong, S.P., Fedonidis, C., Tolmachov, O., Waddington, S.N., Howe, S.J., Niceta, M., Coutelle, C., and Harbottle, R.P. (2011) Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver. J. Mol. Med. *89*, 515-529.

Asakura, A., Seale, P., Girgis-Gabardo, A., and Rudnicki, M.A. (2002). Myogenic specification of side population cells in skeletal muscle. J. Cell Biol. *159,* 123-134.

ATCHISON, R.W., CASTO, B.C., and HAMMON, W.M. (1965). Adenovirus-Associated Defective Virus Particles. Science *149,* 754-756.

Azzoni, A.R., Ribeiro, S.C., Monteiro, G.A., and Prazeres, D.M. (2007). The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. J. Gene Med. *9,* 392-402.

Bachrach, E., Li, S., Perez, A.L., Schienda, J., Liadaki, K., Volinski, J., Flint, A., Chamberlain, J., and Kunkel, L.M. (2004). Systemic delivery of human microdystrophin to regenerating mouse dystrophic muscle by muscle progenitor cells. Proc. Natl. Acad. Sci. U. S. A. *101,* 3581-3586.

Baeuerle, P.A., and Henkel, T. (1994). Function and activation of NF-kappa B in the immune system. Annu. Rev. Immunol. *12,* 141-179.

Baiker, A., Maercker, C., Piechaczek, C., Schmidt, S.B., 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. Nat. Cell Biol. *2,* 182-184.

Barker, P.E. (1982). Double minutes in human tumor cells. Cancer Genet. Cytogenet. *5,* 81-94.

Barnea, E., Zuk, D., Simantov, R., Nudel, U., and Yaffe, D. (1990). Specificity of expression of the muscle and brain dystrophin gene promoters in muscle and brain cells. Neuron *5,* 881-888.

Barresi, R., Moore, S.A., Stolle, C.A., Mendell, J.R., and Campbell, K.P. (2000). Expression of gamma - sarcoglycan in smooth muscle and its interaction with the smooth muscle sarcoglycan-sarcospan complex. J. Biol. Chem. *275,* 38554-38560.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129,* 823-837.

Beard, B.C., Keyser, K.A., Trobridge, G.D., Peterson, L.J., Miller, D.G., Jacobs, M., Kaul, R., and Kiem, H.P. (2007). Unique integration profiles in a canine model of long-term repopulating cells transduced with gammaretrovirus, lentivirus, or foamy virus. Hum. Gene Ther. *18,* 423-434.

Beauchamp, J.R., Heslop, L., Yu, D.S., Tajbakhsh, S., Kelly, R.G., Wernig, A., Buckingham, M.E., Partridge, T.A., and Zammit, P.S. (2000). Expression of CD34 and Myf5 defines the majority of quiescent adult skeletal muscle satellite cells. J. Cell Biol. *151,* 1221-1234.

Bell, S.P., and Dutta, A. (2002). DNA replication in eukaryotic cells. Annu. Rev. Biochem. *71,* 333-374.

Benson, M.A., Newey, S.E., Martin-Rendon, E., Hawkes, R., and Blake, D.J. (2001). Dysbindin, a novel coiled-coil-containing protein that interacts with the dystrobrevins in muscle and brain. J. Biol. Chem. *276,* 24232-24241.

Bertoni, C., Morris, G.E., and Rando, T.A. (2005). Strand bias in oligonucleotide-mediated dystrophin gene editing. Hum. Mol. Genet. *14,* 221-233.

Bettan, M., Emmanuel, F., Darteil, R., Caillaud, J.M., Soubrier, F., Delaere, P., Branelec, D., Mahfoudi, A., Duverger, N., and Scherman, D. (2000). High-level protein secretion into blood circulation after electric pulse-mediated gene transfer into skeletal muscle. Mol. Ther. *2,* 204-210.

Bi, C., and Benham, C.J. (2004). WebSIDD: server for predicting stress-induced duplex destabilized (SIDD) sites in superhelical DNA. Bioinformatics *20,* 1477-1479.

Bickmore, W.A., and Oghene, K. (1996). Visualizing the spatial relationships between defined DNA sequences and the axial region of extracted metaphase chromosomes. Cell *84,* 95-104.

Bies, R.D., Phelps, S.F., Cortez, M.D., Roberts, R., Caskey, C.T., and Chamberlain, J.S. (1992). Human and murine dystrophin mRNA transcripts are differentially expressed during skeletal muscle, heart, and brain development. Nucleic Acids Res. *20,* 1725-1731.

Blake, D.J., and Kroger, S. (2000). The neurobiology of duchenne muscular dystrophy: learning lessons from muscle? Trends Neurosci. *23,* 92-99.

Blake, D.J., Nawrotzki, R., Peters, M.F., Froehner, S.C., and Davies, K.E. (1996). Isoform diversity of dystrobrevin, the murine 87-kDa postsynaptic protein. J. Biol. Chem. *271,* 7802-7810.

Blake, D.J., Schofield, J.N., Zuellig, R.A., Gorecki, D.C., Phelps, S.R., Barnard, E.A., Edwards, Y.H., and Davies, K.E. (1995). G-utrophin, the autosomal homologue of dystrophin Dp116, is expressed in sensory ganglia and brain. Proc. Natl. Acad. Sci. U. S. A. *92,* 3697-3701.

Blake, D.J., Tinsley, J.M., Davies, K.E., Knight, A.E., Winder, S.J., and Kendrick-Jones, J. (1995). Coiled-coil regions in the carboxy-terminal domains of dystrophin and related proteins: potentials for protein-protein interactions. Trends Biochem. Sci. *20,* 133-135.

Blake, D.J., Weir, A., Newey, S.E., and Davies, K.E. (2002). Function and genetics of dystrophin and dystrophin-related proteins in muscle. Physiol. Rev. *82,* 291-329.

Blaveri, K., Heslop, L., Yu, D.S., Rosenblatt, J.D., Gross, J.G., Partridge, T.A., and Morgan, J.E. (1999). Patterns of repair of dystrophic mouse muscle: studies on isolated fibers. Dev. Dyn. *216,* 244-256.

Blow, J.J., and Hodgson, B. (2002). Replication licensing--defining the proliferative state? Trends Cell Biol. *12,* 72-78.

Bode, J., Benham, C., Knopp, A., and Mielke, C. (2000). Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). Crit. Rev. Eukaryot. Gene Expr. *10,* 73-90.

Bode, J., Goetze, S., Heng, H., Krawetz, S.A., and Benham, C. (2003). From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. Chromosome Res. *11,* 435-445.

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 DNAs. Science *255,* 195-197.

Bode, J., Schlake, T., Rios-Ramirez, M., Mielke, C., Stengert, M., Kay, V., and Klehr-Wirth, D. (1995). Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. Int. Rev. Cytol. *162A,* 389-454.

Bode, J., Winkelmann, S., Gotze, S., Spiker, S., Tsutsui, K., Bi, C., A, K.P., and Benham, C. (2006). Correlations between scaffold/matrix attachment region (S/MAR) binding activity and DNA duplex destabilization energy. J. Mol. Biol. *358,* 597-613.

Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B., and Schaffner, W. (1985). A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell *41,* 521-530.

Bouchentouf, M., Benabdallah, B.F., Bigey, P., Yau, T.M., Scherman, D., and Tremblay, J.P. (2008). Vascular endothelial growth factor reduced hypoxia-induced death of human myoblasts and improved their engraftment in mouse muscles. Gene Ther. *15,* 404-414.

Boulikas, T. (1995). Chromatin domains and prediction of MAR sequences. Int. Rev. Cytol. *162A,* 279-388.

Bradley, W.G., Hudgson, P., Larson, P.F., Papapetropoulos, T.A., and Jenkison, M. (1972). Structural changes in the early stages of Duchenne muscular dystrophy. J. Neurol. Neurosurg. Psychiatry. *35,* 451-455.

Bremmer-Bout, M., Aartsma-Rus, A., de Meijer, E.J., Kaman, W.E., Janson, A.A., Vossen, R.H., van Ommen, G.J., den Dunnen, J.T., and van Deutekom, J.C. (2004). Targeted exon skipping in transgenic hDMD mice: A model for direct preclinical screening of human-specific antisense oligonucleotides. Mol. Ther. *10,* 232-240.

Brown, TA (2002). *Genomes, 2nd Edition*. Oxford: Bios Scientific Publishers

Bruckert, P., Kappler, R., Scherthan, H., Link, H., Hagmann, F., and Zankl, H. (2000). Double minutes and c-MYC amplification in acute myelogenous leukemia: Are they prognostic factors? Cancer Genet. Cytogenet. *120,* 73-79.

Burton, E.A., Tinsley, J.M., Holzfeind, P.J., Rodrigues, N.R., and Davies, K.E. (1999). A second promoter provides an alternative target for therapeutic up-regulation of utrophin in Duchenne muscular dystrophy. Proc. Natl. Acad. Sci. U. S. A. *96,* 14025-14030.

Bushby, K.M. (1999). The limb-girdle muscular dystrophies-multiple genes, multiple mechanisms. Hum. Mol. Genet. *8,* 1875-1882.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., and Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. Nat. Rev. Microbiol. *3,* 848-858.

Cai, S., Han, H.J., and Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. Nat. Genet. *34,* 42-51.

Chamberlain, J.R., Schwarze, U., Wang, P.R., Hirata, R.K., Hankenson, K.D., Pace, J.M., Underwood, R.A., Song, K.M., Sussman, M., Byers, P.H., and Russell, D.W. (2004). Gene targeting in stem cells from individuals with osteogenesis imperfecta. Science *303,* 1198-1201.

Chan, Y.M., Bonnemann, C.G., Lidov, H.G., and Kunkel, L.M. (1998). Molecular organization of sarcoglycan complex in mouse myotubes in culture. J. Cell Biol. *143,* 2033-2044.

Chang, A.H., Stephan, M.T., and Sadelain, M. (2006). Stem cell-derived erythroid cells mediate long-term systemic protein delivery. Nat. Biotechnol. *24,* 1017-1021.

Chang, S.C., Tucker, T., Thorogood, N.P., and Brown, C.J. (2006). Mechanisms of X-chromosome inactivation. Front. Biosci. *11,* 852-866.

Chelly, J., Hamard, G., Koulakoff, A., Kaplan, J.C., Kahn, A., and Berwald-Netter, Y. (1990). Dystrophin gene transcribed from different promoters in neuronal and glial cells. Nature *344,* 64-65.

Chen, H.H., Mack, L.M., Kelly, R., Ontell, M., Kochanek, S., and Clemens, P.R. (1997). Persistence in muscle of an adenoviral vector that lacks all viral genes. Proc. Natl. Acad. Sci. U. S. A. *94,* 1645-1650.

Chen, Z.Y., He, C.Y., Ehrhardt, A., and Kay, M.A. (2003). Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo. Mol. Ther. *8,* 495-500.

Cherng, J.Y., Schuurmans-Nieuwenbroek, N.M., Jiskoot, W., Talsma, H., Zuidam, N.J., Hennink, W.E., and Crommelin, D.J. (1999). Effect of DNA topology on the transfection efficiency of poly((2-dimethylamino)ethyl methacrylate)-plasmid complexes. J. Control. Release *60,* 343-353.

Chung, S., Andersson, T., Sonntag, K.C., Bjorklund, L., Isacson, O., and Kim, K.S. (2002). Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines. Stem Cells *20,* 139-145.

Ciuffi, A., Mitchell, R.S., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F.D. (2006). Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. Mol. Ther. *13,* 366-373.

Cole, M.D. (1986). The myc oncogene: its role in transformation and differentiation. Annu. Rev. Genet. *20,* 361-384.

Collins, C.A., Olsen, I., Zammit, P.S., Heslop, L., Petrie, A., Partridge, T.A., and Morgan, J.E. (2005). Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. Cell *122,* 289-301.

Collins, J.M. (1978). Rates of DNA synthesis during the S-phase of HeLa cells. J. Biol. Chem. *253,* 8570-8577.

Comb, M., and Goodman, H.M. (1990). CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. Nucleic Acids Res. *18,* 3975-3982.

Compton, D.A., and Cleveland, D.W. (1993). NuMA is required for the proper completion of mitosis. J. Cell Biol. *120,* 947-957.

Conboy, I.M., and Rando, T.A. (2002). The regulation of Notch signaling controls satellite cell activation and cell fate determination in postnatal myogenesis. Dev. Cell. *3,* 397-409.

Conese, M., Auriche, C., and Ascenzioni, F. (2004). Gene therapy progress and prospects: episomally maintained self-replicating systems. Gene Ther. *11,* 1735-1741.

Cook, P.R. (1999). The organization of replication and transcription. Science *284,* 1790-1795.

Cordier, L., Gao, G.P., Hack, A.A., McNally, E.M., Wilson, J.M., Chirmule, N., and Sweeney, H.L. (2001). Muscle-specific promoters may be necessary for adeno-associated virus-mediated gene transfer in the treatment of muscular dystrophies. Hum. Gene Ther. *12,* 205-215.

Cornelison, D.D., and Wold, B.J. (1997). Single-cell analysis of regulatory gene expression in quiescent and activated mouse skeletal muscle satellite cells. Dev. Biol. *191,* 270-283.

Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O., and Debatisse, M. (2008). Replication fork movement sets chromatin loop size and origin choice in mammalian cells. Nature *455,* 557-560.

Crosbie, R.H., Heighway, J., Venzke, D.P., Lee, J.C., and Campbell, K.P. (1997). Sarcospan, the 25-kDa transmembrane component of the dystrophin-glycoprotein complex. J. Biol. Chem. *272,* 31221-31224.

Cusella De Angelis, M.G., Balconi, G., Bernasconi, S., Zanetta, L., Boratto, R., Galli, D., Dejana, E., and Cossu, G. (2003). Skeletal myogenic progenitors in the endothelium of lung and yolk sac. Exp. Cell Res. *290,* 207-216.

Dai, Y., Roman, M., Naviaux, R.K., and Verma, I.M. (1992). Gene therapy via primary myoblasts: long-term expression of factor IX protein following transplantation in vivo. Proc. Natl. Acad. Sci. U. S. A. *89,* 10892-10895.

Daley, G.Q., Van Etten, R.A., and Baltimore, D. (1990). Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science *247,* 824-830.

Dang, Q., Auten, J., and Plavec, I. (2000). Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector. J. Virol. *74,* 2671-2678.

Davis, G.L., and Hawrisiak, M.M. (1977). Experimental cytomegalovirus infection and the developing mouse inner ear: in vivo and in vitro studies. Lab. Invest. *37,* 20-29.

De Angelis, L., Berghella, L., Coletta, M., Lattanzi, L., Zanchi, M., Cusella-De Angelis, M.G., Ponzetto, C., and Cossu, G. (1999). Skeletal myogenic progenitors originating from embryonic dorsal aorta coexpress endothelial and myogenic markers and contribute to postnatal muscle growth and regeneration. J. Cell Biol. *147,* 869-878.

De Matteis, M.A., and Morrow, J.S. (2000). Spectrin tethers and mesh in the biosynthetic pathway. J. Cell. Sci. *113 (Pt 13),* 2331-2343.

Deato, M.D., and Tjian, R. (2007). Switching of the core transcription machinery during myogenesis. Genes Dev. *21,* 2137-2149.

Deconinck, A.E., Rafael, J.A., Skinner, J.A., Brown, S.C., Potter, A.C., Metzinger, L., Watt, D.J., Dickson, J.G., Tinsley, J.M., and Davies, K.E. (1997). Utrophin-dystrophin-deficient mice as a model for Duchenne muscular dystrophy. Cell *90,* 717-727.

Deconinck, N., Tinsley, J., De Backer, F., Fisher, R., Kahn, D., Phelps, S., Davies, K., and Gillis, J.M. (1997). Expression of truncated utrophin leads to major functional improvements in dystrophin-deficient muscles of mice. Nat. Med. *3,* 1216-1221.

Dennis, C.L., Tinsley, J.M., Deconinck, A.E., and Davies, K.E. (1996). Molecular and functional analysis of the utrophin promoter. Nucleic Acids Res. *24,* 1646-1652.

Dickinson, L.A., and Kohwi-Shigematsu, T. (1995). Nucleolin is a matrix attachment region DNA-binding protein that specifically recognizes a region with high base-unpairing potential. Mol. Cell. Biol. *15,* 456-465.

Djinovic-Carugo, K., Gautel, M., Ylanne, J., and Young, P. (2002). The spectrin repeat: a structural platform for cytoskeletal protein assemblies. FEBS Lett. *513,* 119-123.

Dreyfus, P.A., Chretien, F., Chazaud, B., Kirova, Y., Caramelle, P., Garcia, L., Butler-Browne, G., and Gherardi, R.K. (2004). Adult bone marrow-derived stem cells in muscle connective tissue and satellite cell niches. Am. J. Pathol. *164,* 773-779.

Duan, D. (2006). From the smallest virus to the biggest gene: marching towards gene therapy for duchenne muscular dystrophy. Discov. Med. *6,* 103-108.

Duan, D., Sharma, P., Yang, J., Yue, Y., Dudus, L., Zhang, Y., Fisher, K.J., and Engelhardt, J.F. (1998). Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. J. Virol. *72,* 8568-8577.

Dubowitz, V. (1978). Muscle disorders in childhood. Major Probl. Clin. Pediatr. *16,* iii-xiii, 1-282.

DuBridge, R.B., and Calos, M.P. (1988). Recombinant shuttle vectors for the study of mutation in mammalian cells. Mutagenesis *3,* 1-9.

Dunaway, M., and Ostrander, E.A. (1993). Local domains of supercoiling activate a eukaryotic promoter in vivo. Nature *361,* 746-748.

Dunckley, M.G., Manoharan, M., Villiet, P., Eperon, I.C., and Dickson, G. (1998). Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides. Hum. Mol. Genet. *7,* 1083-1090.

Dyson, P.J., Littlewood, T.D., Forster, A., and Rabbitts, T.H. (1985). Chromatin structure of transcriptionally active and inactive human c-myc alleles. EMBO J. *4,* 2885-2891.

Evans, K., Ott, S., Hansen, A., Koentges, G., and Wernisch, L. (2007). A comparative study of S/MAR prediction tools. BMC Bioinformatics *8,* 71.

Fackelmayer, F.O., and Richter, A. (1994). Purification of two isoforms of hnRNP-U and characterization of their nucleic acid binding activity. Biochemistry *33,* 10416-10422.

Feener, C.A., Koenig, M., and Kunkel, L.M. (1989). Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus. Nature *338,* 509-511.

Ferrari, G., Stornaiuolo, A., and Mavilio, F. (2001). Failure to correct murine muscular dystrophy. Nature *411,* 1014-1015.

Fisher, K.J., Jooss, K., Alston, J., Yang, Y., Haecker, S.E., High, K., Pathak, R., Raper, S.E., and Wilson, J.M. (1997). Recombinant adeno-associated virus for muscle directed gene therapy. Nat. Med. *3,* 306-312.

Fletcher, S., Honeyman, K., Fall, A.M., Harding, P.L., Johnsen, R.D., Steinhaus, J.P., Moulton, H.M., Iversen, P.L., and Wilton, S.D. (2007). Morpholino oligomer-mediated exon skipping averts the onset of dystrophic pathology in the mdx mouse. Mol. Ther. *15,* 1587-1592.

Foster, H., Sharp, P.S., Athanasopoulos, T., Trollet, C., Graham, I.R., Foster, K., Wells, D.J., and Dickson, G. (2008). Codon and mRNA sequence optimization of microdystrophin transgenes improves expression

and physiological outcome in dystrophic mdx mice following AAV2/8 gene transfer. Mol. Ther. *16,* 1825-1832.

Foster, K., Foster, H., and Dickson, J.G. (2006). Gene therapy progress and prospects: Duchenne muscular dystrophy. Gene Ther. *13,* 1677-1685.

Frisch, M., Frech, K., Klingenhoff, A., Cartharius, K., Liebich, I., and Werner, T. (2002). In silico prediction of scaffold/matrix attachment regions in large genomic sequences. Genome Res. *12,* 349-354.

Fuchtbauer, E.M., and Westphal, H. (1992). MyoD and myogenin are coexpressed in regenerating skeletal muscle of the mouse. Dev. Dyn. *193,* 34-39.

Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat. Genet. *24,* 88-91.

Furth, P.A., Hennighausen, L., Baker, C., Beatty, B., and Woychick, R. (1991). The variability in activity of the universally expressed human cytomegalovirus immediate early gene 1 enhancer/promoter in transgenic mice. Nucleic Acids Res. *19,* 6205-6208.

Galvez, B.G., Sampaolesi, M., Brunelli, S., Covarello, D., Gavina, M., Rossi, B., Constantin, G., Torrente, Y., and Cossu, G. (2006). Complete repair of dystrophic skeletal muscle by mesoangioblasts with enhanced migration ability. J. Cell Biol. *174,* 231-243.

Gan, L., Wessel, G.M., and Klein, W.H. (1990). Regulatory elements from the related spec genes of Strongylocentrotus purpuratus yield different spatial patterns with a lacZ reporter gene. Dev. Biol. *142,* 346-359.

Gao, X., Kim, K.S., and Liu, D. (2007). Nonviral gene delivery: what we know and what is next. AAPS J. *9,* E92-104.

Gardner,M.J. (2001) A status report on the sequencing and annotation of the P. falciparum genome. Mol. Biochem. Parasitol. *118*, 133–138.

Garg, S., Oran, A.E., Hon, H., and Jacob, J. (2004). The hybrid cytomegalovirus enhancer/chicken beta-actin promoter along with woodchuck hepatitis virus posttranscriptional regulatory element enhances the protective efficacy of DNA vaccines. J. Immunol. *173,* 550-558.

Gargioli, C., Coletta, M., De Grandis, F., Cannata, S.M., and Cossu, G. (2008). PlGF-MMP-9-expressing cells restore microcirculation and efficacy of cell therapy in aged dystrophic muscle. Nat. Med. *14,* 973-978.

Garrard, W.T. (1991). Histone H1 and the conformation of transcriptionally active chromatin. Bioessays *13,* 87-88.

Gebski, B.L., Mann, C.J., Fletcher, S., and Wilton, S.D. (2003). Morpholino antisense oligonucleotide induced dystrophin exon 23 skipping in mdx mouse muscle. Hum. Mol. Genet. *12,* 1801-1811.

Germann, WJ and Stanfield, CL (2001). *Principles of Human Physiology.* San Francisco, CA: Benjamin Cummings

Ghazal, P., Lubon, H., Fleckenstein, B., and Hennighausen, L. (1987). Binding of transcription factors and creation of a large nucleoprotein complex on the human cytomegalovirus enhancer. Proc. Natl. Acad. Sci. U. S. A. *84,* 3658-3662.

Ghazal, P., Lubon, H., and Hennighausen, L. (1988). Specific interactions between transcription factors and the promoter-regulatory region of the human cytomegalovirus major immediate-early gene. J. Virol. *62,* 1076-1079.

Ghosh, A., Yue, Y., Long, C., Bostick, B., and Duan, D. (2007). Efficient whole-body transduction with trans-splicing adeno-associated viral vectors. Mol. Ther. *15,* 750-755.

Giannakopoulos, A., Stavrou, E.F., Zarkadis, I., Zoumbos, N., Thrasher, A.J., and Athanassiadou, A. (2009). The functional role of S/MARs in episomal vectors as defined by the stress-induced destabilization profile of the vector sequences. J. Mol. Biol. *387,* 1239-1249.

Gilbert, D.M. (2002). Replication timing and transcriptional control: beyond cause and effect. Curr. Opin. Cell Biol. *14,* 377-383.

Gilbert, D.M. (2001). Making sense of eukaryotic DNA replication origins. Science *294,* 96-100.

Gill, D.R., Pringle, I.A., and Hyde, S.C. (2009). Progress and prospects: the design and production of plasmid vectors. Gene Ther. *16,* 165-171.

Girard-Reydet, C., Gregoire, D., Vassetzky, Y., and Mechali, M. (2004). DNA replication initiates at domains overlapping with nuclear matrix attachment regions in the xenopus and mouse c-myc promoter. Gene *332,* 129-138.

Glockner,G., Eichinger,L., Szafranski,K., Pachebat,J.A., Bankier,A.T., Dear,P.H., Lehmann,R., Baumgart,C., Parra,G., Abril,J.F. et al. (2002) Sequence and analysis of chromosome 2 of Dictyostelium discoideum. Nature *418*, 79–85.

Glover, D.J., Lipps, H.J., and Jans, D.A. (2005). Towards safe, non-viral therapeutic gene expression in humans. Nat. Rev. Genet. *6,* 299-310.

Glynn, E.F., Megee, P.C., Yu, H.G., Mistrot, C., Unal, E., Koshland, D.E., DeRisi, J.L., and Gerton, J.L. (2004). Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. PLoS Biol. *2,* E259.

Goetze, S., Baer, A., Winkelmann, S., Nehlsen, K., Seibler, J., Maass, K., and Bode, J. (2005). Performance of genomic bordering elements at predefined genomic loci. Mol. Cell. Biol. *25,* 2260-2272.

Gorecki, D.C., Monaco, A.P., Derry, J.M., Walker, A.P., Barnard, E.A., and Barnard, P.J. (1992). Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. Hum. Mol. Genet. *1,* 505-510.

Gossett, L.A., Kelvin, D.J., Sternberg, E.A., and Olson, E.N. (1989). A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. Mol. Cell. Biol. *9,* 5022-5033.

Gramolini, A.O., Karpati, G., and Jasmin, B.J. (1999). Discordant expression of utrophin and its transcript in human and mouse skeletal muscles. J. Neuropathol. Exp. Neurol. *58,* 235-244.

Grandori, C., Mac, J., Siebelt, F., Ayer, D.E., and Eisenman, R.N. (1996). Myc-Max heterodimers activate a DEAD box gene and interact with multiple E box-related sites in vivo. EMBO J. *15,* 4344-4357.

Grant, P.A. (2001). A tale of histone modifications. Genome Biol. *2,* REVIEWS0003.

Graves, D.C., and Yablonka-Reuveni, Z. (2000). Vascular smooth muscle cells spontaneously adopt a skeletal muscle phenotype: a unique Myf5(-)/MyoD(+) myogenic program. J. Histochem. Cytochem. *48,* 1173-1193.

Gregorevic, P., Blankinship, M.J., Allen, J.M., Crawford, R.W., Meuse, L., Miller, D.G., Russell, D.W., and Chamberlain, J.S. (2004). Systemic delivery of genes to striated muscles using adeno-associated viral vectors. Nat. Med. *10,* 828-834.

Grimm, S., and Baeuerle, P.A. (1993). The inducible transcription factor NF-kappa B: structure-function relationship of its protein subunits. Biochem. J. *290 (Pt 2),* 297-308.

Gronevik, E., von Steyern, F.V., Kalhovde, J.M., Tjelle, T.E., and Mathiesen, I. (2005). Gene expression and immune response kinetics using electroporation-mediated DNA delivery to muscle. J. Gene Med. *7,* 218-227.

Groth, A., Rocha, W., Verreault, A., and Almouzni, G. (2007). Chromatin challenges during DNA replication and repair. Cell *128,* 721-733.

Grounds, M.D., and Davies, K.E. (2007). The allure of stem cell therapy for muscular dystrophy. Neuromuscul. Disord. *17,* 206-208.

Grounds, M.D., Garrett, K.L., Lai, M.C., Wright, W.E., and Beilharz, M.W. (1992). Identification of skeletal muscle precursor cells in vivo by use of MyoD1 and myogenin probes. Cell Tissue Res. *267,* 99-104.

Grounds, M.D., White, J.D., Rosenthal, N., and Bogoyevitch, M.A. (2002). The role of stem cells in skeletal and cardiac muscle repair. J. Histochem. Cytochem. *50,* 589-610.

Gussoni, E., Soneoka, Y., Strickland, C.D., Buzney, E.A., Khan, M.K., Flint, A.F., Kunkel, L.M., and Mulligan, R.C. (1999). Dystrophin expression in the mdx mouse restored by stem cell transplantation. Nature *401,* 390-394.

Guttridge, D.C., Albanese, C., Reuther, J.Y., Pestell, R.G., and Baldwin, A.S.,Jr. (1999). NF-kappaB controls cell growth and differentiation through transcriptional regulation of cyclin D1. Mol. Cell. Biol. *19,* 5785-5799.

Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M., and Fischer, A. (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. N. Engl. J. Med. *348,* 255-256.

Hacker, H., Mischak, H., Miethke, T., Liptay, S., Schmid, R., Sparwasser, T., Heeg, K., Lipford, G.B., and Wagner, H. (1998). CpG-DNA-specific activation of antigen-presenting cells requires stress kinase activity and is preceded by non-specific endocytosis and endosomal maturation. EMBO J. *17,* 6230-6240.

Hanshaw, J.B., Scheiner, A.P., Moxley, A.W., Gaev, L., Abel, V., and Scheiner, B. (1976). School failure and deafness after "silent" congenital cytomegalovirus infection. N. Engl. J. Med. *295,* 468-470.

Harikrishnan, K.N., Chow, M.Z., Baker, E.K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J.M., Jones, P.L., Sif, S., and El-Osta, A. (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat. Genet. *37,* 254-264.

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., and Akira, S. (2000). A Toll-like receptor recognizes bacterial DNA. Nature *408,* 740-745.

Henegariu, O., Heerema, N.A., Lowe Wright, L., Bray-Ward, P., Ward, D.C., and Vance, G.H. (2001). Improvements in cytogenetic slide preparation: controlled chromosome spreading, chemical aging and gradual denaturing. Cytometry *43,* 101-109.

Heng, H.H., Goetze, S., Ye, C.J., Liu, G., Stevens, J.B., Bremer, S.W., Wykes, S.M., Bode, J., and Krawetz, S.A. (2004). Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. J. Cell. Sci. *117,* 999-1008.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nat. Rev. Genet. *9,* 15-26.

Henikoff, S., Furuyama, T., and Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. Trends Genet. *20,* 320-326.

Heslop, L., Beauchamp, J.R., Tajbakhsh, S., Buckingham, M.E., Partridge, T.A., and Zammit, P.S. (2001). Transplanted primary neonatal myoblasts can give rise to functional satellite cells as identified using the Myf5nlacZl+ mouse. Gene Ther. *8,* 778-783.

Heslop, L., Morgan, J.E., and Partridge, T.A. (2000). Evidence for a myogenic stem cell that is exhausted in dystrophic muscle. J. Cell. Sci. *113 (Pt 12),* 2299-2308.

Hnia, K., Zouiten, D., Cantel, S., Chazalette, D., Hugon, G., Fehrentz, J.A., Masmoudi, A., Diment, A., Bramham, J., Mornet, D., and Winder, S.J. (2007). ZZ domain of dystrophin and utrophin: topology and mapping of a beta-dystroglycan interaction site. Biochem. J. *401,* 667-677.

Hodges, B.L., Taylor, K.M., Joseph, M.F., Bourgeois, S.A., and Scheule, R.K. (2004). Long-term transgene expression from plasmid DNA gene therapy vectors is negatively affected by CpG dinucleotides. Mol. Ther. *10,* 269-278.

Hoffman, E.P., and Dressman, D. (2001). Molecular pathophysiology and targeted therapeutics for muscular dystrophy. Trends Pharmacol. Sci. *22,* 465-470.

Hohenester, E., Tisi, D., Talts, J.F., and Timpl, R. (1999). The crystal structure of a laminin G-like module reveals the molecular basis of alpha-dystroglycan binding to laminins, perlecan, and agrin. Mol. Cell *4,* 783-792.

Holder, E., Maeda, M., and Bies, R.D. (1996). Expression and regulation of the dystrophin Purkinje promoter in human skeletal muscle, heart, and brain. Hum. Genet. *97,* 232-239.

Hozak, P., Hassan, A.B., Jackson, D.A., and Cook, P.R. (1993). Visualization of replication factories attached to nucleoskeleton. Cell *73,* 361-373.

Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., and Eck, M.J. (2000). Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. Nat. Struct. Biol. *7,* 634-638.

Huard, J., Cao, B., and Qu-Petersen, Z. (2003). Muscle-derived stem cells: potential for muscle regeneration. Birth Defects Res. C. Embryo. Today *69,* 230-237.

Hunninghake, G.W., Monick, M.M., Liu, B., and Stinski, M.F. (1989). The promoter-regulatory region of the major immediate-early gene of human cytomegalovirus responds to T-lymphocyte stimulation and contains functional cyclic AMP-response elements. J. Virol. *63,* 3026-3033.

Hyde, S.C., Pringle, I.A., Abdullah, S., Lawton, A.E., Davies, L.A., Varathalingam, A., Nunez-Alonso, G., Green, A.M., Bazzani, R.P., Sumner-Jones, S.G*., et al.* (2008). CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression. Nat. Biotechnol. *26,* 549-551.

Ibraghimov-Beskrovnaya, O., Milatovich, A., Ozcelik, T., Yang, B., Koepnick, K., Francke, U., and Campbell, K.P. (1993). Human dystroglycan: skeletal muscle cDNA, genomic structure, origin of tissue specific isoforms and chromosomal localization. Hum. Mol. Genet. *2,* 1651-1657.

Igoucheva, O., Alexeev, V., and Yoon, K. (2006). Differential cellular responses to exogenous DNA in mammalian cells and its effect on oligonucleotide-directed gene modification. Gene Ther. *13,* 266-275.

Igoucheva, O., Alexeev, V., and Yoon, K. (2001). Targeted gene correction by small single-stranded oligonucleotides in mammalian cells. Gene Ther. *8,* 391-399.

Ikemoto, M., Fukada, S., Uezumi, A., Masuda, S., Miyoshi, H., Yamamoto, H., Wada, M.R., Masubuchi, N., Miyagoe-Suzuki, Y., and Takeda, S. (2007). Autologous transplantation of SM/C-2.6(+) satellite cells transduced with micro-dystrophin CS1 cDNA by lentiviral vector into mdx mice. Mol. Ther. *15,* 2178-2185.

Inagaki, K., Lewis, S.M., Wu, X., Ma, C., Munroe, D.J., Fuess, S., Storm, T.A., Kay, M.A., and Nakai, H. (2007). DNA palindromes with a modest arm length of greater, similar 20 base pairs are a significant target for recombinant adeno-associated virus vector integration in the liver, muscles, and heart in mice. J. Virol. *81,* 11290-11303.

Irintchev, A., Zeschnigk, M., Starzinski-Powitz, A., and Wernig, A. (1994). Expression pattern of M-cadherin in normal, denervated, and regenerating mouse muscles. Dev. Dyn. *199,* 326-337.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409,* 533-538.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409,* 533-538.

James, M., Nuttall, A., Ilsley, J.L., Ottersbach, K., Tinsley, J.M., Sudol, M., and Winder, S.J. (2000). Adhesion-dependent tyrosine phosphorylation of (beta)-dystroglycan regulates its interaction with utrophin. J. Cell. Sci. *113 (Pt 10),* 1717-1726.

Jearawiriyapaisarn, N., Moulton, H.M., Buckley, B., Roberts, J., Sazani, P., Fucharoen, S., Iversen, P.L., and Kole, R. (2008). Sustained dystrophin expression induced by peptide-conjugated morpholino oligomers in the muscles of mdx mice. Mol. Ther. *16,* 1624-1629.

Jenke, A.C., Eisenberger, T., Baiker, A., Stehle, I.M., Wirth, S., and Lipps, H.J. (2005). The nonviral episomal replicating vector pEPI-1 allows long-term inhibition of bcr-abl expression by shRNA. Hum. Gene Ther. *16,* 533-539.

Jenke, A.C., Stehle, I.M., Herrmann, F., Eisenberger, T., Baiker, A., Bode, J., Fackelmayer, F.O., and Lipps, H.J. (2004). Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. Proc. Natl. Acad. Sci. U. S. A. *101,* 11322-11327.

Jenke, A.C., Wilhelm, A.D., Orth, V., Lipps, H.J., Protzer, U., and Wirth, S. (2008). Long-term suppression of hepatitis B virus replication by short hairpin RNA expression using the scaffold/matrix attachment region-based replicating vector system pEPI-1. Antimicrob. Agents Chemother. *52,* 2355-2359.

Jenke, B.H., Fetzer, 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. EMBO Rep. *3,* 349-354.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev. *21,* 1519-1529.

Kapsa, R., Quigley, A., Lynch, G.S., Steeper, K., Kornberg, A.J., Gregorevic, P., Austin, L., and Byrne, E. (2001). In vivo and in vitro correction of the mdx dystrophin gene nonsense mutation by short-fragment homologous replacement. Hum. Gene Ther. *12,* 629-642.

Kardon, G., Campbell, J.K., and Tabin, C.J. (2002). Local extrinsic signals determine muscle and endothelial cell fate and patterning in the vertebrate limb. Dev. Cell. *3,* 533-545.

Kaufman, S.J., and Robert-Nicoud, M. (1985). DNA replication and differentiation in rat myoblasts studied with monoclonal antibodies against 5-bromodeoxyuridine, actin, and alpha 2-macroglobulin. Cytometry *6,* 570-577.

Kay, V., and Bode, J. (1994). Binding specificity of a nuclear scaffold: supercoiled, single-stranded, and scaffold-attached-region DNA. Biochemistry *33,* 367-374.

Khurana, T.S., Watkins, S.C., Chafey, P., Chelly, J., Tome, F.M., Fardeau, M., Kaplan, J.C., and Kunkel, L.M. (1991). Immunolocalization and developmental expression of dystrophin related protein in skeletal muscle. Neuromuscul. Disord. *1,* 185-194.

Kimura, E., Han, J.J., Li, S., Fall, B., Ra, J., Haraguchi, M., Tapscott, S.J., and Chamberlain, J.S. (2008). Cell-lineage regulated myogenesis for dystrophin replacement: a novel therapeutic approach for treatment of muscular dystrophy. Hum. Mol. Genet. *17,* 2507-2517.

Kimura, H., and Cook, P.R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J. Cell Biol. *153,* 1341-1353.

Kinali, M., Arechavala-Gomeza, V., Feng, L., Cirak, S., Hunt, D., Adkin, C., Guglieri, M., Ashton, E., Abbs, S., Nihoyannopoulos, P.*, et al.* (2009). Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. Lancet Neurol. *8,* 918-928.

Kipp, M., Gohring, F., Ostendorp, T., van Drunen, C.M., van Driel, R., Przybylski, M., and Fackelmayer, F.O. (2000). SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. Mol. Cell. Biol. *20,* 7480-7489.

Kitzmann, M., Carnac, G., Vandromme, M., Primig, M., Lamb, N.J., and Fernandez, A. (1998). The muscle regulatory factors MyoD and myf-5 undergo distinct cell cycle-specific expression in muscle cells. J. Cell Biol. *142,* 1447-1459.

Kobayashi, Y.M., Rader, E.P., Crawford, R.W., Iyengar, N.K., Thedens, D.R., Faulkner, J.A., Parikh, S.V., Weiss, R.M., Chamberlain, J.S., Moore, S.A., and Campbell, K.P. (2008). Sarcolemma-localized nNOS is required to maintain activity after mild exercise. Nature *456,* 511-515.

Kochanek, S., Renz, D., and Doerfler, W. (1995). Transcriptional silencing of human Alu sequences and inhibition of protein binding in the box B regulatory elements by 5'-CG-3' methylation. FEBS Lett. *360,* 115-120.

Koedood, M., Fichtel, A., Meier, P., and Mitchell, P.J. (1995). Human cytomegalovirus (HCMV) immediate-early enhancer/promoter specificity during embryogenesis defines target tissues of congenital HCMV infection. J. Virol. *69,* 2194-2207.

Koenig, M., and Kunkel, L.M. (1990). Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. J. Biol. Chem. *265,* 4560-4566.

Koenig, M., Monaco, A.P., and Kunkel, L.M. (1988). The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell *53,* 219-228.

Kothari, S., Baillie, J., Sissons, J.G., and Sinclair, J.H. (1991). The 21bp repeat element of the human cytomegalovirus major immediate early enhancer is a negative regulator of gene expression in undifferentiated cells. Nucleic Acids Res. *19,* 1767-1771.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128,* 693-705.

Kramarcy, N.R., Vidal, A., Froehner, S.C., and Sealock, R. (1994). Association of utrophin and multiple dystrophin short forms with the mammalian M(r) 58,000 dystrophin-associated protein (syntrophin). J. Biol. Chem. *269,* 2870-2876.

Kramer, M.G., Barajas, M., Razquin, N., Berraondo, P., Rodrigo, M., Wu, C., Qian, C., Fortes, P., and Prieto, J. (2003). In vitro and in vivo comparative study of chimeric liver-specific promoters. Mol. Ther. *7,* 375-385.

Kricker, M.C., Drake, J.W., and Radman, M. (1992). Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. Proc. Natl. Acad. Sci. U. S. A. *89,* 1075-1079.

Kuang, S., Charge, S.B., Seale, P., Huh, M., and Rudnicki, M.A. (2006). Distinct roles for Pax7 and Pax3 in adult regenerative myogenesis. J. Cell Biol. *172,* 103-113.

LaBarge, M.A., and Blau, H.M. (2002). Biological progression from adult bone marrow to mononucleate muscle stem cell to multinucleate muscle fiber in response to injury. Cell *111,* 589-601.

Lai, Y., Yue, Y., Liu, M., Ghosh, A., Engelhardt, J.F., Chamberlain, J.S., and Duan, D. (2005). Efficient in vivo gene expression by trans-splicing adeno-associated viral vectors. Nat. Biotechnol. *23,* 1435-1439.

Laumonier, T., Yang, S., Konig, S., Chauveau, C., Anegon, I., Hoffmeyer, P., and Menetrey, J. (2008). Lentivirus mediated HO-1 gene transfer enhances myogenic precursor cell survival after autologous transplantation in pig. Mol. Ther. *16,* 404-410.

Lavitrano, M., Bacci, M.L., Forni, M., Lazzereschi, D., Di Stefano, C., Fioretti, D., Giancotti, P., Marfe, G., Pucci, L., Renzi, L*., et al.* (2002). Efficient production by sperm-mediated gene transfer of human decay accelerating factor (hDAF) transgenic pigs for xenotransplantation. Proc. Natl. Acad. Sci. U. S. A. *99,* 14230-14235.

Leach,D. and Lindsey,J. (1986) In vivo loss of supercoiled DNA carrying a palindromic sequence. Mol. Gen. Genet. *204*, 322–327.

Lee, J.Y., Qu-Petersen, Z., Cao, B., Kimura, S., Jankowski, R., Cummins, J., Usas, A., Gates, C., Robbins, P., Wernig, A., and Huard, J. (2000). Clonal isolation of muscle-derived cells capable of enhancing muscle regeneration and bone healing. J. Cell Biol. *150,* 1085-1100.

Lei, J.X., Liu, Q.Y., Sodja, C., LeBlanc, J., Ribecco-Lutkiewicz, M., Smith, B., Charlebois, C., Walker, P.R., and Sikorska, M. (2005). S/MAR-binding properties of Sox2 and its involvement in apoptosis of human NT2 neural precursors. Cell Death Differ. *12,* 1368-1377.

Li, F., Chen, J., Solessio, E., and Gilbert, D.M. (2003). Spatial distribution and specification of mammalian replication origins during G1 phase. J. Cell Biol. *161,* 257-266.

Li, S., Kimura, E., Fall, B.M., Reyes, M., Angello, J.C., Welikson, R., Hauschka, S.D., and Chamberlain, J.S. (2005). Stable transduction of myogenic cells with lentiviral vectors expressing a minidystrophin. Gene Ther. *12,* 1099-1108.

Li, S., MacLaughlin, F.C., Fewell, J.G., Li, Y., Mehta, V., French, M.F., Nordstrom, J.L., Coleman, M., Belagali, N.S., Schwartz, R.J., and Smith, L.C. (1999). Increased level and duration of expression in muscle by co-expression of a transactivator using plasmid systems. Gene Ther. *6,* 2005-2011.

Liebich, I., Bode, J., Reuter, I., and Wingender, E. (2002). Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). Nucleic Acids Res. *30,* 3433-3442.

Lim, L.E., and Campbell, K.P. (1998). The sarcoglycan complex in limb-girdle muscular dystrophy. Curr. Opin. Neurol. *11,* 443-452.

Lin, C.M., Fu, H., Martinovsky, M., Bouhassira, E., and Aladjem, M.I. (2003). Dynamic alterations of replication timing in mammalian cells. Curr. Biol. *13,* 1019-1028.

Liu, S., Calderwood, D.A., and Ginsberg, M.H. (2000). Integrin cytoplasmic domain-binding proteins. J. Cell. Sci. *113 (Pt 20),* 3563-3571.

Loser, P., Jennings, G.S., Strauss, M., and Sandig, V. (1998). Reactivation of the previously silenced cytomegalovirus major immediate-early promoter in the mouse liver: involvement of NFkappaB. J. Virol. *72,* 180-190.

Loyola, A., and Almouzni, G. (2007). Marking histone H3 variants: how, when and why? Trends Biochem. Sci. *32,* 425-433.

Loyola, A., Bonaldi, T., Roche, D., Imhof, A., and Almouzni, G. (2006). PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol. Cell *24,* 309-316.

Lu, Q.L., Rabinowitz, A., Chen, Y.C., Yokota, T., Yin, H., Alter, J., Jadoon, A., Bou-Gharios, G., and Partridge, T. (2005). Systemic delivery of antisense oligoribonucleotide restores dystrophin expression in body-wide skeletal muscles. Proc. Natl. Acad. Sci. U. S. A. *102,* 198-203.

Luderus, M.E., den Blaauwen, J.L., de Smit, O.J., Compton, D.A., and van Driel, R. (1994). Binding of matrix attachment regions to lamin polymers involves single-stranded regions and the minor groove. Mol. Cell. Biol. *14,* 6297-6305.

Lufino, M.M., Edser, P.A., and Wade-Martins, R. (2008). Advances in high-capacity extrachromosomal vector technology: episomal maintenance, vector delivery, and transgene expression. Mol. Ther. *16,* 1525-1538.

Lumeng, C.N., Phelps, S.F., Rafael, J.A., Cox, G.A., Hutchinson, T.L., Begy, C.R., Adkins, E., Wiltshire, R., and Chamberlain, J.S. (1999). Characterization of dystrophin and utrophin diversity in the mouse. Hum. Mol. Genet. *8,* 593-599.

Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. Nature *382,* 646-649.

Macville, M., Schrock, E., Padilla-Nash, H., Keck, C., Ghadimi, B.M., Zimonjic, D., Popescu, N., and Ried, T. (1999). Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. Cancer Res. *59,* 141-150.

Mah, C., Cresawn, K.O., Fraites, T.J.,Jr, Pacak, C.A., Lewis, M.A., Zolotukhin, I., and Byrne, B.J. (2005). Sustained correction of glycogen storage disease type II using adeno-associated virus serotype 1 vectors. Gene Ther. *12,* 1405-1409.

Mal, A., and Harter, M.L. (2003). MyoD is functionally linked to the silencing of a muscle-specific regulatory gene prior to skeletal myogenesis. Proc. Natl. Acad. Sci. U. S. A. *100,* 1735-1739.

Malagon,F. and Aguilera,A. (1998) Genetic stability and DNA rearrangements associated with a 2 x 1.1-Kb perfect palindrome in Escherichia coli. Mol. Gen. Genet. *259*, 639–644

Manzini, S., Vargiolu, A., Stehle, I.M., Bacci, M.L., Cerrito, M.G., Giovannoni, R., Zannoni, A., Bianco, M.R., Forni, M., Donini, P*., et al.* (2006). Genetically modified pigs produced with a nonviral episomal vector. Proc. Natl. Acad. Sci. U. S. A. *103,* 17672-17677.

Martens, J.H., Verlaan, M., Kalkhoven, E., Dorsman, J.C., and Zantema, A. (2002). Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. Mol. Cell. Biol. *22,* 2598-2606.

Matsumura, K., Ervasti, J.M., Ohlendieck, K., Kahl, S.D., and Campbell, K.P. (1992). Association of dystrophin-related protein with dystrophin-associated proteins in mdx mouse muscle. Nature *360,* 588-591.

Maucksch, C., Bohla, A., Hoffmann, F., Schleef, M., Aneja, M.K., Elfinger, M., Hartl, D., and Rudolph, C. (2009). Transgene expression of transfected supercoiled plasmid DNA concatemers in mammalian cells. J. Gene Med. *11,* 444-453.

McClorey, G., Fall, A.M., Moulton, H.M., Iversen, P.L., Rasko, J.E., Ryan, M., Fletcher, S., and Wilton, S.D. (2006). Induced dystrophin exon skipping in human muscle explants. Neuromuscul. Disord. *16,* 583-590.

McDearmon, E.L., Burwell, A.L., Combs, A.C., Renley, B.A., Sdano, M.T., and Ervasti, J.M. (1998). Differential heparin sensitivity of alpha-dystroglycan binding to laminins expressed in normal and dy/dy mouse skeletal muscle. J. Biol. Chem. *273,* 24139-24144.

McGarvey, K.M., Greene, E., Fahrner, J.A., Jenuwein, T., and Baylin, S.B. (2007). DNA methylation and complete transcriptional silencing of cancer genes persist after depletion of EZH2. Cancer Res. *67,* 5097-5102.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc. Natl. Acad. Sci. U. S. A. *101,* 1525-1530.

Mechali, M. (2001). DNA replication origins: from sequence specificity to epigenetics. Nat. Rev. Genet. *2,* 640-645.

Mehler, M.F. (2000). Brain dystrophin, neurogenetics and mental retardation. Brain Res. Brain Res. Rev. *32,* 277-307.

Mesner, L.D., Hamlin, J.L., and Dijkwel, P.A. (2003). The matrix attachment region in the Chinese hamster dihydrofolate reductase origin of replication may be required for local chromatid separation. Proc. Natl. Acad. Sci. U. S. A. *100,* 3281-3286.

Miranda, T.B., and Jones, P.A. (2007). DNA methylation: the nuts and bolts of repression. J. Cell. Physiol. *213,* 384-390.

Mizuno, Y., Thompson, T.G., Guyon, J.R., Lidov, H.G., Brosius, M., Imamura, M., Ozawa, E., Watkins, S.C., and Kunkel, L.M. (2001). Desmuslin, an intermediate filament protein that interacts with alpha -dystrobrevin and desmin. Proc. Natl. Acad. Sci. U. S. A. *98,* 6156-6161.

Mlynarova, L., Jansen, R.C., Conner, A.J., Stiekema, W.J., and Nap, J.P. (1995). The MAR-Mediated Reduction in Position Effect Can Be Uncoupled from Copy Number-Dependent Expression in Transgenic Plants. Plant Cell *7,* 599-609.

Mocarski, E.S.,Jr, Abenes, G.B., Manning, W.C., Sambucetti, L.C., and Cherrington, J.M. (1990). Molecular genetic analysis of cytomegalovirus gene regulation in growth, persistence and latency. Curr. Top. Microbiol. Immunol. *154,* 47-74.

Molnar, M.J., Gilbert, R., Lu, Y., Liu, A.B., Guo, A., Larochelle, N., Orlopp, K., Lochmuller, H., Petrof, B.J., Nalbantoglu, J., and Karpati, G. (2004). Factors influencing the efficacy, longevity, and safety of electroporation-assisted plasmid-based gene transfer into mouse muscles. Mol. Ther. *10,* 447-455.

Moores, C.A., and Kendrick-Jones, J. (2000). Biochemical characterisation of the actin-binding properties of utrophin. Cell Motil. Cytoskeleton *46,* 116-128.

Mori, Y., Matsubara, H., Folco, E., Siegel, A., and Koren, G. (1993). The transcription of a mammalian voltage-gated potassium channel is regulated by cAMP in a cell-specific manner. J. Biol. Chem. *268,* 26482-26493.

Morris, G.E., Nguyen, T.M., Nguyen, T.N., Pereboev, A., Kendrick-Jones, J., and Winder, S.J. (1999). Disruption of the utrophin-actin interaction by monoclonal antibodies and prediction of an actin-binding surface of utrophin. Biochem. J. *337 (Pt 1),* 119-123.

Moulton, H.M., Fletcher, S., Neuman, B.W., McClorey, G., Stein, D.A., Abes, S., Wilton, S.D., Buchmeier, M.J., Lebleu, B., and Iversen, P.L. (2007). Cell-penetrating peptide-morpholino conjugates alter pre-mRNA splicing of DMD (Duchenne muscular dystrophy) and inhibit murine coronavirus replication in vivo. Biochem. Soc. Trans. *35,* 826-828.

Mu, Z.M., Le, X.F., Vallian, S., Glassman, A.B., and Chang, K.S. (1997). Stable overexpression of PML alters regulation of cell cycle progression in HeLa cells. Carcinogenesis *18,* 2063-2069.

Muir, L.A., and Chamberlain, J.S. (2009). Emerging strategies for cell and gene therapy of the muscular dystrophies. Expert Rev. Mol. Med. *11,* e18.

Murakami, T., Nishi, T., Kimura, E., Goto, T., Maeda, Y., Ushio, Y., Uchino, M., and Sunada, Y. (2003). Full-length dystrophin cDNA transfer into skeletal muscle of adult mdx mice by electroporation. Muscle Nerve *27,* 237-241.

Murray, A.W., and Szostak, J.W. (1983). Construction of artificial chromosomes in yeast. Nature *305,* 189-193.

Nadal-Ginard, B. (1978). Commitment, fusion and biochemical differentiation of a myogenic cell line in the absence of DNA synthesis. Cell *15,* 855-864.

Nagata, Y., Kobayashi, H., Umeda, M., Ohta, N., Kawashima, S., Zammit, P.S., and Matsuda, R. (2006). Sphingomyelin levels in the plasma membrane correlate with the activation state of muscle satellite cells. J. Histochem. Cytochem. *54,* 375-384.

Nakai, H., Storm, T.A., and Kay, M.A. (2000). Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors. Nat. Biotechnol. *18,* 527-532.

Nayler, O., Stratling, W., Bourquin, J.P., Stagljar, I., Lindemann, L., Jasper, H., Hartmann, A.M., Fackelmayer, F.O., Ullrich, A., and Stamm, S. (1998). SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. Nucleic Acids Res. *26,* 3542-3549.

Newey, S.E., Benson, M.A., Ponting, C.P., Davies, K.E., and Blake, D.J. (2000). Alternative splicing of dystrobrevin regulates the stoichiometry of syntrophin binding to the dystrophin protein complex. Curr. Biol. *10,* 1295-1298.

Newey, S.E., Howman, E.V., Ponting, C.P., Benson, M.A., Nawrotzki, R., Loh, N.Y., Davies, K.E., and Blake, D.J. (2001). Syncoilin, a novel member of the intermediate filament superfamily that interacts with alpha-dystrobrevin in skeletal muscle. J. Biol. Chem. *276,* 6645-6655.

Nguyen, A.T., Dow, A.C., Kupiec-Weglinski, J., Busuttil, R.W., and Lipshutz, G.S. (2008). Evaluation of gene promoters for liver expression by hydrodynamic gene transfer. J. Surg. Res. *148,* 60-66.

Nguyen, T.M., Ellis, J.M., Love, D.R., Davies, K.E., Gatter, K.C., Dickson, G., and Morris, G.E. (1991). Localization of the DMDL gene-encoded dystrophin-related protein using a panel of nineteen monoclonal antibodies: presence at neuromuscular junctions, in the sarcolemma of dystrophic skeletal muscle, in vascular and other smooth muscles, and in proliferating brain cell lines. J. Cell Biol. *115,* 1695-1700.

Niller, H.H., and Hennighausen, L. (1991). Formation of several specific nucleoprotein complexes on the human cytomegalovirus immediate early enhancer. Nucleic Acids Res. *19,* 3715-3721.

Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene *108,* 193-199.

Nowak, K.J., and Davies, K.E. (2004). Duchenne muscular dystrophy and dystrophin: pathogenesis and opportunities for treatment. EMBO Rep. *5,* 872-876.

Odom, G.L., Gregorevic, P., Allen, J.M., Finn, E., and Chamberlain, J.S. (2008). Microutrophin delivery through rAAV6 increases lifespan and improves muscle function in dystrophic dystrophin/utrophin-deficient mice. Mol. Ther. *16,* 1539-1545.

Ogino, H., Fujii, M., Satou, W., Suzuki, T., Michishita, E., and Ayusawa, D. (2002). Binding of 5-bromouracil-containing S/MAR DNA to the nuclear matrix. DNA Res. *9,* 25-29.

Ohlendieck, K., Ervasti, J.M., Matsumura, K., Kahl, S.D., Leveille, C.J., and Campbell, K.P. (1991). Dystrophin-related protein is localized to neuromuscular junctions of adult skeletal muscle. Neuron *7,* 499-508.

Orlando, V., Strutt, H., and Paro, R. (1997). Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods *11,* 205-214.

O'Rourke, J.P., Newbound, G.C., Kohn, D.B., Olsen, J.C., and Bunnell, B.A. (2002). Comparison of gene transfer efficiencies and gene expression levels achieved with equine infectious anemia virus- and human immunodeficiency virus type 1-derived lentivirus vectors. J. Virol. *76,* 1510-1515.

Pacak, C.A., Conlon, T., Mah, C.S., and Byrne, B.J. (2008). Relative persistence of AAV serotype 1 vector genomes in dystrophic muscle. Genet. Vaccines Ther. *6,* 14.

Painter, R.B., and Schaefer, A. (1969). State of newly synthesized HeLa DNA. Nature *221,* 1215-1217.

Pall, E.A., Bolton, K.M., and Ervasti, J.M. (1996). Differential heparin inhibition of skeletal muscle alpha-dystroglycan binding to laminins. J. Biol. Chem. *271,* 3817-3821.

Papapetrou, E.P., Ziros, P.G., Micheva, I.D., Zoumbos, N.C., and Athanassiadou, A. (2006). Gene transfer into human hematopoietic progenitor cells with an episomal vector carrying an S/MAR element. Gene Ther. *13,* 40-51.

Pardee, A.B. (1974). A restriction point for control of normal animal cell proliferation. Proc. Natl. Acad. Sci. U. S. A. *71,* 1286-1290.

Pardoll, D.M., Vogelstein, B., and Coffey, D.S. (1980). A fixed site of DNA replication in eucaryotic cells. Cell *19,* 527-536.

Park, F., and Kay, M.A. (2001). Modified HIV-1 based lentiviral vectors have an effect on viral transduction efficiency and gene expression in vitro and in vivo. Mol. Ther. *4,* 164-173.

Pearce, M., Blake, D.J., Tinsley, J.M., Byth, B.C., Campbell, L., Monaco, A.P., and Davies, K.E. (1993). The utrophin and dystrophin genes share similarities in genomic structure. Hum. Mol. Genet. *2,* 1765-1772.

Peng, B., Zhao, Y., Lu, H., Pang, W., and Xu, Y. (2005). In vivo plasmid DNA electroporation resulted in transfection of satellite cells and lasting transgene expression in regenerated muscle fibers. Biochem. Biophys. Res. Commun. *338,* 1490-1498.

Peters, M.F., Adams, M.E., and Froehner, S.C. (1997). Differential association of syntrophin pairs with the dystrophin complex. J. Cell Biol. *138,* 81-93.

Peters, M.F., Sadoulet-Puccio, H.M., Grady, M.R., Kramarcy, N.R., Kunkel, L.M., Sanes, J.R., Sealock, R., and Froehner, S.C. (1998). Differential membrane localization and intermolecular associations of alpha-dystrobrevin isoforms in skeletal muscle. J. Cell Biol. *142,* 1269-1278.

Petrov, A., Pirozhkova, I., Carnac, G., Laoudj, D., Lipinski, M., and Vassetzky, Y.S. (2006). Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. Proc. Natl. Acad. Sci. U. S. A. *103,* 6982-6987.

Piechaczek, C., Fetzer, C., Baiker, A., Bode, J., and Lipps, H.J. (1999). A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. Nucleic Acids Res. *27,* 426-428.

Piluso, G., Mirabella, M., Ricci, E., Belsito, A., Abbondanza, C., Servidei, S., Puca, A.A., Tonali, P., Puca, G.A., and Nigro, V. (2000). Gamma1- and gamma2-syntrophins, two novel dystrophin-binding proteins localized in neuronal cells. J. Biol. Chem. *275,* 15851-15860.

Podsakoff, G., Wong, K.K.,Jr, and Chatterjee, S. (1994). Efficient gene transfer into nondividing cells by adeno-associated virus-based vectors. J. Virol. *68,* 5656-5666.

Pogliaghi, G., Tacchini, L., Anzon, E., Radice, L., and Bernelli-Zazzera, A. (1995). Heat shock activation of NFkB in rat liver is mediated by interleukin-1. FEBS Lett. *372,* 181-184.

Polesskaya, A., Seale, P., and Rudnicki, M.A. (2003). Wnt signaling induces the myogenic specification of resident CD45+ adult stem cells during muscle regeneration. Cell *113,* 841-852.

Popplewell, L.J., Trollet, C., Dickson, G., and Graham, I.R. (2009). Design of phosphorodiamidate morpholino oligomers (PMOs) for the induction of exon skipping of the human DMD gene. Mol. Ther. *17,* 554-561.

Price, G.B., Allarakhia, M., Cossons, N., Nielsen, T., Diaz-Perez, M., Friedlander, P., Tao, L., and Zannis-Hadjopoulos, M. (2003). Identification of a cis-element that determines autonomous DNA replication in eukaryotic cells. J. Biol. Chem. *278,* 19649-19659.

Pringle, I.A., McLachlan, G., Collie, D.D., Sumner-Jones, S.G., Lawton, A.E., Tennant, P., Baker, A., Gordon, C., Blundell, R., Varathalingam, A*., et al.* (2007). Electroporation enhances reporter gene expression following delivery of naked plasmid DNA to the lung. J. Gene Med. *9,* 369-380.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat. Rev. Mol. Cell Biol. *10,* 192-206.

Quenneville, S.P., Chapdelaine, P., Skuk, D., Paradis, M., Goulet, M., Rousseau, J., Xiao, X., Garcia, L., and Tremblay, J.P. (2007). Autologous transplantation of muscle precursor cells modified with a lentivirus for muscular dystrophy: human cells and primate models. Mol. Ther. *15,* 431-438.

Ravnan, J.B., Gilbert, D.M., Ten Hagen, K.G., and Cohen, S.N. (1992). Random-choice replication of extrachromosomal bovine papillomavirus (BPV) molecules in heterogeneous, clonally derived BPV-infected cell lines. J. Virol. *66,* 6946-6952.

Reese, B.E., Bachman, K.E., Baylin, S.B., and Rountree, M.R. (2003). The methyl-CpG binding protein MBD1 interacts with the p150 subunit of chromatin assembly factor 1. Mol. Cell. Biol. *23,* 3226-3236.

Renz, A., and Fackelmayer, F.O. (1996). Purification and molecular cloning of the scaffold attachment factor B (SAF-B), a novel human nuclear protein that specifically binds to S/MAR-DNA. Nucleic Acids Res. *24,* 843-849.

Riu, E., Chen, Z.Y., Xu, H., He, C.Y., and Kay, M.A. (2007). Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo. Mol. Ther. *15,* 1348-1355.

Roberds, S.L., Anderson, R.D., Ibraghimov-Beskrovnaya, O., and Campbell, K.P. (1993). Primary structure and muscle-specific expression of the 50-kDa dystrophin-associated glycoprotein (adhalin). J. Biol. Chem. *268,* 23739-23742.

Roberds, S.L., Ervasti, J.M., Anderson, R.D., Ohlendieck, K., Kahl, S.D., Zoloto, D., and Campbell, K.P. (1993). Disruption of the dystrophin-glycoprotein complex in the cardiomyopathic hamster. J. Biol. Chem. *268,* 11496-11499.

Robertson, K.D., and Wolffe, A.P. (2000). DNA methylation in health and disease. Nat. Rev. Genet. *1,* 11-19.

Rodino-Klapac, L.R., Chicoine, L.G., Kaspar, B.K., and Mendell, J.R. (2007). Gene therapy for duchenne muscular dystrophy: expectations and challenges. Arch. Neurol. *64,* 1236-1241.

Rountree, M.R., Bachman, K.E., and Baylin, S.B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat. Genet. *25,* 269-277.

Rudd, S., Frisch, M., Grote, K., Meyers, B.C., Mayer, K., and Werner, T. (2004). Genome-wide in silico mapping of scaffold/matrix attachment regions in Arabidopsis suggests correlation of intragenic scaffold/matrix attachment regions with gene expression. Plant Physiol. *135,* 715-722.

Rusconi, S., Severne, Y., Georgiev, O., Galli, I., and Wieland, S. (1990). A novel expression assay to study transcriptional activators. Gene *89,* 211-221.

Sadoulet-Puccio, H.M., Khurana, T.S., Cohen, J.B., and Kunkel, L.M. (1996). Cloning and characterization of the human homologue of a dystrophin related phosphoprotein found at the Torpedo electric organ post-synaptic membrane. Hum. Mol. Genet. *5,* 489-496.

Sadoulet-Puccio, H.M., Rajala, M., and Kunkel, L.M. (1997). Dystrobrevin and dystrophin: an interaction through coiled-coil motifs. Proc. Natl. Acad. Sci. U. S. A. *94,* 12413-12418.

Sakuta, M. (2009). One hundred books which built up neurology (31)--William Richard Gowers "Pseudo-hypertrophic muscular paralysis. A clinical lecture" (1879). Brain Nerve *61,* 880-881.

Sampaolesi, M., Blot, S., D'Antona, G., Granger, N., Tonlorenzi, R., Innocenzi, A., Mognol, P., Thibaud, J.L., Galvez, B.G., Barthelemy, I.*, et al.* (2006). Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. Nature *444,* 574-579.

Sansom, O.J., Maddison, K., and Clarke, A.R. (2007). Mechanisms of disease: methyl-binding domain proteins as potential therapeutic targets in cancer. Nat. Clin. Pract. Oncol. *4,* 305-315.

Sarraf, S.A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol. Cell *15,* 595-605.

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. EMBO J. *23,* 191-201.

Schaarschmidt, D., Ladenburger, E.M., Keller, C., and Knippers, R. (2002). Human Mcm proteins at a replication origin during the G1 to S phase transition. Nucleic Acids Res. *30,* 4176-4185.

SCHERER, W.F. (1954). Studies on the propagation in vitro of poliomyelitis viruses. VI. Effect on virus yield of cell population, virus inoculum and temperature of incubation. J. Immunol. *73,* 331-336.

Schnepp, B.C., Jensen, R.L., Clark, K.R., and Johnson, P.R. (2009). Infectious molecular clones of adeno-associated virus isolated directly from human tissues. J. Virol. *83,* 1456-1464.

Schor, A.M., Canfield, A.E., Sutton, A.B., Allen, T.D., Sloan, P., and Schor, S.L. (1992). The behaviour of pericytes in vitro: relevance to angiogenesis and differentiation. EXS *61,* 167-178.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110,* 521-529.

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.

Seale, P., Sabourin, L.A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., and Rudnicki, M.A. (2000). Pax7 is required for the specification of myogenic satellite cells. Cell *102,* 777-786.

Seidman RJ. Muscle biopsy and introduction to the pathology of skeletal muscle- chapter for eMedicine, Neurology, an online textbook, http://www.emedicine.com/, 2001, 2005, 2006

Shefer, G., Van de Mark, D.P., Richardson, J.B., and Yablonka-Reuveni, Z. (2006). Satellite-cell pool size does matter: defining the myogenic potency of aging skeletal muscle. Dev. Biol. *294,* 50-66.

Sherwood, R.I., Christensen, J.L., Conboy, I.M., Conboy, M.J., Rando, T.A., Weissman, I.L., and Wagers, A.J. (2004). Isolation of adult mouse myogenic progenitors: functional heterogeneity of cells within and engrafting skeletal muscle. Cell *119,* 543-554.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119,* 941-953.

Shibahara, K., and Stillman, B. (1999). Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell *96,* 575-585.

Shinin, V., Gayraud-Morel, B., Gomes, D., and Tajbakhsh, S. (2006). Asymmetric division and cosegregation of template DNA strands in adult muscle satellite cells. Nat. Cell Biol. *8,* 677-687.

Sinclair, J.H., Baillie, J., Bryant, L.A., Taylor-Wiedeman, J.A., and Sissons, J.G. (1992). Repression of human cytomegalovirus major immediate early gene expression in a monocytic cell line. J. Gen. Virol. *73 (Pt 2),* 433-435.

Skuk, D., and Tremblay, J.P. (2003). Myoblast transplantation: the current status of a potential therapeutic tool for myopathies. J. Muscle Res. Cell. Motil. *24,* 285-300.

Sotgia, F., Lee, J.K., Das, K., Bedford, M., Petrucci, T.C., Macioce, P., Sargiacomo, M., Bricarelli, F.D., Minetti, C., Sudol, M., and Lisanti, M.P. (2000). Caveolin-3 directly interacts with the C-terminal tail of beta -dystroglycan. Identification of a central WW-like domain within caveolin family members. J. Biol. Chem. *275,* 38048-38058.

Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H., and Cardoso, M.C. (2002). DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. Mol. Cell *10,* 1355-1365.

Stamminger, T., Fickenscher, H., and Fleckenstein, B. (1990). Cell type-specific induction of the major immediate early enhancer of human cytomegalovirus by cyclic AMP. J. Gen. Virol. *71 (Pt 1),* 105-113.

Stehle, I.M., Postberg, J., Rupprecht, S., Cremer, T., Jackson, D.A., and Lipps, H.J. (2007). Establishment and mitotic stability of an extra-chromosomal mammalian replicon. BMC Cell Biol. *8,* 33.

Stehle, I.M., Scinteie, M.F., Baiker, A., Jenke, A.C., and Lipps, H.J. (2003). Exploiting a minimal system to study the epigenetic control of DNA replication: the interplay between transcription and replication. Chromosome Res. *11,* 413-421.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol. Mol. Biol. Rev. *64,* 435-459.

Straub, V., Ettinger, A.J., Durbeej, M., Venzke, D.P., Cutshall, S., Sanes, J.R., and Campbell, K.P. (1999). Epsilon-Sarcoglycan Replaces Alpha-Sarcoglycan in Smooth Muscle to Form a Unique Dystrophin-Glycoprotein Complex. J. Biol. Chem. *274,* 27989-27996.

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. Genes Dev. *12,* 599-606.

Sumer, H., Saffery, R., Wong, N., Craig, J.M., and Choo, K.H. (2004). Effects of scaffold/matrix alteration on centromeric function and gene expression. J. Biol. Chem. *279,* 37631-37639.

Taddei, A., Roche, D., Sibarita, J.B., Turner, B.M., and Almouzni, G. (1999). Duplication and maintenance of heterochromatin domains. J. Cell Biol. *147,* 1153-1166.

Tajbakhsh, S., Rocancourt, D., Cossu, G., and Buckingham, M. (1997). Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. Cell *89,* 127-138.

Takeshita, F., Takase, K., Tozuka, M., Saha, S., Okuda, K., Ishii, N., and Sasaki, S. (2007). Muscle creatine kinase/SV40 hybrid promoter for muscle-targeted long-term transgene expression. Int. J. Mol. Med. *19,* 309-315.

Tamaki, T., Akatsuka, A., Ando, K., Nakamura, Y., Matsuzawa, H., Hotta, T., Roy, R.R., and Edgerton, V.R. (2002). Identification of myogenic-endothelial progenitor cells in the interstitial spaces of skeletal muscle. J. Cell Biol. *157,* 571-577.

Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H., and Lassar, A.B. (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science *242,* 405-411.

Tetko, I.V., Haberer, G., Rudd, S., Meyers, B., Mewes, H.W., and Mayer, K.F. (2006). Spatiotemporal expression control correlates with intragenic scaffold matrix attachment regions (S/MARs) in Arabidopsis thaliana. PLoS Comput. Biol. *2,* e21.

Thomas, G.D., Sander, M., Lau, K.S., Huang, P.L., Stull, J.T., and Victor, R.G. (1998). Impaired metabolic modulation of alpha-adrenergic vasoconstriction in dystrophin-deficient skeletal muscle. Proc. Natl. Acad. Sci. U. S. A. *95,* 15090-15095.

Thomas, L., Stamberg, J., Gojo, I., Ning, Y., and Rapoport, A.P. (2004). Double minute chromosomes in monoblastic (M5) and myeloblastic (M2) acute myeloid leukemia: two case reports and a review of literature. Am. J. Hematol. *77,* 55-61.

Thompson, T.G., Chan, Y.M., Hack, A.A., Brosius, M., Rajala, M., Lidov, H.G., McNally, E.M., Watkins, S., and Kunkel, L.M. (2000). Filamin 2 (FLN2): A muscle-specific sarcoglycan interacting protein. J. Cell Biol. *148,* 115-126.

Tinsley, J., Deconinck, N., Fisher, R., Kahn, D., Phelps, S., Gillis, J.M., and Davies, K. (1998). Expression of full-length utrophin prevents muscular dystrophy in mdx mice. Nat. Med. *4,* 1441-1444.

Tinsley, J.M., Blake, D.J., Roche, A., Fairbrother, U., Riss, J., Byth, B.C., Knight, A.E., Kendrick-Jones, J., Suthers, G.K., and Love, D.R. (1992). Primary structure of dystrophin-related protein. Nature *360,* 591-593.

Tripathy, S.K., Svensson, E.C., Black, H.B., Goldwasser, E., Margalith, M., Hobart, P.M., and Leiden, J.M. (1996). Long-term expression of erythropoietin in the systemic circulation of mice after intramuscular injection of a plasmid DNA vector. Proc. Natl. Acad. Sci. U. S. A. *93,* 10876-10880.

Trollet, C., Athanasopoulos, T., Popplewell, L., Malerba, A., and Dickson, G. (2009). Gene therapy for muscular dystrophy: current progress and future prospects. Expert Opin. Biol. Ther. *9,* 849-866.

Turner, B.M. (2002). Cellular memory and the histone code. Cell *111,* 285-291.

Tye, B.K. (1999). MCM proteins in DNA replication. Annu. Rev. Biochem. *68,* 649-686.

Vacek, M., Sazani, P., and Kole, R. (2003). Antisense-mediated redirection of mRNA splicing. Cell Mol. Life Sci. *60,* 825-833.

van Deutekom, J.C., Bremmer-Bout, M., Janson, A.A., Ginjaar, I.B., Baas, F., den Dunnen, J.T., and van Ommen, G.J. (2001). Antisense-induced exon skipping restores dystrophin expression in DMD patient derived muscle cells. Hum. Mol. Genet. *10,* 1547-1554.

van Deutekom, J.C., Janson, A.A., Ginjaar, I.B., Frankhuizen, W.S., Aartsma-Rus, A., Bremmer-Bout, M., den Dunnen, J.T., Koop, K., van der Kooi, A.J., Goemans, N.M.*, et al.* (2007). Local dystrophin restoration with antisense oligonucleotide PRO051. N. Engl. J. Med. *357,* 2677-2686.

Vassetzky, Y., Hair, A., and Mechali, M. (2000). Rearrangement of chromatin domains during development in Xenopus. Genes Dev. *14,* 1541-1552.

Vater, R., Young, C., Anderson, L.V.B., Lindsay, S., Blake, D.J., Davies, K.E., Zuellig, R., and Slater, C.R. (1998). Utrophin mRNA Expression in Muscle Is Not Restricted to the Neuromuscular Junction. Mol. Cell. Neurosci. *10,* 229-242.

Vermeulen, K., Van Bockstaele, D.R., and Berneman, Z.N. (2003). The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. Cell Prolif. *36,* 131-149.

Vilaysane, A., and Muruve, D.A. (2009). The innate immune response to DNA. Semin. Immunol. *21,* 208-214.

Vilquin, J.T., Kennel, P.F., Paturneau-Jouas, M., Chapdelaine, P., Boissel, N., Delaere, P., Tremblay, J.P., Scherman, D., Fiszman, M.Y., and Schwartz, K. (2001). Electrotransfer of naked DNA in the skeletal muscles of animal models of muscular dystrophies. Gene Ther. *8,* 1097-1107.

Vitiello, C., Faraso, S., Sorrentino, N.C., Di Salvo, G., Nusco, E., Nigro, G., Cutillo, L., Calabro, R., Auricchio, A., and Nigro, V. (2009). Disease rescue and increased lifespan in a model of cardiomyopathy and muscular dystrophy by combined AAV treatments. PLoS One *4,* e5051.

Volonte, D., Liu, Y., and Galbiati, F. (2005). The modulation of caveolin-1 expression controls satellite cell activation during muscle repair. FASEB J. *19,* 237-239.

Wagner, K.R. (2008). Approaching a new age in Duchenne muscular dystrophy treatment. Neurotherapeutics *5,* 583-591.

Wagner, K.R., Cohen, J.B., and Huganir, R.L. (1993). The 87K postsynaptic membrane protein from Torpedo is a protein-tyrosine kinase substrate homologous to dystrophin. Neuron *10,* 511-522.

Wakeford, S., Watt, D.J., and Partridge, T.A. (1991). X-irradiation improves mdx mouse muscle as a model of myofiber loss in DMD. Muscle Nerve *14,* 42-50.

Walsh, C.A., and Goffinet, A.M. (2000). Potential mechanisms of mutations that affect neuronal migration in man and mouse. Curr. Opin. Genet. Dev. *10,* 270-274.

Wan, K.M., Nickerson, J.A., Krockmalnic, G., and Penman, S. (1999). The nuclear matrix prepared by amine modification. Proc. Natl. Acad. Sci. U. S. A. *96,* 933-938.

Wang, Z., Chamberlain, J.S., Tapscott, S.J., and Storb, R. (2009). Gene therapy in large animal models of muscular dystrophy. ILAR J. *50,* 187-198.

Wang, Z., Zhu, T., Qiao, C., Zhou, L., Wang, B., Zhang, J., Chen, C., Li, J., and Xiao, X. (2005). Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. Nat. Biotechnol. *23,* 321-328.

Wanisch, K., and Yanez-Munoz, R.J. (2009). Integration-deficient lentiviral vectors: a slow coming of age. Mol. Ther. *17,* 1316-1332.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. *39,* 457-466.

Weintraub, H. (1993). The MyoD family and myogenesis: redundancy, networks, and thresholds. Cell *75,* 1241-1244.

Wells, D.J. (2008). Treatments for muscular dystrophy: increased treatment options for Duchenne and related muscular dystropies. Gene Ther. *15,* 1077-1078.

Whetstine, J.R., and Matherly, L.H. (2001). The basal promoters for the human reduced folate carrier gene are regulated by a GC-box and a cAMP-response element/AP-1-like element. Basis for tissue-specific gene expression. J. Biol. Chem. *276,* 6350-6358.

Williamson, R.A., Henry, M.D., Daniels, K.J., Hrstka, R.F., Lee, J.C., Sunada, Y., Ibraghimov-Beskrovnaya, O., and Campbell, K.P. (1997). Dystroglycan is essential for early embryonic development: disruption of Reichert's membrane in Dag1-null mice. Hum. Mol. Genet. *6,* 831-841.

Wilson, J., Putt, W., Jimenez, C., and Edwards, Y.H. (1999). Up71 and up140, two novel transcripts of utrophin that are homologues of short forms of dystrophin. Hum. Mol. Genet. *8,* 1271-1278.

Wilton, S.D., Fall, A.M., Harding, P.L., McClorey, G., Coleman, C., and Fletcher, S. (2007). Antisense oligonucleotide-induced exon skipping across the human dystrophin gene transcript. Mol. Ther. *15,* 1288-1296.

Wilton, S.D., and Fletcher, S. (2008). Exon skipping and Duchenne muscular dystrophy: hope, hype and how feasible? Neurol. India *56,* 254-262.

Winder, S.J., Gibson, T.J., and Kendrick-Jones, J. (1995). Dystrophin and utrophin: the missing links! FEBS Lett. *369,* 27-33.

Winder, S.J., Hemmings, L., Maciver, S.K., Bolton, S.J., Tinsley, J.M., Davies, K.E., Critchley, D.R., and Kendrick-Jones, J. (1995). Utrophin actin binding domain: analysis of actin binding and cellular targeting. J. Cell. Sci. *108 (Pt 1),* 63-71.

Wong, E., and Wei, C.L. (2009). ChIP'ing the mammalian genome: technical advances and insights into functional elements. Genome Med. *1,* 89.

Wong, S.P., Argyros, O., Coutelle, C., and Harbottle, R.P. (2011). Non-viral S/MAR vectors replicate episomally in vivo when provided with a selective advantage. Gene Ther. *18,* 82-87.

Wong, S.P., Argyros, O., Coutelle, C., and Harbottle, R.P. (2011). Non-viral S/MAR vectors replicate episomally in vivo when provided with a selective advantage. Gene Ther. *18,* 82-87.

Wu, B., Moulton, H.M., Iversen, P.L., Jiang, J., Li, J., Li, J., Spurney, C.F., Sali, A., Guerron, A.D., Nagaraju, K.*, et al.* (2008). Effective rescue of dystrophin improves cardiac function in dystrophin-deficient mice by a modified morpholino oligomer. Proc. Natl. Acad. Sci. U. S. A. *105,* 14814-14819.

Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300,* 1749-1751.

Xiao, X., Li, J., and Samulski, R.J. (1996). Efficient long-term gene transfer into muscle tissue of immunocompetent mice by adeno-associated virus vector. J. Virol. *70,* 8098-8108.

Yablonka-Reuveni, Z., Day, K., Vine, A., and Shefer, G. (2008). Defining the transcriptional signature of skeletal muscle stem cells. J. Anim. Sci. *86,* E207-16.

Yablonka-Reuveni, Z., and Rivera, A.J. (1994). Temporal expression of regulatory and structural muscle proteins during myogenesis of satellite cells on isolated adult rat fibers. Dev. Biol. *164,* 588-603.

Yablonka-Reuveni, Z., Rudnicki, M.A., Rivera, A.J., Primig, M., Anderson, J.E., and Natanson, P. (1999). The transition from proliferation to differentiation is delayed in satellite cells from mice lacking MyoD. Dev. Biol. *210,* 440-455.

Yaffe, D., and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature *270,* 725-727.

Yamaji, A., Sekizawa, Y., Emoto, K., Sakuraba, H., Inoue, K., Kobayashi, H., and Umeda, M. (1998). Lysenin, a novel sphingomyelin-specific binding protein. J. Biol. Chem. *273,* 5300-5306.

Yan, Z., Zhang, Y., Duan, D., and Engelhardt, J.F. (2000). Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. Proc. Natl. Acad. Sci. U. S. A. *97,* 6716-6721.

Yew, N.S., Zhao, H., Przybylska, M., Wu, I.H., Tousignant, J.D., Scheule, R.K., and Cheng, S.H. (2002). CpG-depleted plasmid DNA vectors with enhanced safety and long-term gene expression in vivo. Mol. Ther. *5,* 731-738.

Yokota, T., Lu, Q.L., Morgan, J.E., Davies, K.E., Fisher, R., Takeda, S., and Partridge, T.A. (2006). Expansion of revertant fibers in dystrophic mdx muscles reflects activity of muscle precursor cells and serves as an index of muscle regeneration. J. Cell. Sci. *119,* 2679-2687.

Yoshida, M., Hama, H., Ishikawa-Sakurai, M., Imamura, M., Mizuno, Y., Araishi, K., Wakabayashi-Takai, E., Noguchi, S., Sasaoka, T., and Ozawa, E. (2000). Biochemical evidence for association of dystrobrevin with the sarcoglycan-sarcospan complex as a basis for understanding sarcoglycanopathy. Hum. Mol. Genet. *9,* 1033-1040.

Yoshida, M., Suzuki, A., Yamamoto, H., Noguchi, S., Mizuno, Y., and Ozawa, E. (1994). Dissociation of the complex of dystrophin and its associated proteins into several unique groups by n-octyl beta-D-glucoside. Eur. J. Biochem. *222,* 1055-1061.

Yoshida, N., Yoshida, S., Koishi, K., Masuda, K., and Nabeshima, Y. (1998). Cell heterogeneity upon myogenic differentiation: down-regulation of MyoD and Myf-5 generates 'reserve cells'. J. Cell. Sci. *111 (Pt 6),* 769-779.

Yue, Y., Ghosh, A., Long, C., Bostick, B., Smith, B.F., Kornegay, J.N., and Duan, D. (2008). A single intravenous injection of adeno-associated virus serotype-9 leads to whole body skeletal muscle transduction in dogs. Mol. Ther. *16,* 1944-1952.

Zammit, P.S. (2008). All muscle satellite cells are equal, but are some more equal than others? J. Cell. Sci. *121,* 2975-2982.

Zammit, P.S., Golding, J.P., Nagata, Y., Hudon, V., Partridge, T.A., and Beauchamp, J.R. (2004). Muscle satellite cells adopt divergent fates: a mechanism for self-renewal? J. Cell Biol. *166,* 347-357.

Zammit, P.S., Heslop, L., Hudon, V., Rosenblatt, J.D., Tajbakhsh, S., Buckingham, M.E., Beauchamp, J.R., and Partridge, T.A. (2002). Kinetics of myoblast proliferation show that resident satellite cells are competent to fully regenerate skeletal muscle fibers. Exp. Cell Res. *281,* 39-49.

Zammit, P.S., Partridge, T.A., and Yablonka-Reuveni, Z. (2006). The skeletal muscle satellite cell: the stem cell that came in from the cold. J. Histochem. Cytochem. *54,* 1177-1191.

Zammit, P.S., Relaix, F., Nagata, Y., Ruiz, A.P., Collins, C.A., Partridge, T.A., and Beauchamp, J.R. (2006). Pax7 and myogenic progression in skeletal muscle satellite cells. J. Cell. Sci. *119,* 1824-1832.

Zernicka-Goetz, M., Pines, J., McLean Hunter, S., Dixon, J.P., Siemering, K.R., Haseloff, J., and Evans, M.J. (1997). Following cell fate in the living mouse embryo. Development *124,* 1133-1137.

Zhao, K., Kas, E., Gonzalez, E., and Laemmli, U.K. (1993). SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. EMBO J. *12,* 3237-3247.