Development of Viral & Non-viral Episomal Vectors For Gene Therapy Applications

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE UNIVERSITY OF LONDON

MARCH 2012

BY

HANNA KYMÄLÄINEN

SCHOOL OF BIOLOGICAL SCIENCES

ROYAL HOLLOWAY UNIVERSITY OF LONDON

EGHAM, SURREY

Statement

Unless otherwise stated in the text, all the work contained in this thesis was conducted by the author, working in the School of Biological Sciences, Royal Holloway University of London, between October 2008 and November 2011.

All the work is original unless acknowledged by references in the text. This work has not been submitted for any other degree at this or any other university.

Signed:

Date:

Abstract

Gene therapy consists of methods which attempt to repair or replace defective genes responsible for disease, or to add genes to a therapeutic effect. To achieve this, two episomally maintained recombinant viral vectors have shown promising results: integration-deficient lentiviral vectors (IDLVs), and adeno-associated virus (AAV) vectors. The non-integrating nature of these vectors improves their safety profile but also limits transgene retention as nuclear episomes generally get lost during cell division.

In the present study, the establishment of stable replicating episomes via transduction with AAV and IDLV gene therapy vectors was examined in CHO cells. Different DNA elements and cell culture conditions were evaluated, and in particular the effects of (i) DNA elements called S/MARs (scaffold/matrix attachment regions) which are involved in chromatin organisation, transcription and replication, and (ii) induction of transient cell cycle arrest in transfected and transduced cell populations. In the case of both AAV and IDLV vectors, the incorporation of S/MAR elements into vector transcription units had only marginal effects on the establishment of stable transgene-positive cell populations, either with or without induction of transient cell cycle arrest.

However, a striking general result was observed in cell populations transduced with IDLVs and subjected to a transient cell cycle arrest soon after transduction. Under these conditions, following release from cell cycle arrest and in the absence of any selection pressure, substantial populations (10-25%) of proliferating and stably transduced cells emerged and were maintained over at least 100 population doublings. This establishment of stable transduction was seen only with IDLVs, was crucially dependent on the induction of a period of transient cell cycle arrest, occurred independently of the presence of S/MAR elements, and resulted in transgene-positive cell populations which could be isolated and propagated as stable clonal cell lines. In these polyclonal and clonal IDLV-transduced cell lines, the existence of non-integrated vector genomes in the form of multi-copy nuclear episomes was confirmed by evidence from linear amplification –mediated PCR, deep sequencing, Southern blotting and FISH (fluorescent *in situ* hybridisation).

The cumulative evidence suggests that transduction of CHO cells with IDLVs followed by a short period of induced cell cycle arrest leads to the establishment of stable IDLV-based nuclear episomes which are transcriptionally active and undergo replication and segregation during cell division without the need for antibiotic-based or other positive selection pressure.

Preliminary investigations were also done to test the capacity of combined IDLV transduction and transient cell cycle arrest to establish stable episome HeLa cells and murine haematopoietic stem cells. However, further experiments are required either to optimise the protocol in these cells or to find other clinically relevant cell types in which the protocol can be implemented. The transfer of this technology to a variety of clinically relevant human stem or progenitor cell populations could improve the safety profile of a range of gene therapy strategies currently under investigation.

Acknowledgements

Firstly, I would like to thank my supervisor George Dickson for giving me the opportunity to undertake my PhD. His continuing support, encouragement and sage advice have made my introduction to a career in research science undoubtedly more successful and enjoyable. I would also like to thank my second supervisor, Rafael Yáñez-Muñoz, for his valuable academic input and advice.

I would like to thank all the members of the Dickson lab for making it such a pleasant working environment. I would especially like to thank Takis, Susie, Anita, Linda and Martin for their friendship and personal support, for taking the time to plough through my thesis, and for generally making the last three years a lot more fun. I am also thankful for the continuing scientific assistance, kindness, and coffee company of Keith and Helen. A special thanks goes to Alberto and Martin for having made the train journeys fly by.

I would also like to acknowledge the practical assistance of Mariama Kabba and the help, advice and hospitality of Caroline Ogilvie and Angela Davies at the Cytogenetics Department at King's College Hospital during my months of FISHing.

The company, advice and practical assistance of Christina Lehrer, Frank Giordano and Uwe Appelt were invaluable for the success of the qRT-PCR and LAM-PCR at the NCT in Heidelberg. I would also like to thank Manfred Schmidt and Christof von Kalle for their hospitality.

A massive thank you goes to my parents, Eino and Helinä Kymäläinen, without whom none of this would have been possible. You have been the bedrock of all my pursuits, and it is with your love and support that I have got this far.

I would also like to thank Henry for his love and support during the stressful times working on this thesis, for the flowers, hugs, and for generally making it all better.

Abbreviations

AAV	Adeno-associated virus
BLAST	Basic local alignment search tool
bp	Base pair(s)
CAG	CMV enhancer / chicken β-actin promoter
CHO	Chinese hamster ovary (cell line)
CMV	Cytomegalovirus
CNS	Central nervous system
Cre	Recombinase enzyme (**c**auses **re**combination)
DMEM	Dulbecco's modified Eagle's medium
DMEM/-Met	Dulbecco's Modified Eagle's Medium without methionine and 2% FCS
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
FACS	Fluorescence-activated cell sorting
FCS	Foetal calf serum
FISH	Fluorescent *in situ* hybridisation
FLP	Flippase recombination enzyme
Flt-3	Fms-related tyrosine kinase 3 ligand
FRT	FLP recognition target
GFP	Green fluorescent protein
HeLa	Cervical cancer cell line derived from Henrietta Lacks
HIV	Human immunodeficiency virus
HSC	Haematopoietic stem cell

ICLV	Integration-competent lentiviral vector
IDLV	Integration-deficient lentiviral vector
IL	Interleukin
ITR	Inverted terminal repeat
kb	Kilobase(s)
LAM-PCR	Linear amplification-mediated polymerase chain reaction
LTR	Long terminal repeat
MFI	Mean fluorescence intensity
mHSC	murine haematopoietic stem cell
MLV	Murine leukaemia virus
MOI	Multiplicity of infection
MSC	Mesenchymal stem cell
NHEJ	Non-homologous end joining
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pEPI	Episomal plasmid vector
PPT	Polypurine tract
qRT-PCR	Quantitative real-time polymerase chain reaction
rAAV	Recombinant AAV
RBS	Rep-binding site
RE	Restriction enzyme
RNA	Ribonucleic acid
RT	Room temperature
S/MAR	Scaffold/Matrix attachment region

scAAV	Self-complementary AAV
SCF	Stem cell factor
SCID	Severe combined immunodeficiency
SIDD	Stress-induced strand separation
ssAAV	Single-stranded AAV
SSC	Saline sodium citrate buffer
SV40	Simian virus 40
TRS	Terminal resolution site
VSV-g	Vesicular stomatitis virus glycoprotein G
WPRE	Woodchuck hepatitis virus post-transcriptional element
wt	Wild-type
293A	Human embryonic kidney cell line, adherent
293T	Human embryonic kidney cell line, containing SV40 large T-antigen

Table of Contents

Chapter 1: Introduction

1.1 Introduction of vector systems for episomal gene addition therapy

The aim of gene therapy is to correct a disease phenotype by introducing a therapeutic gene into the appropriate cells. This approach is currently being investigated for the treatment of several types of disorders, including cancer, acquired diseases such as Parkinson's and Alzheimer's disease, and genetically inherited diseases. Whereas the gene therapy methods for treatment of acquired and inherited diseases are very similar and consist of the introduction of a therapeutic gene into the patient, cancer gene therapy may differ somewhat in that it often aims to promote an immune reaction against the tumour cells. This thesis concentrates on the investigation of treatments for genetically inherited diseases.

The genetic diseases under investigation for gene addition therapy are generally caused by one dysfunctional gene, in which the therapy aims to provide a functional copy. The diseases are usually rare, chronic and incurable, and have limited treatment options through traditional medicine. Some currently under investigation include the haemophilias, Duchenne muscular dystrophy, and cystic fibrosis amongst others.

There are multiple attractions and advantages to treating a genetically inherited disease by introducing a functional copy of the affected gene into the patient. Firstly, many of the diseases that gene therapy seeks to cure are otherwise untreatable, so that any method of providing relief of the symptoms would be beneficial. In the case of a disease for which some treatment exists, gene therapy is likely to be less disruptive to the patient compared to current medical regimens. Secondly, gene therapy seeks to cure the patient with one or, at the most, very few interventions, making the prospect cost-effective and applicable also to the developing world. Lastly, the patient is less likely to suffer from complications relating to the rejection of the gene product or treated cells by the immune system, as it is synthesised endogenously without the need for allotransplantation.

The key goals of gene therapy are long-term expression of a transgene at levels high enough to ameliorate or correct the disease phenotype. Currently, there are two ways of achieving efficient long-term expression; either an integrating delivery vector can be used to transduce a stem or progenitor cell population so that the resulting daughter cell populations will all carry the introduced gene, or non-integrating vectors can be used to transduce post-mitotic or slowly dividing tissues such as the CNS, muscle and liver. Generally, lentiviral vectors are employed for the integrating strategies and they are often performed *ex vivo*. The imported gene can then either integrate into the host genome, randomly or non-randomly by site-specific recombination, or remain episomal. Non-integrating vectors including the adeno-associated virus as well as non-viral methods are favoured for non-integrating gene transfer strategies that can be executed both *in vivo* and *ex vivo*.

Each of these methods has its advantages and challenges: integration into the host genome provides stable gene transfer but random integration may cause insertional mutagenesis, as transgenes delivered by retroviral vectors have been found to preferentially integrate close to genes in transcriptionally active sites (Hacein-Bey-Abina *et al.*, 2008). Homologous recombination, where the integration site of an incoming gene is directed by including homologous sequences, is safer but as of yet fairly inefficient (Wanisch *et al.*, 2009). Direct replacing of a dysfunctional gene segment with a functional copy is an attractive goal but the techniques need to improve before it can be considered an efficacious form of therapy, and thus far, for many diseases, gene addition appears more advantageous when compared to gene replacement strategies.

The investigations in this thesis are centred on providing a non-integrating transgene which will remain as a stable and replicating extrachromosomal addition to the recipient's genome. To this end, adeno-associated virus and integration-deficient lentivirus are assessed as a basis for episomal viral vector systems, and plasmids and minicircles as a basis for non-viral systems. These vectors were chosen due to their capacity for providing long-term transgene expression without integration. Other popular vectors such as adenoviral vectors are better suited to other applications such as cancer gene therapy.

There are some areas which need further development until gene therapy can become a mainstream treatment option. Potential problems with gene therapy include initiating an adverse immune response and associated acute toxicity, which mostly concern vectors to which the patient may have pre-existing antibodies. The first reported death due to a gene therapy vector was caused by an inflammatory reaction to an adenoviral vector (Hollon, 2000). However, such difficulties can be circumvented by choosing vectors and serotypes which are not likely to encounter pre-existing immunity.

Insertional mutagenesis and potentially resulting genotoxicity and activation of oncogenes may result from gene therapy with integrating vectors such as lentiviral vectors. Both integrating and non-integrating vectors may also generate phenotoxicity associated with ectopic or overexpression of the transgene. The transgene may also be silenced, especially if it is integrated into a transcriptionally inactive part of the genome or adjacent to elements subject to silencing. Risks for vertical and horizontal gene transfer must also be considered before bringing a therapy into the clinic; horizontal gene transfer may occur, for example, upon superinfection with a wild-type virus and subsequent generation of recombinant replication-competent viruses, and vertical gene transfer may arise from gene therapy resulting in the transduction of the germ cells of the recipient; both of these forms of unintended gene transfer are to be avoided at all costs. Current methods demonstrably do so, and therefore the safety aspect of gene therapy is mainly focused on the individual receiving treatment (Pauwels *et al.*, 2009; Schuettrumpf *et al.*, 2006). Altogether, gene addition therapy represents a collection of new and exciting ways to provide safe and efficient strategies for to treat and cure disease.

1.2 Adeno-associated virus

1.2.1 AAV and its life cycle

In recent years, adeno-associated virus has emerged as one of the most promising vectors for viral gene therapy. AAV is a small, helper-dependent parvovirus that has not convincingly been associated with any known human disease. It is only capable of replicative infection upon co-infection with a helper virus (adenovirus or herpes virus) (Buller *et al.*, 1981; Rose *et al.*, 1972). Infection without the presence of a helper virus may result in a latent infection by chromosomal integration.

The AAV capsid is small, non-enveloped and resistant to inactivation. It consists of a total of 60 copies of structural proteins VP1, VP2 and VP3 that are all variants of the same protein. Of these, VP1 and VP2 result from alternatively spliced transcripts, and VP3 is cleaved from VP2 in DNA-containing capsids. The capsids are icosahedral in shape, and the topology of the capsid depends on the serotype; hence, the raised regions are favoured binding sites for antibodies (Figure 1.1 A). In the mature virion, VP3 accounts for approximately 80% of the capsid protein, and VP1 and VP2 share the rest equally. Of the total weight of the virus particle, about 50 % is protein and the rest is DNA (Berns *et al*, 2007).

The first AAV to be characterised was AAV2, which was discovered in the 1960's as a contaminating agent in adenovirus preparations (Berns *et al.*, 2007). To date, 11 different AAV serotypes and more than 100 variants have been identified from adenovirus stocks, and from human and non-human primate sources (Wu *et al.*, 2006). Of these, at least AAV2 and AAV3 (which are closely related) are prevalent infections in the adult human population, as evidenced both by the presence of neutralising antibodies and also viral sequences in tissue samples (Gao *et al.*, 2004). A recent study indicates that across all ages, the prevalence of neutralising antibodies is 22% for AAV2 and 15% for AAV8 (Calcedo *et al.*, 2011).

The cell surface receptors utilised by AAV2 for cell entry include heparan sulphate proteoglycan and human fibroblast growth factor (Opie *et al.*, 2003; Qing *et al.*, 1999), however different serotypes bind different receptors and infection can involve the participation of more than one cell surface molecule.

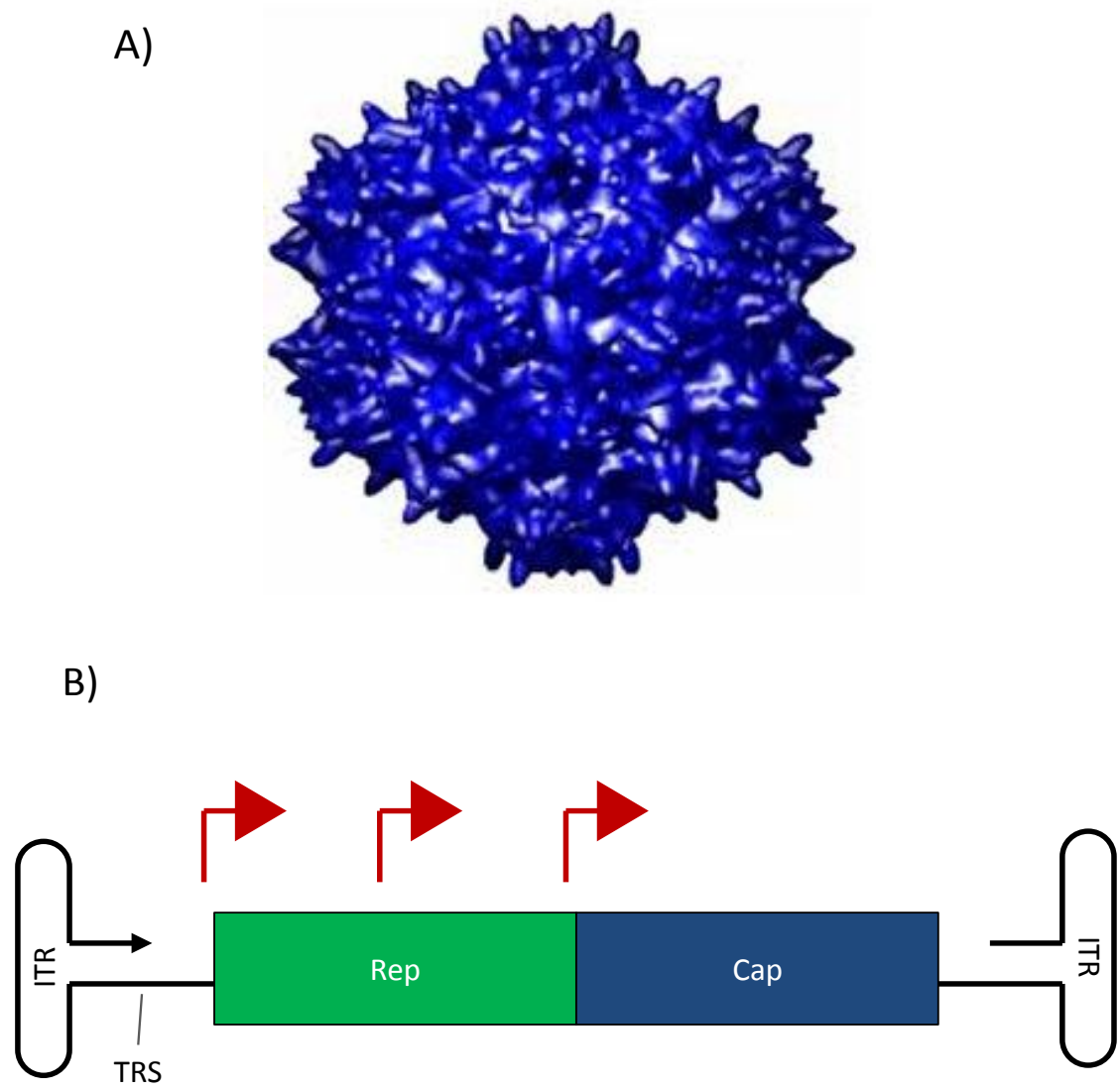


Figure 1.1 Diagrammatic representation of the AAV2 capsid (A) and genome organisation (B).

A) The AAV capsid has an icosahedral surface topology, and the extent and location of the depressions and protrusions vary between serotypes. Graphical reconstruction of the AAV viral particle topology by University of California in San Francisco Computer Graphics Lab, www.medicineworld.org.

B) The AAV genome consists of two 3-dimensional hairpin formations (Inverted terminal repeats, ITRs) flanking a single-stranded DNA genome, with 2 main genes Rep and Cap containing 3 open reading frames. Promoters denoted with red arrows; TRS, terminal resolution site.

Whether the viral genomes persist in integrated or episomal form in various cells and tissues is not yet well defined. The tissues most commonly positive for AAV sequences are muscle, perhaps due to the resistance of this tissue to helper virus infection and subsequent AAV mobilisation, and the female genitourinary tract (Gao *et al.*, 2004). AAV infection has not been linked to any known pathology.

Upon infection, the AAV capsid is internalised into endosomes which are then trafficked through the cytoplasm to the vicinity of the nucleus. A unique region of surface protein VP1 which exhibits phospholipase activity is exposed upon cell entry and is postulated to assist in capsid escape from the endosomes (Girod *et al.*, 2002). The N-terminal of VP1 also contains a nuclear localisation signal. The 5' end of the viral genome in AAV2 has been shown to protrude from the capsid, and this may serve as the template for initiating viral DNA replication with or without capsid disassembly (Summerford *et al.*, 1999).

The wild-type AAV genome is a single stranded DNA, some 4.7 kb in length, and codes for the Rep and Cap genes flanked on either side by 145 bp inverted terminal repeats (ITRs) (Figure 1.1 B). The genome contains 3 transcription units, with the two 3' proximal promoters initiating transcription of Rep genes and the 5' proximal promoter the Cap gene. Both Rep transcription units are alternatively spliced, resulting in 4 different Rep non-structural proteins involved in DNA replication: Rep 78/68 and Rep 52/40. The Cap gene encodes for the 3 capsid proteins, with VP1 and VP2 resulting from alternatively spliced transcripts and VP3 cleaved from the VP2 protein. The ITRs fold into T-shaped hairpin loop structures and are critical for viral genome packaging, stability and replication. The palindromic ITR sequences also act as primers during DNA replication (Figure 1.2 A).

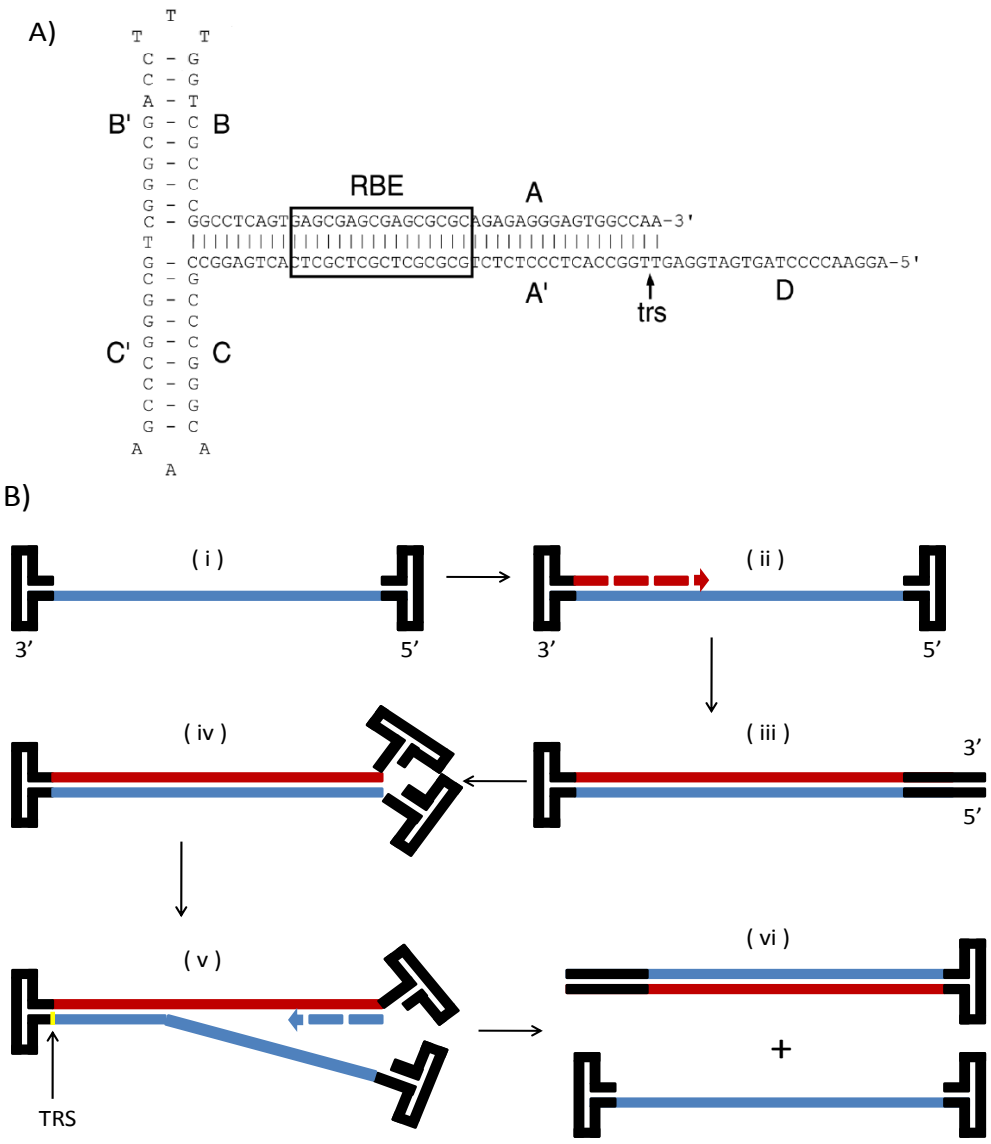


Figure 1.2 AAV2 ITRs and their role in AAV replication.

A) The ITRs function as origins of replication and form T-shaped structures consisting of two arm palindromes, B - B' and C- C', and a longer stem palindrome A - A'. The D sequence is present only once and remains single-stranded. RBE: Rep-binding element, TRS: terminal resolution site. Modified from (Goncalves, 2005).

B) The stages of AAV replication. i) A single-stranded AAV genome, ii) DNA polymerase synthesises second strand, iii) DNA synthesis goes to completion at the 5'ITR, iv) ITRs regain secondary hairpin structure, v) DNA synthesis is initiated again in the opposite direction resulting in displacement of original strand, which is separated from newly formed duplex at the terminal resolution site (TRS), vi) resulting complete single-stranded genome and duplex genome continuing replication. Modified from (McCarty, 2008).

The rolling hairpin model of AAV DNA replication stipulates that replication is initiated from the double-stranded hairpin structure at the 3' terminal repeat without the need for RNA priming (Figure 1.2 B). If the helper virus is adenovirus the DNA polymerase utilised is cellular, and in the case of herpes simplex as helper virus, AAV uses the herpes -encoded polymerase (Weindler *et al.*, 1991). DNA synthesis continues to the end of the complementary strand forming a linear molecule with a hairpin structure at one end. The viral Rep 78/68 protein then binds the terminal resolution site (TRS) close to the ITR and nicks the dsDNA of the original parental strand. The 5' end of the parental ITR is covalently joined by Rep to the 5' end of the newly synthesised strand, resulting in an inversion of the ITR.

The helicase activity of the Rep functions to actively unwind the hairpin structure at the ITR, and the resulting linear single-stranded DNA acts as a template for the completion of the second strand synthesis by DNA polymerase. Each of the strands then refold into mirror ITR structures, and DNA synthesis is re-initiated in the strand containing a double-stranded ITR structure at the 3' end as before. When the synthesis has proceeded to the end of the template, the strands are displaced to produce one single-stranded genome and one duplex structure. Each of these can then be used as a template for further rounds of DNA replication (Berns *et al.*, 2007).

Transcription of the viral genome is dependent on regulation by viral, helper and cellular factors. The switch between latent and productive infection is provided by either super-infection by a helper virus or genotoxic stimuli such as heat, ultraviolet irradiation or ionizing radiation. Of the AAV proteins, Rep 78/68 is responsible for both repression and activation of productive infection in a manner dependent on expression level and binding sites.

During non-productive infection, a small amount of Rep 78/68 is expressed which then binds its own promoter preventing further expression. Rep 78/68 bound to its own promoter also inhibits expression of the Rep 52/40 transcripts, but not the Cap genes. The promoter region for the Cap genes contains binding sites for cellular and helper virus regulatory proteins. Adenoviral helper protein E1A displaces Rep 78/68 to the ITR and allows for transcription to recommence, resulting in a productive infection. During productive infection, capsid proteins are translated in the cytoplasm and then transported into the nucleus and assembled into virions. Both plus- and minus -sense

single-stranded genomes are packaged into the preassembled virions with the help of both the large and small Rep proteins (King *et al.*, 2001).

Latent infection may occur when AAV infects a healthy cell in the absence of a helper virus. AAV2 shows a tendency towards integration into the human genome at a defined region on the long arm of chromosome 19 at 19q13.42, a site known as AAVS1(Kotin *et al.*, 1992; Kotin *et al.*, 1990). Integration does not occur at a precise site but rather anywhere within a > 1 kb region (Huser *et al.*, 2002). The region contains a Rep-binding site (RBS) and a TRS required for the integration event, which is facilitated by the Rep 78/68; in the absence of Rep, the integration occurs infrequently and at random sites. In either case, the viral genomes are usually found to have been integrated as concatemers. Concatenated AAV genomes can also survive for prolonged periods as extrachromosomal circles, possibly due to the unusual ITR structures preventing exonuclease binding and DNA modification (Nakai *et al.*, 2000a).

1.2.2 Single-stranded AAV vectors for gene therapy

AAVs are a good candidate for safe and stable episomal gene transfer, due to their natural ability to form concatenated episomal structures upon transduction into cells and lack of any known pathogenicity. The AAVs display a very wide tissue tropism, and have been used to transduce liver, skeletal muscle, retina, cardiac tissues and central nervous system. Different serotypes display differing tissue tropisms, making it possible to engineer recombinant AAVs pseudotyped with recombinant capsid serotypes to target different cell types (Li *et al.*, 2008; Wu *et al.*, 2000). Currently, recombinant AAV8 and AAV9 are the serotypes of choice for many applications due to lack of pre-existing immunity and wide tissue tropism. Vectors incorporating capsids of these serotypes also transduce many tissues more efficiently than AAV2, at least in the mouse (Wu *et al.*, 2006). Importantly, AAV9 has been shown to cross the blood-brain barrier, making it a good candidate vector for CNS gene therapy (Manfredsson *et al.*, 2009). AAV2 and AAV3 infections occur naturally in a large part of adult human population and hence vectors derived from these serotypes are likely to encounter existing neutralising antibodies. For AAV2 and AAV8 the neutralising antibody titre is at its lowest in infants aged 7-11 months, possibly indicating a suitable therapeutic window for vector administration (Calcedo *et al.*, 2011).

The AAV vectors in use today contain only the ITRs *in cis* orientation, adjoining the vector cassette. The Rep and Cap genes without ITRs and the adenoviral early gene helper functions are provided *in trans* by two additional plasmids during vector production, ensuring that no viral genes are expressed by the vector. Manufacturing vectors in this manner ensures an extremely low frequency of the likelihood that replication-competent virus is created, although rare recombination events leading to the formation of such particles may still be possible.

Like wild-type AAVs, AAV vectors form episomal concatemers by homologous recombination between the ITRs. The AAV genomes are primarily maintained in such extrachromosomal forms, as the viral Rep that usually mediates integration into the AAVS1 site is not present in the transduced cells. In the absence of the Rep protein, the rate of integration in muscle tissue is below detection (< 0.5 % of total vector DNA) and so the risk of insertional oncogenesis by the recombinant AAV vectors is low and perhaps negligible (Schnepp *et al.*, 2003).

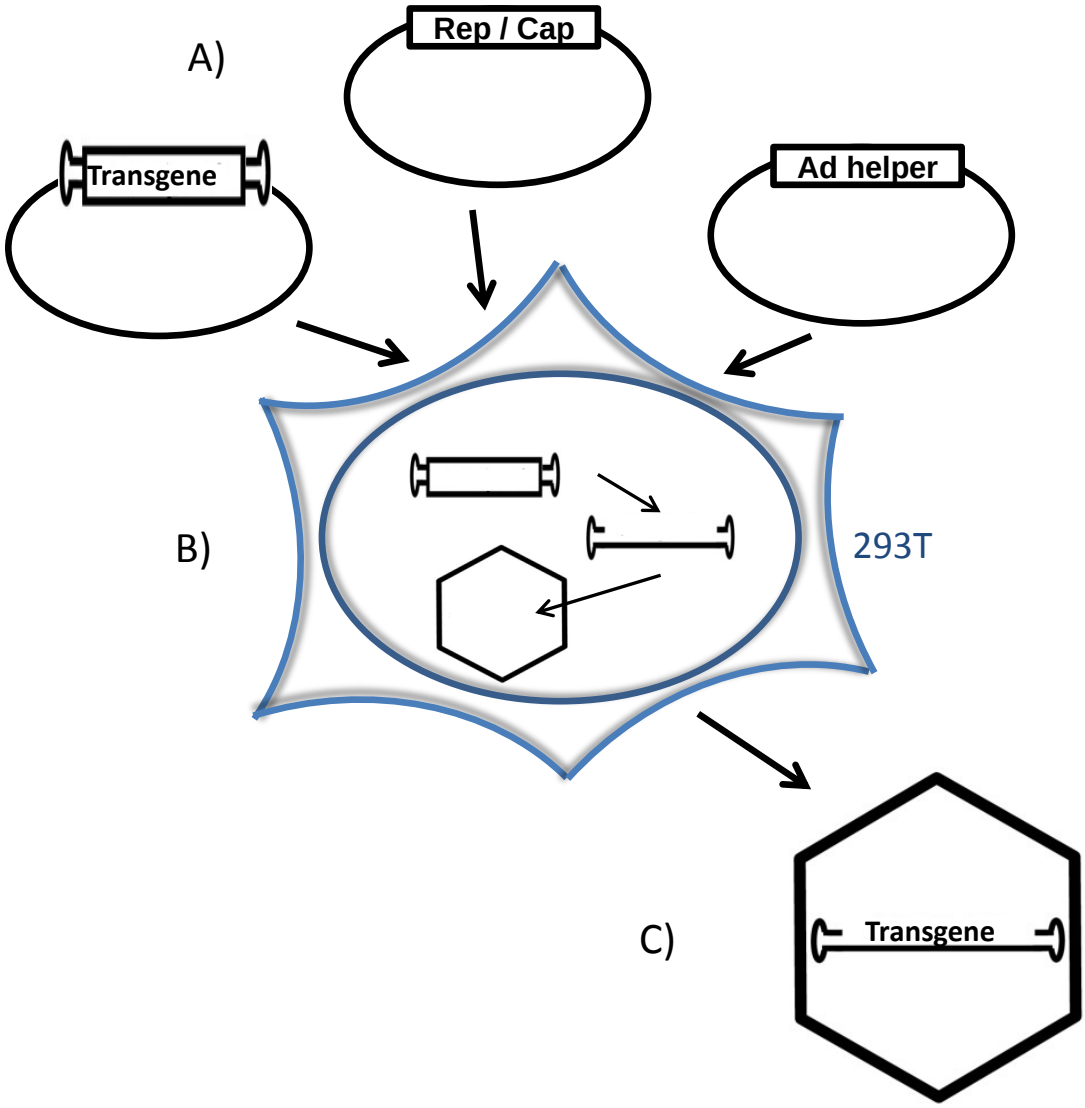


Figure 1.3 AAV vectors are produced by transient transfection of 293T cells.

A) The transgene flanked by ITRs, the Rep and Cap genes, and the adenoviral helper functions are provided on separate plasmids that are transfected into 293T cells.

B) Cellular factors and viral proteins work together to replicate the viral genome and package it into capsids.

C) Viral capsids containing the transgene are harvested from the cells.

AAV vectors are thus capable of establishing stable long-term transgene expression though episomal concatemers, with transgene expression in post-mitotic tissues observed for over 5 years post-transduction in dog liver (Herzog *et al.*, 1999). This offers an attractive alternative for long-term expression in post-mitotic tissues without integration. However, AAV genomes are maintained in the cells as extrachromosomal structures, these are diluted and often lost in cells undergoing multiple rounds of division. One possibility for overcoming the loss of the AAV transgene in dividing cells is to utilise the natural tendency of AAV to integrate at a defined site AAVS1; providing the Rep 68/78 gene *in trans* during transduction combined with a functional RBS on the transgene cassette results in a site-specific integration event at AAVS1 on chromosome 19 (Howden *et al.*, 2008). Another option is to provide the Rep protein in a hybrid Ad/AAV vector with Rep expression tightly regulated using an inducible system. This method has shown promise both *in vitro* and in mouse models (Recchia *et al.*, 2011; Recchia *et al.*, 2004). However, Rep-mediated integration is still relatively inefficient *in vivo*, and some risk of non-specific integration as well as genotoxicity resulting from double strand breaks remains.

One of the main challenges presented by AAV vectors is their limited packaging capacity, which according to current agreement is ~5 kb. In 2008 it was shown that AAV5 specifically was capable of packaging an 8.9 kb DNase-resistant genome in a sequence-independent manner and that transduction by these vectors resulted in an 8.9 kb gene product capable of ameliorating a disease phenotype in mice (Allocca *et al.*, 2008). However, subsequent studies have found that although AAV genomes larger than 5 kb can be found in transduced cells, this is more likely to be due to intracellular reassembly of partial genome fragments than packaging of large fragments into the capsid (Alcaro *et al.*, 2010; Dong *et al.*, 2010; Wu *et al.*). The efficiency of the genome reassembly may depend on the cell type and genome sequence, and larger than wild-type genomes (4.7 kb to 5.2 kb) result in at least 10-fold reduction in the titre achieved during AAV vector production (Lai *et al.*, 2009). Furthermore, the genomes of large AAV vectors have been shown to be packaged in truncated forms, although recombination after viral entry can recreate functional, large genomes (Dong *et al.*, 2010).

Another challenge for AAV-mediated gene therapy is the difficulty in high-yield vector production. Since the amount of vector needed per kilogram of weight is fairly high (e.g. 2 x 10^{12} vg / kg), producing AAV in amounts sufficient for clinical trials poses logistical challenges which are being addressed by several groups. New production methods including bioreactors and the use of baculovirus Sf9 system are making high-throughput vector production easier (Durocher *et al.*, 2007; Smith *et al.*, 2009).

1.2.3 Self-complementary AAV vectors for gene therapy

To increase expression and persistence of the AAV vector transgene in transduced cells, self-complementary AAV vectors (scAAVs) have been developed to overcome a crucial rate-limiting step of infection with conventional ssAAV vectors, which is the second-strand synthesis or recruitment of an antisense strand to form a double-stranded genome (Ferrari *et al.*, 1996). After virion uncoating, ssAAV vector genomes either have to synthesise a complementary strand or pair up with an antisense genome (Nakai *et al.*, 2000b). This process of double-strand formation can take days to reach completion, delaying the onset of gene expression, and since single-stranded genomes are unstable, leading to loss of genomes and reduced transduction efficiency (McCarty *et al.*, 2001). Conversely, once uncoated, scAAV vector genomes are able to self-hybridise by folding back upon themselves, producing an immediate double-stranded (ds) genome. The scAAV vectors have also been reported to result in higher level transduction and more rapid onset of transgene expression in several tissue types, including liver (McCarty *et al.*, 2001), retina (Yang *et al.*, 2002), brain and muscle (McCarty *et al.*, 2003).

The scAAV vector genomes are created by removing the TRS site from one of the ITRs. When the DNA is replicated by the rolling hairpin method, the replication machinery is initiated at the wild-type (wt) ITR and continues towards the mutated ITR. In the absence of a TRS or a break in DNA, second strand synthesis continues throughout the mutated ITR and back again along the genome using the opposite strand as template. The synthesis continues to the end of the wt ITR which is then resolved at the TRS, resulting in a dimeric genome with a mutated ITR in the middle (McCarty *et al.*, 2001). The scAAV genomes are packaged into AAV particles as single-stranded forms consisting of sense and antisense repeats of the expression cassette, flanked by wild-type ITRs and separated by the mutated ITR (King *et al.*, 2001; McCarty *et al.*,

2003). Upon transduction and uncoating, the single-stranded genome is released into the nucleus, and the palindromic wild-type ITRs pair up forming a double-stranded genome, with a mutated ITR hairpin structure at one end and two wild-type ITRs at the other end (McCarty *et al.*, 2001).

The most common *in vivo* forms of scAAV genomes have been found to consist of 1-2 genomes circularised though homologous recombination between the ITRs (Choi *et al.*, 2005). Intramolecular recombination, where the ITRs at the ends of one genome engage to form circles, is about 2-fold more frequent than intermolecular recombination between ITRs resulting in concatemerised genomes, and the majority of genomes were found to be circularised within 24 h of transduction (Choi *et al.*, 2005). In scAAVs, intermolecular recombination was more common between the ends of the genome containing the mutated, fully double-stranded ITRs (closed ends), rather than the other ends containing two copies of the wild type ITRs and featuring a break in the continuity of the dsDNA (open ends).

In terms of the double-stranded expression cassette, the packaging capacity of the palindromic self-complementary AAV transgene cassette is limited to approximately 2.1 kb. This places constraints on the promoters and other regulatory regions that can be used in AAV vectors. However, it is possible to divide the transgene onto two separate vectors that will be spliced together into one transcript upon co-infection into a cell (Duan *et al.*, 2001; Lai *et al.*, 2008). This is particularly of interest when developing treatment for diseases requiring large transgenes, such as Duchenne Muscular Dystrophy. However, the efficiency of trans-splicing often remains modest due to mRNA accumulation, and work is currently underway to eliminate this barrier by improving the splice site design (Lai *et al.*, 2008; Lostal *et al.*, 2010).

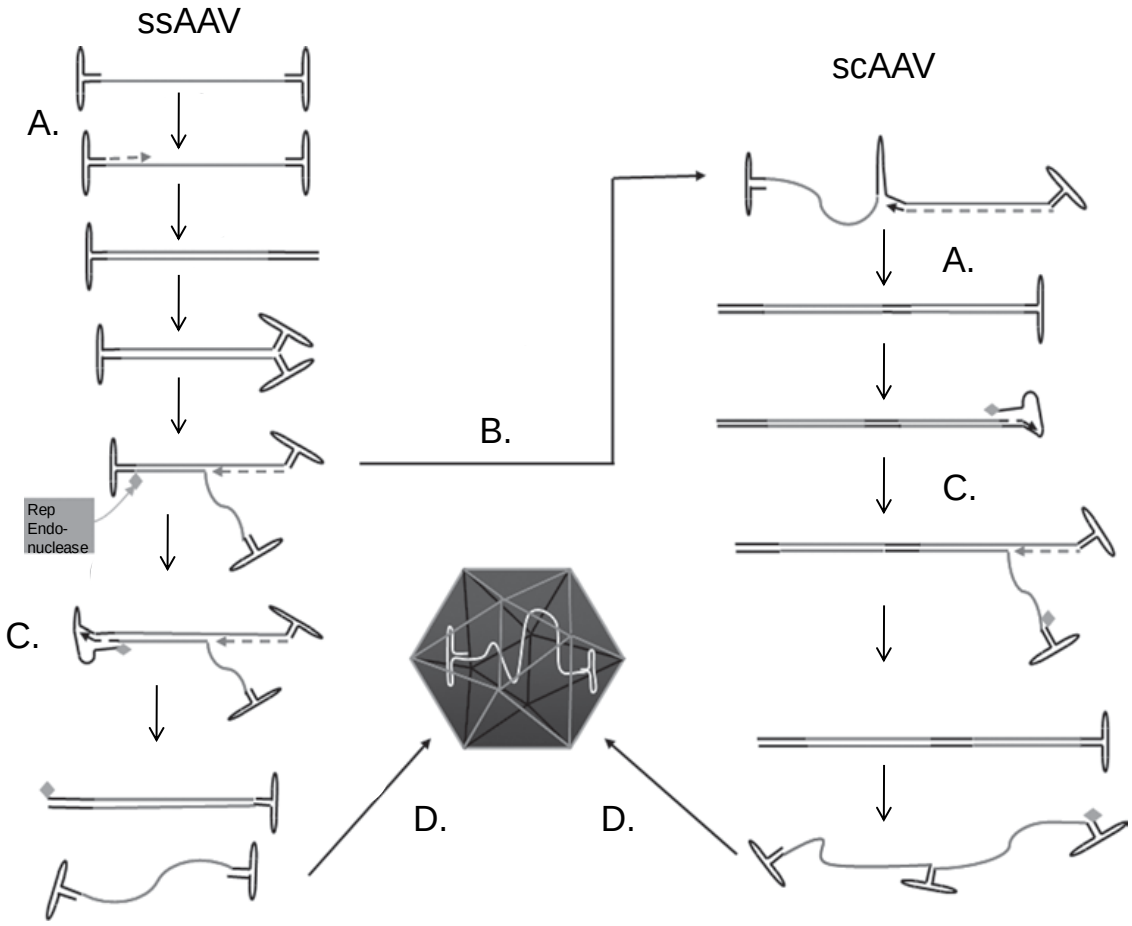


Figure 1.4 Comparison of ssAAV and scAAV life cycles, and formation of scAAV genomes.

A) Complementary strand synthesis. B) Failure to resolve mutated hairpin end leads to dimer formation. C) Terminal resolution and strand displacement. D) Monomeric and dimeric genome packaging. Modified from (McCarty, 2008).

1.2.4. Studies and trials with AAV vectors

Prior to 2005, more than 20 clinical trials using AAV2 as a vector had been initiated, mostly for the treatment of chronic diseases. The vector has proven safe, showing minimal evidence of toxicity. There are currently ~30 clinical trials registered, either completed, in progress or recruiting, using recombinant AAV as the gene delivery vector. Of these, most are Phase I studies testing the safety and dosage of the vector. Vector delivery methods vary; intra-articular delivery has been used to treat alpha-1-antitrypsin deficiency (Brantly *et al.,* 2006), intravenous delivery has been used for the treatment of haemophilia (Nathwani *et al.*, 2011), intramuscular delivery to treat muscular dystrophies (Mendell *et al.,* 2009), and by stereotactic injection into the brain for treatment of neurological disorders such as Parkinson's (Kaplitt *et al.,* 2007).

A recent Phase I clinical trial assessed the safety of delivering a minidystrophin gene for the treatment of Duchenne muscular dystrophy using intramuscular delivery of AAV2 vector (Mendell *et al.,* 2010). No moderate or severe adverse reactions were reported, but although some minidystrophin protein was detected in treated muscle 42 days post-transduction, no long-term transgene expression was achieved. This was indicated to result from cell-mediated immunity against the transgene.

Encouragingly, a Phase III study investigating the potential of a recombinant AAV2 vector to treat Leber's Congenital Amaurosis has been reported to provide safe and efficient treatment for the disease for 1.5 years after vector administration (Maguire *et al.*, 2009; Simonelli *et al.,* 2010). Improvement in clinical disability ratings was also reported in a Phase I/II study using subthalamic injections of AAV to treat Parkinson's disease (Feigin *et al.*, 2007; Kaplitt *et al.*, 2007).

Most excitingly, long-term transgene expression and amelioration of the clinical phenotype has recently been reported in a study using self-complementary AAV vectors for the treatment of haemophilia B. AAV-mediated delivery of codon-optimised FIX resulted in FIX expression at 2-11% of normal levels in all participants; 4 out of 6 patients were able to entirely discontinue FIX prophylaxis, and the remaining 2 patients also showed a clinical improvement (Nathwani *et al.,* 2011). The authors also suggested using a short course of glucocorticoids to suppress any immune reaction associated with the treatment.

1.3 Lentiviral Vectors

1.3.1 Human Immunodeficiency virus 1 (HIV-1)

HIV was identified as the cause of Acquired Immunodeficiency Syndrome (AIDS) in the early 1980s, and has since proved to be one of the most challenging pathogens in modern medicine (Gallo *et al.*, 1985). HIV belongs to the genus Lentiviruses of the *Retroviridae* family. Retroviruses are distinguished from other viruses by their extraordinary life cycle, which involves an RNA genome that is reverse transcribed into DNA and integrated into the host genome (Khoury *et al.*, 1976; Temin *et al.*, 1970). The integrated proviral form of DNA serves as a template for host transcription machinery during replication and formation of new viral particles. Structurally, mature lentiviral particles consist of a distinctive conical capsid containing the RNA genome. The capsid is surrounded by a shell consisting of a matrix, which, in turn, is coated by a lipid bilayer envelope (Barre-Sinoussi *et al.*, 1983; Gelderblom *et al.*, 1987). The envelope contains an embedded envelope glycoprotein (Env) which is crucial to the interaction with host receptors for viral entry (Figure 1.5 A).

The lentiviral genome contains 3 primary open reading frames: Gag, Pol and Env (Figure 1.5 B) (Leis *et al.*, 1988; Ratner *et al.*, 1987). These are initially translated into precursor polyproteins, which are subsequently processed by cellular or viral proteases into mature proteins (Vogt *et al.*, 1975). The Gag precursor is cleaved into the matrix, capsid, nucleocapsid and p6 proteins during or after the budding of progeny virions. The Pol polyprotein is cleaved down to yield the viral protease, integrase and reverse transcriptase enzymes, and the glycosylated Env precursor is cleaved into the surface and transmembrane proteins gp120 and gp41. It has been suggested that the heavy glycosylation of the extracellular domain of the Env protein may assist the virus in evading recognition by the host immune system (Habeshaw *et al.*, 1990). The remaining 6 HIV-1 gene products are translated using spliced mRNA and include the regulatory proteins Tat and rev, which promote viral gene expression, and accessory proteins vpr, vpu and nef involved in virulence and replication.

The HIV genome also contains c*is*-acting nucleic acid sequence elements, many of which are clustered near the 5' untranslated region. This region is highly structured,

containing several hairpin loops, and is absolutely required *in cis* for the packaging and reverse transcription of the genome. The features found here include the primer binding site (PBS), which is required for the initiation of reverse transcription, and the encapsidation signal (ψ) which consists of 4 stem loops. This packaging domain also contains the main splice donor site. Another feature is the central polypurine tract (cPPT) located in the central region of the genome. Together with the PPT found in the 3' untranslated region, it is utilised during the synthesis of plus-strand DNA. The Rev-responsive element (RRE) is another complex structure located proximal to the 3' end of the genome, and is involved in RNA transport.

HIV entry into the cells is facilitated by the Env glycoproteins, which interact with the CD4 receptors on the surface of target cells to mediate virus binding and with co-receptors of the CCR5 and CXCR4 family to promote the fusion of the viral envelope with the cell membrane (Feng *et al.*, 1996; McDougal *et al.*, 1986). After entry, the viral particles are partially uncoated in the cytoplasm and begin the reverse transcription of the RNA genomes (described below). The resulting partially double-stranded DNA is transported into the nucleus by the pre-integration complex (PIC) in a process dependent on both host and viral factors (Stevenson *et al.*, 1992). Assisted by the Integrase (*IN*) enzyme, the viral genome is integrated into the chromosomal DNA. The integrated provirus is used to produce transcripts of the viral genome, which are exported into the cytoplasm and some are spliced to produce all necessary viral proteins. The lentiviral genome is packaged into the virion as two copies of single-stranded RNA, and progeny virions form by budding from the cell membrane. Final capsid processing and maturation occurs after the release of the virions (Gelderblom *et al.*, 1987; Katsumoto *et al.*, 1988).

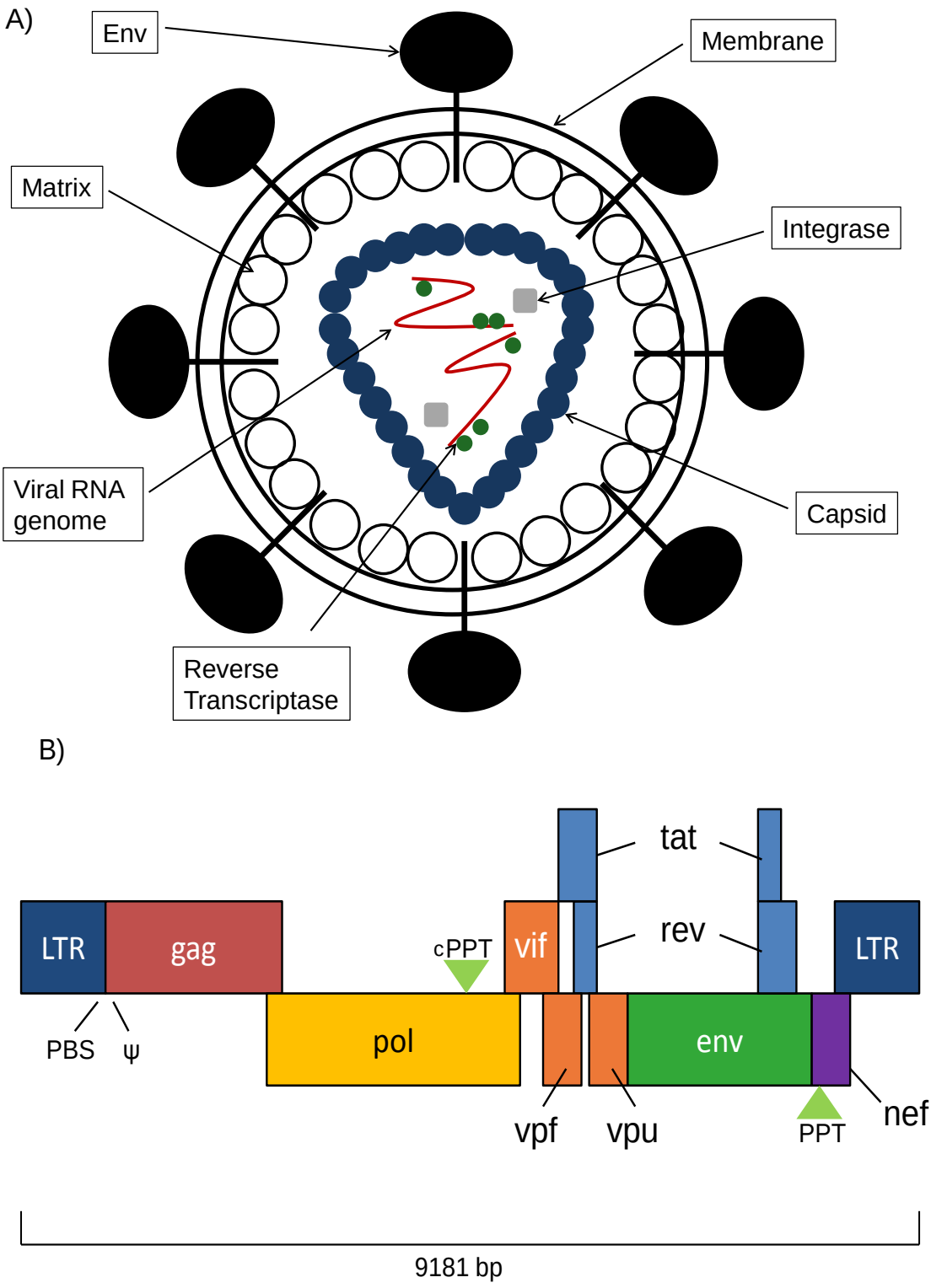


Figure 1.5 Diagrammatical representation of the HIV-1 capsid structure (A) and genome organisation (B).

PBS: Primer binding site, ψ: RNA packaging signal, PPT: Polypurine tract. Not to scale. Diagrams adapted from (Berns et al., 2007).

Reverse transcription of the retroviral genome is a key step in the process of infection. It is catalysed by the viral heterodimeric reverse transcriptase (RT) enzyme, which has both RNaseH and DNA polymerase functions, using either RNA or DNA as a template and removing ribonucleotides. The process begins by the initiation of the minus-strand DNA synthesis from tRNA bound to the primer binding site. DNA synthesis proceeds to the 5' end of the genome, at which point the RNase function of the RT digests the RNA portion of the DNA-RNA hybrid and the short DNA fragment is transferred to the 3' end of the genome where it hybridises with a short a homologous repeated region. DNA synthesis then continues until the primer binding site at the 5' end of the genome, and this is accompanied by degradation of the template RNA. The RNA fragments bound to the two PPTs in the central and distal regions of the genome are not removed by RNaseH and act as main primers for the plus-strand synthesis. RNaseH also displaces the tRNA primer, allowing the plus- and minus-strand DNAs to hybridise, and strand syntheses to go to completion.

The virus dsDNA is transported into the nucleus as part of the PIC. The processing and integration of the viral genome is catalysed by the *IN*, and although the chromosomal location of the integration event is not defined, the steps in the viral genome processing are. Integrase begins by removing two nucleotides from the blunt ends of the viral genome to create 5' overhangs. It then creates a staggered cut in the genomic DNA, and the viral DNA is joined by the overhangs to the matching recessed ends in the genomic DNA (Brown *et al.*, 1989; Fujiwara *et al.*, 1988). Cellular repair machinery assists by filling the gaps, hence the process requires both viral and host factors to go to completion.

1.3.2 The lentivirus as a gene therapy vector

In recent decades there has been growing interest in harnessing the capability of HIV and other lentiviruses for permanently infecting non-dividing cells for the purposes of gene therapy. Lentiviral gene therapy vectors have been designed such that the risk of generating replication-competent viruses is minimal, while the capacity for effectively introducing the transgene into the target cells is retained.

To overcome some of the safety concerns, the first generation lentiviral vector systems were designed such that the envelope, packaging and vector cassettes were all provided *in trans* on 3 separate plasmids (Naldini *et al.*, 1996); firstly the transfer vector plasmid contained only minimal *cis*-acting elements needed for packaging; secondly, the packaging plasmid containing all HIV genes apart from Env; and thirdly the envelope plasmid carrying the gene for vesicular stomatitis virus glycoprotein G (VSV-g). This 3-plasmid system confines the production of infectious particles to transfected producer cells and reduces the chance of producing viable recombinant virions via homologous recombination.

Second generation vectors were made safer by removing all four accessory HIV genes, *vif*, *vpr*, *vpu* and *nef*, from the packaging plasmid in a modification which was not found to affect infectivity (Zufferey *et al.*, 1997). In the third generation lentivector system, the safety profile was further improved by removing the need for Tat expression by placing a constitutive promoter near the 5' long terminal repeat (LTR) of the transfer vector, and expressing Rev on a separate, fourth plasmid (Dull *et al.*, 1998).

Another modification aimed to improve vector biosafety is the deletion of the promoter and enhancer elements in the LTR, preventing LTR-driven transcription and limiting the expression of the transgene only from the constitutive promoter in the transfer construct. These so-called self-inactivating (SIN) vectors carry a deletion in the U3 region encompassing the TATA box in the 3' untranslated region, which after reverse transcription in the host target cell is flipped into the U3 region of the 5'LTR of the proviral DNA. This deletion impairs the promoter function of the LTR and therefore prevents the expression of full-length vector RNA (Logan *et al.*, 2004; Miyoshi *et al.*, 1998; Zufferey *et al.*, 1998). The 4th generation SIN-vectors are even less likely to generate replication-competent lentivectors and have a reduced risk of mobilisation upon inadvertent super-infection with wild-type virus.

As a gene delivery vector, lentiviral vectors have several advantages over other systems; firstly, they have a relatively large packaging capacity of at least 8 Kb of DNA, which is an important feature when packaging sizeable expression cassettes of tissue-specific promoters and transgenes (Naldini *et al.*, 1996). Secondly, they differ from simpler retroviruses not only in the genome organisation but also in that they are able to transduce non-dividing cells, a very useful quality when considering their application to

non-proliferating tissues such as muscle, neurons and haematopoietic stem cells (Lewis *et al.*, 1992). Thirdly, through genomic integration of the transgene, they are capable of providing stable long-term gene expression, negating the need for repeat administration. Lentivectors also have reduced immunogenicity compared to other vectors, making it possible to consider systemic delivery routes and reducing the likelihood of anti-vector antibody formation.

Lentiviral vectors are commonly produced by transient transfection of 293T cells and and harvesting of the virus-containing supernatant (Figure 1.6). Useful titres, typically 10^6-10^7 TU/ml, are routinely achieved. Substituting the HIV-1 envelope protein with the G glycoprotein of vesicular stomatitis virus (VSV-G) confers the viral particles with a highly stable capsid structure, allowing the vector preparations to be further concentrated by ultracentrifugation. VSV-G pseudotyped vectors also exhibit a dramatically increased tissue tropism compared to wild-type lentivirus vectors, which is why it is commonly utilised for both *in vitro* and *in vivo* applications (Burns *et al.*, 1993).

In 1999 it was reported that the addition of a posttranscriptional element from the woodchuck hepatitis virus (WPRE) enhances transgene expression from lentiviral vectors (Zufferey *et al.*, 1999). It was subsequently also reported to enhance viral titres (Werner *et al.*, 2004). Its addition to the 3' untranslated region downstream of the coding region has been postulated to improve mRNA nuclear export and genomic transcription (Higashimoto *et al.*, 2007; Zufferey *et al.*, 1999). However, concerns about potential oncogenic activity of the WPRE have since been raised, and therefore the safety aspect of its inclusion in a gene therapy vector need careful consideration (Kingsman *et al.*, 2005).

In addition to HIV-based lentiviral vectors, the equine infectious anaemia virus (EIAV) has also been adapted for use as a gene therapy vector (Mitrophanous *et al.*, 1999).The main advantages of EIAV vectors are their ability to transduce non-dividing cells and the avoidance of a disease-causing entity in humans.

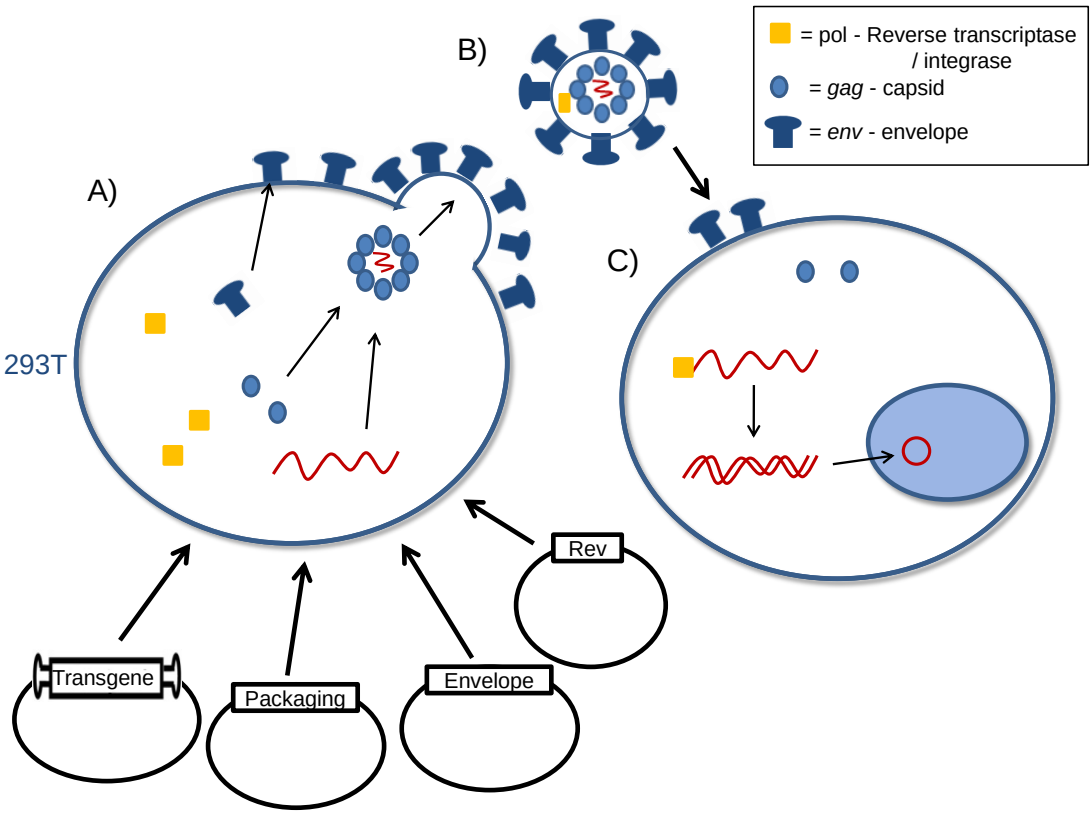


Figure 1.6 Packaging and infection by non-integrating lentiviral vectors.

A) IDLVs are produced by transient transfection of 293T cells using a 3-plasmid system. One plasmid contains the transgene, the packaging plasmid contains the gag and pol genes, and the envelope plasmid containing the gene for the envelope protein. B) Encapsidated virions are released from the cell. C) After IDLV enters the cell, reverse transcriptase converts RNA genome into dsDNA which then forms circular episomes in the nucleus.

1.3.3 Integration-deficient lentiviral vectors

HIV-1 derived vectors display a propensity for integrating close to or within transcriptionally active genes, possibly as a result of interactions between the PIC and chromatin-binding factors. Unfortunately, this also affects the potential for insertional mutagenesis and subsequent oncogenesis. As the integrating genome is more likely to interfere with gene expression, it may cause major problems when considering the safety aspects of potential gene therapy vectors.

The development of integrase-deficient lentiviral vectors (IDLVs) provides many of the advantages of previous generations of lentivectors together with an improved safety profile. IDLVs utilise the natural tendency of lentiviruses to form episomal circles as an intermediate or by-product during infection. The occurrence of these circular forms can be greatly increased by impairing the ability of the virus to integrate, and the most efficient way of doing this is by introducing mutations to Integrase (*IN*), the enzyme which catalyses the integration of the viral cassette into the host genome.

Since *IN* is also involved in several other processes including reverse transcription and nuclear import of the genome, many mutant versions result in the impairment of several functions and consequently reduced amounts of viral DNA (Engelman *et al.*, 1995). Class I mutants, in which only the DNA cleavage and integration functions of *IN* are affected, represent an improvement and result in normal levels of viral DNA (Leavitt *et al.*, 1996). Class I *IN* mutants are generated by substituting one of the three amino acids in the catalytic site of the enzyme, D64, D116, and E152. Currently, many groups favour the D64V mutation (Wanisch *et al.*, 2009).

On the other hand, an increase in safety may result in a decrease in gene expression from IDLVs in many cells types. Initial studies reported lower expression than from the integrating counterparts, however, more recently efficient expression has been achieved (Bayer *et al.*, 2008). Furthermore, non-integrating lentivectors are also protected from epigenetic silencing which may occur upon integration into an inactive part of the genome. Expression levels both *in vitro* and *in vivo* have been significantly augmented by including the SIN deletion, which removes the negative regulatory effect arising from the full-length LTR (Bayer *et al.*, 2008). Encouragingly, IDLVs have been shown to produce efficient and sustained transgene expression in some systems such as murine brain and retina (Vargas *et al.*, 2004; Yanez-Munoz *et al.*, 2006).

The main challenge presented by IDLVs is their tendency to get diluted out in dividing cells. As episomal structures they do not segregate during mitosis, and transgene expression is lost (Wanisch *et al.*, 2009). This restricts their use to non-dividing cell types such as the CNS and the retina. Also, challenges such as the potential for integration, vector mobilisation, and generation of replication competent lentiviruses are not entirely overcome by the use of IDLVs, although the risk for all of them is significantly reduced. Deep sequencing analysis of IDLV integration sites has shown that most integration events occur in non-coding regions and most likely via cell-mediated double strand break repair mechanisms rather than residual integrase activity (Matrai *et al.*, 2011).

1.3.4 Studies & trials using retroviral vectors

One of the first diseases to be treated in the clinic using gene therapy was X-linked severe combined immunodeficiency (SCID). It is caused by lack of expression of interleukin 2 receptor γ chain, resulting in a failure in T- and NK-cell differentiation. The bone marrow derived HSCs of 5 affected boys was treated ex vivo using a retroviral vector based on murine leukaemia virus (MLV), and correction of disease phenotype was soon reported in 4 of the 5 patients (Hacein-Bey-Abina *et al.*, 2002). The success of these trials understandably generated great interest around possibilities for gene therapy, and the field began to expand rapidly with treatment options for different single gene disorders being explored. However, the enthusiasm was somewhat tempered by the revelation that while 9 out of a total of 10 patients were successfully treated, 4 of the 9 acquired T-cell leukaemia as a side effect (Hacein-Bey-Abina *et al.*, 2008; Hacein-Bey-Abina *et al.*, 2003; McCormack *et al.*, 2004). In all of the cases, the leukaemia developed as proto-oncogenes were activated as a result of vector integration. It has been shown that the MLV vector used in the trials preferentially integrated near the 5' end of the growth-promoting LIM domain only 2 gene (LMO2). Overexpression of this gene followed by secondary genome rearrangements lead to signalling cascades ending up in oncogenesis (Hacein-Bey-Abina *et al.*, 2008).

Gene therapy trials to treat a different form of SCID, adenosine deaminase (ADA) deficiency, were also launched in the 1990s. In the trial, initiated more than 20 years ago, the *ADA* gene was delivered into T-cells using a retroviral vector. Four years after the initiation of the trial, both of the 2 patients were reported to have experienced

marked clinical improvement (Blaese *et al.*, 1995). Interestingly, successful long-term treatment of SCID-ADA has been reported without observations of the clonal lymphoproliferation observed with X-SCID despite integration events near the same proto-oncogenes, raising questions about the potentially oncogenic role of the transgene used to treat the X-linked form of SCID (Woods *et al.*, 2006). Recently, two clinical trials using *ex vivo* retroviral transduction of the *ADA* gene into CD34+ bone marrow followed by infusion of the transduced cells back into the patients concluded that 8 out of 10 patients did not need enzyme replacement therapy and have not experienced adverse effects to date (Aiuti *et al.*, 2009).

Lentiviral vectors have been used for transduction of HSCs for the treatment of X-linked adrenoleukodystrophy. Successful amelioration of the clinical phenotype was reported in a trial of 2 patients whose autologous cells were transduced *ex vivo* with a lentivector encoding for a wild-type *ABCD1* gene, followed by re-infusion of the transduced cells after myeloablative treatment (Cartier *et al.*, 2009). In another study using lentiviral vectors for *ex vivo* treatment of stem cells, boys suffering from the Wiskott-Aldrich syndrome were treated using autotransplantation of lentivector – transduced haematopoietic stem cells. Marked improvement of the patients' clinical condition was reported (Boztug *et al.*, 2010). In another recent success story, a clinical trial using lentivectors to transduce HSCs for the treatment of β-thalassemia reported transfusion independence of the patient 33 months after treatment (Cavazzana-Calvo *et al.*, 2010).

Due to their ability to efficiently deliver genes to post-mitotic tissues, EIAV-based lentivectors are the vector of choice for targeting the central nervous system. Pre-clinical studies have suggested the efficacy of lentiviral vectors for the treatment of diseases such as spinal muscular atrophy and Parkinson's disease (Ahmed *et al.*, 2010; Azzouz *et al.*, 2004). A Phase I/II clinical trial assessing the safety and efficacy of an EIAV vector for the treatment of Parkinson's is underway (Schambach *et al.*, 2008).

1.4 Non-viral vector systems

1.4.1 Overview of non-viral vector systems

As an alternative to engineering viruses to act as gene delivery vectors, naked DNA can also be used in gene transfer. The main advantage of non-viral gene delivery over viral methods is the possibility for repeat administration, as limited immune response is generated by the vector itself. Also, in theory there is no strict limit for the size of the transgene cassette, so even entire genomic loci might be used. DNA delivered into the cells by non-viral methods does not integrate into the genome unless very large amounts of DNA per cell or a specific integration mechanism are used, and therefore it may be considered safer than retroviral transfer methods (Ledwith *et al.*, 2000; Wang *et al.*, 2004).

To deliver the DNA into the target cells, either chemical or physical delivery methods must be employed. Chemical methods include complexing DNA with agents such as cationic lipids or co-polymers, or encapsulating it in liposomes. Physical delivery methods such as electroporation and sonoporation involve temporarily disrupting the cell membrane, thus allowing the entry of DNA into the cells; these methods are currently limited to use *in vitro* and to easily accessible tissues such as skin and muscle. Other physical methods include aerosolisation for gene transfer into the lungs, and hydrodynamic delivery to the liver in pre-clinical systems (Herweijer *et al.*, 2003).

Intracellular targeting of the DNA into the nuclei can be limiting, especially in non-dividing cells where the nuclear membrane remains intact. Nuclear entry can be improved by the attachment of nuclear localisation signals to the plasmid, enabling transport through the nuclear pore complexes (Miller *et al.*, 2009). For purposes of gene addition therapy, not only is it desirable for the plasmid DNA to enter the target nuclei, it also needs to remain there long enough to produce sufficient amount of gene product to correct the disease phenotype. For mitotic stability and permanent transfection to occur, a functional origin of replication needs to be present in the plasmid. Several plasmid systems have been designed which rely on viral origins of replication, activated by the expression of appropriate viral gene products. Such systems include the EBV-based vectors containing OriP and EBNA1, and vectors containing OriT and the SV40 large T-antigen (Van Craenenbroeck *et al.*, 2000).

However, the expression of viral *trans*-acting factors like EBNA1 and large T-antigen in mammalian cells can be problematic, as this can lead to the immortalisation of the transduced primary cells and subsequent tumour formation (Di Mayorca *et al.*, 1969). Alternative systems for episomal maintenance of non-viral vectors that do not rely on viral factors, such as the pEPI plasmid vector (see section 1.5.3), are therefore attractive from a safety perspective.

1.4.2 Applications for non-viral vectors

The main challenge facing non-viral DNA delivery methods is the relatively low efficiency of gene transfer. Whilst serviceable transfection efficiencies can be achieved with certain cell lines *in vitro, in vivo* applications are currently limited due to relatively poor efficiency of gene transfer. This is partly due to the more sensitive nature of the primary cells and tissues, and also the difficulty with getting the DNA close enough to the target tissues for efficient entry by endocytosis or permeabilisation and transfer across the cell membrane. As a result, non-viral approaches are particularly suitable to treat tissues easily accessible to vectors, such as skin, lungs, muscle and the intestine.

As DNA is negatively charged, it can form condensed particles with cationic lipids or polymers. Such lipoplexes or polyplexes can then be induced to associate with cell membranes, enter cells and carry the DNA with them. However, the charged DNA complexes have a tendency to aggregate in biological fluids and lack tissue specificity. Liposomes, small vesicles composed of phospholipid bilayer, can somewhat overcome these problems. They can be targeted to specific cell types via surface receptors and deliver their contents by fusing with the target cell membrane, dispatching the DNA contained within. Liposome targeting can be further improved by incorporating ligands into their structure (Gopal *et al.*, 2011). However, although lipofection is an efficient *in vitro* method of transfection, the efficiency *in vivo* is fairly low, reducing its applicability to epithelia and tissues directly underneath. In addition, the DNA-lipid complexes can induce significant toxicity when employed *in vivo*, which limits the dose that can be used (Barteau *et al.*, 2008).

Lipofection is currently being employed for gene delivery into both skin and lung tissues, both of which do not require extensive vector dissemination post-entry. Plasmids complexed with various gene transfer agents such as cationic lipids can be used to deliver DNA into the pulmonary epithelium, and have been shown to result in relatively efficient gene transfer into the lung although it is associated with significant toxicity (Emerson *et al.*, 2003). Liposomes can be delivered into the lungs using aerosols. Liposome-based gene transfer methods are currently being developed through pre-clinical and safety studies in preparation for the clinic (McLachlan *et al.*, 2011). Research into gene delivery into the skin is mainly concentrated in the area of vaccinations, as the high occurrence of antigen-presenting cells lends itself well to immunisation. Lipofection of liposome –encased antigens into the skin presents an attractive alternative to subcutaneous or intramuscular injections (Hansen *et al.*, 2011). Non-viral vectors have also been used to deliver therapeutic genes into skeletal muscle, using chemical methods as well as electroporation and direct injection (Braun, 2008). The advantages of skeletal muscle as a target for gene therapy include the possibility for treating non-muscle related illnesses as the tissue is able to secrete proteins into the bloodstream. Myofibers are also mostly post-mitotic, enabling long-term retention of even non-integrating vectors.

In pre-clinical mouse models, hydrodynamic injection can be used to deliver DNA to the liver. It entails a rapid injection or infusion of a large volume of naked DNA solution into the tail vein, resulting predominantly in the transfection of hepatocytes, likely via both receptor-mediated and mechanically induced pathways (Niidome *et al.*, 2002). Although the technique is easier to execute in smaller animals such as rodents, advances have been made in translating the method for use in larger animals such as pigs (Fabre *et al.*, 2008). Computer-controlled hydrodynamic injections have been developed for gene delivery to the pig liver, to enable exactly the right speed of injection of the larger volumes necessary to achieve the correct pressure in the target tissue (Suda *et al.*, 2008). A Phase I clinical trial using hydrodynamic gene delivery has demonstrated the safety and feasibility of this approach also in humans (Khorsandi *et al.*, 2008).

1.4.3 Plasmid vectors

Non-viral DNA delivery vectors are commonly based on plasmids as they can readily be engineered to contain the transgene cassette of choice, propagated in bacteria and purified. Such vectors contain a bacterial origin of replication as well as a selection marker, usually for antibiotic resistance. In addition to the transgene cassette, plasmid vectors have been engineered to contain elements to allow for either episomal persistence or for integration into host genome, in order to prolong and enhance transgene expression (Van Craenenbroeck *et al.*, 2000).

One episomal vector system is derived from Simian Virus 40 (SV40), where the plasmid contains a viral replication origin which is activated in the presence of the large T-antigen, a viral protein. Another popular episomally retained vector is based on the Epstein-Barr virus (EBV) and contains a viral origin of replication, OriP, and a corresponding activating factor, EBV nuclear antigen 1 (EBNA1). Both OriT/T-antigen and OriP/EBNA1 vectors replicate and segregate during mitosis and can be used to provide long-term gene expression even in dividing tissues. However, their translation into the clinic is hampered by the potential oncogenic properties of the viral factors. The SV40 vector has been found to integrate into the host genome and the T-antigen can be oncogenic (Strayer *et al.*, 2002). Similarly, EBNA-1 has been implicated as an oncogene (Schulz *et al.*, 2009). EBV is also the causative agent of infectious mononucleosis, which raises more potential risks both during manufacture and treatment (Conese *et al.*, 2004; Perez-Luz *et al.*, 2010).

Integration into the host genome can be achieved through the expression of a transposase and a responsive element, such as the Sleeping Beauty transposon system. In theory, it combines the advantages of a plasmid-based system such as large capacity and ease of vector production with the long-term expression potential of an integrating viral vector system. Nevertheless, the efficiency of integration remains significantly lower than that of retroviral vectors, and is mainly being explored for *ex vivo* gene therapy (Ivics *et al.*, 2011).

Tissue specificity of plasmid vectors can be improved by the use of tissue-specific promoters, which may also improve transgene retention *in vivo* by reducing immune response resulting from ectopic gene expression, especially in antigen-presenting cells.

For instance, plasmid-based vectors have been shown to achieve high levels of clotting factor IX expression both *in vivo* and *in vitro* through the use of the human α-1 antitrypsin promoter and additional *cis*-acting control regions (Intra *et al.*, 2011; Miao *et al.*, 2000).

1.4.4 DNA minicircle vectors

Minicircles are vectors derived from plasmids but eventually lacking prokaryotic backbone sequences such as antibiotic resistance markers and bacterial origins of replication. It has been shown that the presence of bacterial sequences can lead to the elimination of the vector in mammalian cells (Chen *et al.*, 2003). Moreover, minicircles provide enhanced transgene expression compared with plasmids, ostensibly due to lack of heterochromatin formation and resulting silencing caused by CpG islands (Riu *et al.*, 2007). Unlike mammalian DNA, plasmid DNA that has been propagated in bacteria contains unmethylated CpG islands. These motifs may cause an inflammatory response in mammalian cells, making them somewhat unsuitable for inclusion into gene delivery vectors (Bauer *et al.*, 2001; Hyde *et al.*, 2008).

Minicircles are created from parental plasmid vectors through an intramolecular recombination process in which two smaller circles are formed from one larger one, resulting in one minicircle containing the transgene cassette and a miniplasmid containing the remaining bacterial sequences (Figure 3.2). Several alternative systems have been devised to achieve a reliable and controllable recombination system for the creation of the minicircles; bacteriophage λ integrase, *Cre* or FLP recombinase or, most recently, φC31 recombinase have all been utilised (Bigger *et al.*, 2001; Chen *et al.*, 2003; Darquet *et al.*, 1997). The recombination reactions can be either uni- or bi-directional; unidirectional reactions occur between two different sequences resulting in a modified site and cannot be reversed, whereas bidirectional reactions occur between identical sequences resulting in a site that can act as a template for another recombination reaction.

C*re* recombinase, derived from bacteriophage P1, catalyses site-specific recombination between two 34 bp long direct repeats (LoxP sites) (Hoess *et al.*, 1985). An analogous method uses the FLP (Flippase) enzyme derived from the 2μm plasmid of yeast which catalyses recombination between two 34 bp FRT (Flippase recognition target) sites (Jakobsen *et al.*, 2010). Both of these methods catalyse recombination between identical

sites, and thus the reaction is both reversible and bidirectional. Many recombination systems, including *Cre*, also result in the creation of concatemeric structures containing several minicircle genomes. Concatemer formation can be reduced by modifying the *LoxP* sites, which also shifts the recombination equilibrium in favour of minicircle formation (Araki *et al.*, 1997).

Bacteriophage λ recombinase catalyses recombination between two attachment (*att*) sites, *att*B (originally derived from bacterial genome) and *att*P (derived from phage). The reaction requires two additional host proteins, IHF and FIS, and results in a supercoiled minicircle containing a recombined site *att*R (Darquet *et al.*, 1997). The reaction catalysed by the λ recombinase is unidirectional. However, concatenated dimeric genomes are generally found contaminating the final product, suggesting this method may not be suitable for the production of pure monomeric minicircles (Darquet *et al.*, 1999).

The φC31 integrase from *Streptomyces* temperate (integrating) phage catalyses recombination between *att*B and *att*P sites, which have minimal functional sizes of 34 and 39 bp, respectively (Groth *et al.*, 2000). As the reaction catalysed by φC31 is unidirectional, it is in theory more efficient than recombinases catalysing a bidirectional reaction; however, the initially observed recombination efficacy of the enzyme was low (Chen *et al.*, 2003). An improved protocol utilised 2 copies of the φC31 recombinase as well as an optimised medium and incubation temperature, overcoming the concerns regarding sufficient efficacy (Chen *et al.*, 2005).

The expression of the recombinase needs to be temporally controllable, so that the production of the parent plasmid is not hampered by untimely recombination events. Placing the gene for the recombinase under the control of the arabinose control regulon (araC) enables expression only when the growth medium is supplemented with L-arabinose. The system has been shown to inhibit unregulated recombinase expression, and is favoured by many laboratories (Bigger *et al.*, 2001; Chen *et al.*, 2003).

After the recombination event, the minicircles need to be separated from the miniplasmids and remaining parental plasmids. This can be done by using a restriction enzyme to linearise and/or fragment the mini- and parental plasmids, and the supercoiled minicircle is then labelled with ethidium bromide or propidium iodide and separated from the other two species by cesium chloride density gradient ultracentrifugation. The intercalating agent is removed by running the minicircle DNA

through a cation exchange column (Bigger *et al.*, 2001), or by isopropane extraction followed by ethanol precipitation (Chen *et al.*, 2003). However, the cost-to-yield ratio of these methods are high due to the several stages and processes involved, and density gradient –based methods are not acceptable for the production of clinical grade gene transfer agents. Thus the rate limiting step in the development of minicircle gene therapy has been the difficulty in large-scale production, and options to improve this are currently being explored (Schleef *et al.*, 2010).

Another approach is to remove the mini- and parental plasmids from the producer *E.coli* cells after transfection by including a conditional restriction enzyme that targets the backbone sequences, thus avoiding the cumbersome purification stages altogether. The minicircle producer plasmid contains the transgene cassette flanked by *att*P and *att*B sites and a set of araC -inducible genes including two copies of the gene encoding φC31 recombinase, and a gene for the I-SceI homing endonuclease. After induction by L-arabinose, efficient recombination between the *att* sites is provided by the recombinase, whilst I-SceI initiates the degradation of the plasmid backbone by host exonucleases by making a double-stranded nick in the engineered I-SceI site (Chen *et al.*, 2005). More recently, a stable *E.coli* producer strain containing the transgene cassette, *att* sites and the inducible recombination operon has been developed to circumvent the transfection stages. In both cases, minicircle DNA can be purified directly from the producer cultures using standard affinity column purification, at an efficiency comparable to standard plasmid production (Chen *et al.*, 2005; Kay *et al.*, 2010).

Many of the advantages of minicircles over plasmid vectors are related to their small size. As they are about half the size of standard plasmid vectors, the complexes of minicircles and cationic lipids are also significantly smaller in size, which may result in increased endocytosis as well as intracellular mobility and consequently transfection efficiency (Darquet *et al.*, 1997; Zabner *et al.*, 1995). Transgene expression from minicircle vectors has been reported to be 10 – 1000 fold higher than from standard plasmid vectors, both *in vitro* and *in vivo* (Chen *et al.*, 2003; Jia *et al.*, 2010). Moreover, minicircles containing S/MAR elements have been reported to achieve high levels of long-lasting transgene expression in the absence of antibiotic selection (Broll *et al.*, 2010; Rangasamy, 2010).

Hydrodynamic delivery of minicircles has been used to transport DNA into the mouse liver, with some very promising results. A minicircle encoding the FIX gene driven by

the human α_1-antitrypsin promoter was resistant to silencing by methylation compared to an equivalent plasmid vector (Schuttrumpf *et al.*, 2011). Long-term phenotype correction was also achieved in mice with mucopolysaccharidosis type I when a minicircle delivery of the *IDUA* gene was coupled with a co-stimulatory pathway blockade and the inclusion of microRNA target sequences (Osborn *et al.*, 2011). Recently, minicircles have been used to provide the necessary factors to generate induced pluripotent stem cells from adult human adipose stem cells (Jia *et al.*, 2010).

Minicircles are a safe alternative compared to plasmid-mediated gene transfer due to the lack of antibiotic markers as well as most bacterial sequences that could act as templates for recombination with wild-type bacteria. That said, the *attR* site present in minicircles generated using bacteriophage λ recombinase could interact with an *E.coli* genome containing the same sequence, and the minicircle could become integrated into the bacterial genome through homologous recombination; however, this would require two low probability events (the lysis of an infected *E.coli* and intermolecular recombination) to take place in the same cell (Darquet *et al.*, 1997). The φC31 system could potentially present with the same situation, something which needs to be considered before bringing the minicircle system into the clinic.

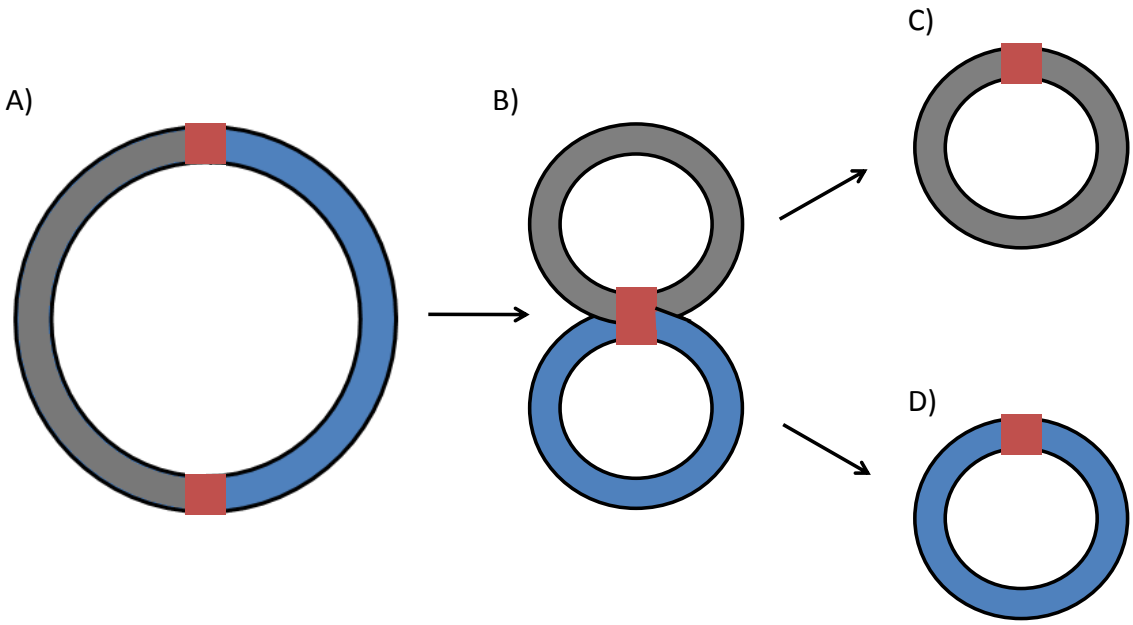


Figure 1.7 Production of minicircles.

The minicircles are produced from a parental plasmid (A) by intramolecular recombination between two recombination sites (B) facilitated by a catalysing enzyme. The resulting molecules are the minicircle (C) containing the transgene cassette (in grey) and little bacterial sequences, and the miniplasmid (D) which contains the bacterial origin of replication and antibiotic resistance genes. Figure modified from PlasmidFactory McBox insert.

1.5 Scaffold/Matrix Attachment Region

Important considerations when delivering a non-integrating transgene for stable gene therapy are how to (i) sustain long-term gene expression without silencing, and (ii) induce regulated episome replication during S-phase, and (iii) ensure efficient episome segregation into the daughter cells during mitosis. From the point of view of nuclear architecture, the key mediators of efficient transgene expression and retention are the inclusion of necessary regulatory elements into the cassette, and maintaining the cassette in a state of active chromatin by controlling DNA methylation and histone acetylation states.

In the nucleus, transcriptional activities are localised into compartments where RNA polymerase is complexed with the nucleoskeleton together with active chromatin regions. Different regions of the nucleus are associated with different epigenetic states of chromatin, and active regions are grouped together in such transcription compartments (Cook, 1999). In the case of transgene expression from extrachromosomal episomes, it is important to target the episome to such nuclear regions to enable efficient transcription.

Episomally maintained transgenes, whether delivered into the cell by non-viral vectors, AAV vectors or IDLVs, are a safer option compared with using integrating vectors. However, an inherent problem that arises with this strategy is the loss of episomes and hence transgene expression through cell division. This problem needs addressing, especially when the therapeutic aim is the treatment of a disease that requires the transgene to be expressed in dividing cells such as hepatocytes. Thus far, one of the most promising methods is the inclusion of a scaffold/matrix attachment region (S/MAR) in the episomal cassette.

1.5.1 Functions of S/MARs

S/MARs are structural components of eukaryotic DNA involved in arranging chromatin into loops or domains. Each domain represents an independent regulatory unit, flanked by bordering elements that insulate the transcription unit from outside influences and limit the range of action for enhancers within the domain. These bordering elements,

S/MARs, are architectural components of genomic DNA that anchor each domain into its place by interacting with components of the nuclear scaffolding, the intranuclear framework made of proteinaceous filaments that provides support for chromatin loops and participates in processes such as DNA replication and transcription.

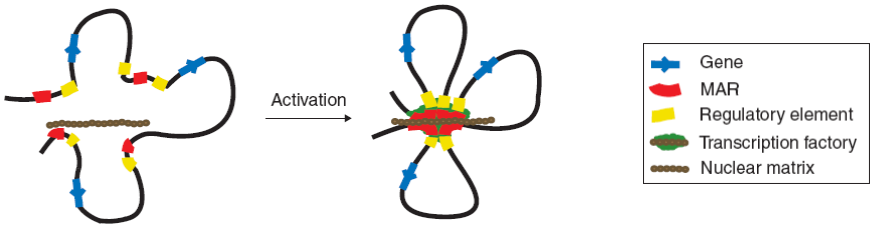


Figure 1.8 A diagram of S/MAR function in transcription.
Upon activation, the S/MAR elements bind the nuclear matrix, bringing together gene coding sequences, regulatory elements and the transcription machinery. Modified from (Ottaviani et al., 2008).

S/MARs are principally defined by their affinity to proteins associated with the nuclear scaffolding, such as SAF-A and SAF-B scaffold attachment factors and DNA-binding proteins shown to bind S/MARs exclusively (Kipp *et al.*, 2000; Renz *et al.*, 1996; Romig *et al.*, 1992). They have also been identified as major components of the nuclear matrix crucial for nuclear architecture *in vivo*, facilitating the attachment of chromatin to other matrix proteins. The association of chromatin with the nuclear matrix is also thought to be crucial in mitotic segregation, with the matrix filaments collaborating with the spindle fibres in guiding chromatids to the daughter cells.

Although the characteristics of S/MARs are shared in all eukaryotes, common sequence motifs have not been found, and the homologies appear to be structural (Benham *et al.*, 1997; Bode *et al.*, 2006). A common feature of S/MARs is a high AT-content (~70%), contributing to an increased propensity for strand separation. This may lead to the formation of secondary structures such as hairpins, which are recognised by numerous enzymes such as topoisomerases and polymerases, and also enzymes involved in histone modification. Accordingly, a large number of proteins have been shown to bind to S/MARs, including topoisomerase II, histone H1, as well as matrix proteins lamin B1 and SATB1 (Jenke *et al.*, 2002).

S/MARs participate in several DNA processes, including the initiation of replication (Bode *et al.*, 1992; Schaarschmidt *et al.*, 2004). Through their function in organising DNA into domains they can act as insulators, by keeping each transcription unit separated and establishing local access of transcription factors to enhancer and promoter sequences. S/MARs have been shown to augment transcription rates by shielding integrated transgenes from negative environmental influences (Goetze *et al.*, 2005), and they can support histone acetylation thereby decondensing the domain and activating transcription (Rupprecht *et al.*, 2009). S/MARs can prevent silencing by inhibiting *de novo* methylation and also alleviate pre-existing methylation, resulting in improved transgene expression (Dang *et al.*, 2000; Rangasamy, 2010).

1.5.2 Utilising S/MARs in episomal vector systems

The implied ability of S/MARs to initiate replication and enhance gene expression has led to investigations in their use in transgene cassettes to assist in the establishment of mitotically stable, replicating non-integrating episomes. Many of the studies into the structure and function of S/MARs have concentrated on the strong 5kb S/MAR element downstream of the human interferon β (IFNB1) gene on chromosome 9 (Bode *et al.*, 1988; Mielke *et al.*, 1990). One of the first S/MAR elements to be characterised, it has been incorporated into a variety of vectors. A CMV-GFP plasmid construct incorporating this element was found to be episomally maintained in transfected CHO cells for over a hundred generations without selection pressure, albeit at a very low copy number (Piechaczek *et al.*, 1999). Several follow-up studies combining this S/MAR element with tissue-specific promoters in plasmid-based vectors found that the S/MAR was able to provide persistent expression by preventing transgene silencing, although the episomal status of the vector or its mitotic retention was not necessarily unequivocally established (Argyros *et al.*, 2008; Lufino *et al.*, 2008; Sgourou *et al.*, 2009). It appears that the cell type and affiliated sequences influence the potential for episomal retention induced by S/MARs, and it is possible that this element will turn out to have limited applicability.

An *in vitro* investigation searching for mammalian origin of replication sites has led to the discovery of a potential consensus sequence in episomally replicating DNA circles

(Price *et al.*, 2003). Plasmid DNA containing the consensus sequence was maintained under selection, and highly retained without selection in HeLa cells. Whether this sequence could confer episomal stability in other systems remains to be explored.

A synthetic tetramer of a 155-bp module that comprises the core unwinding element of the IFNB1 S/MAR was able to functionally replace the larger original element and when linked to an upstream transcription unit, its introduction into the cell in a plasmid construct resulted in replication and episomal retention in the absence of selection in CHO cells (Jenke *et al.*, 2004). A naturally occurring truncated mutant also based on the β-IFN S/MAR element has recently been shown to enable episome establishment in CHO cells using plasmid-based minicircles, which are small non-viral episomes devoid of any bacterial sequences (Broll *et al.*, 2010). However, in both cases an initial period of antibiotic selection was required to acquire stable episomes, something which cannot easily be applied when transferring the method to clinical gene therapy.

There is some evidence indicating that the S/MAR element can have synergistic effects depending on the promoter included in the expression cassette, such that the S/MAR element is more efficient in protecting against silencing when combined with tissue-specific promoters rather than ubiquitous ones like CMV (Argyros *et al.*, 2008). It has also been shown that not only is the S/MAR linked to the transcription unit but the transcription needs to run into the S/MAR element, presumably to maintain an open chromatin conformation (Rupprecht *et al.*, 2010; Stehle *et al.*, 2003).

The vast majority of the studies published so far on S/MAR -containing vectors have utilised non-viral delivery systems such as plasmids and minicircles. One study investigated the potential of the S/MAR element to act as an insulator against transgene silencing in integrating lentiviral vectors and found the S/MAR element resulted in significantly improved long-term transgene expression when combined with another insulator element (Ramezani *et al.*, 2003). In another study, a minicircle containing the SV40 origin of replication and the S/MAR element was delivered into the liver by hydrodynamic injection, but no episomal replication was observed following a partial hepatectomy (Jacobs *et al.*, 2008). Furthermore, no studies have been published to date using S/MAR vectors to assist in mitotic retention of and long-term transgene expression from non-integrating viral episomes.

The insulator properties of the S/MAR element have been utilised in constructing integrating retroviral vectors to maintain open chromatin conformation at the site of the transgene. Inclusion of the S/MAR element from immunoglobulin-κ gene in a lentiviral expression cassette transduced into mouse hepatocytes *in vitro* resulted in significantly increased transduction efficiency, as measured by both transgene expression and proviral DNA copy number, compared to the non-S/MAR containing vector. *In vivo* transduction of the S/MAR -containing vector following a partial hepatectomy resulted in sustained therapeutic levels of transgene expression by the liver (Park *et al.*, 2001).

1.5.3 The pEPI plasmid vector

The pEPI plasmid vector has been widely investigated as an episomally replicating plasmid vector not dependent on viral factors. The vector is based on the commercial pGFP-C1 vector with the addition of the S/MAR element from the 5' region of the human β-IFN gene, placed downstream of the GFP gene driven by the CMV promoter (Piechaczek *et al.*, 1999). The plasmid also contains a second mammalian transcription unit with the neomycin phosphotransferase gene driven by the SV40 promoter, for the ease of antibiotic selection of successfully transfected cells.

Instead of utilising the OriT / large T-antigen system required for the replication of the parent plasmid, the pEPI plasmid was postulated not to need viral *trans*-acting factors as the S/MAR element would carry out equivalent tasks relating to Ori functions instead. This turned out to be the case, and the original pEPI vector was found to replicate episomally for hundreds of cell generations; following an initial period of selection. Long-term transgene retention has been reported in Chinese hamster ovary (CHO) cells, HeLa cells, and K562 cells (Jenke *et al.*, 2004; Piechaczek *et al.*, 1999; Schaarschmidt *et al.*, 2004), and the vector is maintained in circular extrachromosomal structures at a copy number of 5-10 per cell (Baiker *et al.*, 2000).

The pEPI vector has also been shown to successfully provide long-term gene expression *in vivo*. A vector incorporating the α-1 antitrypsin promoter was used to transfect murine liver by hydrodynamic injection, and luciferase expression was observed in the tissues for 6 months following delivery (Argyros *et al.*, 2008). However, a partial hepatectomy resulted in an almost complete loss of transgene expression, indicating that initial selection pressure as applied in the *in vitro* studies is required for the establishment of the episomes (Argyros *et al.*, 2008). In contrast, stable *in vivo*

expression from a pEPI vector was achieved in pig foetuses following sperm-mediated gene transfer during fertilisation (Manzini *et al.,* 2006). The development of an *in vivo* selection strategy for the pEPI vector resulted in proliferating hepatocytes in the adult mouse retaining the S/MAR -containing vector when provided with a selective advantage, emphasizing the importance of selection pressure during the establishment of pEPI vectors (Wong *et al.*, 2011).

A modified pEPI vector had also been developed in which the transcription is initiated from an inducible promoter that is switched on in the presence of doxycycline. This way, transgene expression is retained as long as doxycycline is provided in the culture medium. When doxycycline is no longer provided the transcription ceases, leading to the disappearance of the episomes during mitosis as the S/MAR element requires transcription read-through in order to promote episome retention. This system could be applied for transient gene expression such as the modification of stem cells (Rupprecht *et al.*, 2010).

Recently, an improved pEPI vector has been described, in which the tendency for transgene silencing has been further diminished by CpG depletion in the plasmid backbone (Haase *et al.*, 2010). The original pEPI plasmid contains 305 CpG islands, which are known to contribute to loss of gene expression by attracting methylation and therefore silencing, whereas the new pEPIto vector contains 60% less CpG motifs and is also smaller in size due to the omission of the second transcription unit coding for the antibiotic resistance gene. The CMV promoter is replaced with the human CMV enhancer/ human elongation factor α1 promoter, which has also been shown to be less prone to silencing (Hyde *et al.*, 2008). In preliminary experiments pEPIto has already exhibited higher transgene expression levels and more persistent expression than the parent vector, making it an interesting new alternative for the pEPI (Haase *et al.*, 2010).

1.5.4 Stress-induced duplex destabilisation

The DNA double helix is a flexible structure that undergoes conformational changes during the various physiological activities it partakes in, and separation of the two DNA strands is required for the binding of the RNA or DNA polymerase during transcription and replication. Superhelical stress imposed on the DNA duplex is a biologically important way of regulating local DNA duplex stability, and changes in superhelicity are regulated by an array of processes including DNA helicase, gyrase and topoisomerase activities, and constraints brought on by other DNA-binding events such as histone acetylation (Benham *et al.*, 2004; Klar *et al.*, 2005).

Stress-induced DNA duplex destabilisation (SIDD) can be achieved by imposing negative superhelicity. The resulting stress experienced by DNA sequences differs depending on the intrinsic thermodynamic stability of the particular sequence, such that in physiological conditions AT-rich regions tend to be less stable. However, duplex transitions in topologically constrained molecules are governed by much more complex factors than temperature-induced transitions in unconstrained molecules, and so whether a transition occurs at a given site depends not just on the local sequence properties but also on how the site competes with all the others in the domain (Benham *et al.*, 2004).

SIDD has been implicated in the mechanisms of activity of numerous biological processes, most notably the initiation of transcription and replication (Benham, 2001). Research has indicated that a site that is susceptible to superhelical strand separation is required in both prokaryotes and eukaryotes for the initiation of replication. In some instances a SIDD region is sufficient to initiate these processes, without the help of DNA binding proteins (Benham *et al.*, 2004; Bode *et al.*, 2006). SIDD sites have also been shown to play an important part in transcription termination (Benham, 1996). In addition to these regions, promoters and polyadenylation signals show a propensity for duplex destabilisation (Bode *et al.*, 1995).

Many of the functions attributed to destabilised DNA regions are shared with the functions of S/MARs, and indeed one of the hallmarks of S/MARs is their tendency for strand separation. S/MARs have been shown to exhibit a succession of evenly spaced sites of duplex destabilisation that would render part or all of the sequence single stranded at sufficient superhelicity (Benham *et al.*, 1997). The modular design of authentic S/MARs is demonstrated by the fact that an artificially constructed tetramer of

a 150 bp part of a natural S/MAR region exhibits S/MAR-like activity, and suggests that the underlying sequence elements interact with scaffold proteins at multiple sites that require duplex destabilisation to occur over a region of sufficient length (Benham *et al.*, 1997). The human genome is estimated to contain approximately ~ 30,000 – 70,000 S/MARs, but only 159 of these have been characterised and listed in the S/MAR database, S/MARt DB (Liebich *et al.*, 2002). Although SIDD cannot yet be used as a basis for predicting S/MAR regions in wild-type genomic DNA (Bode *et al.*, 2006), computational analysis of DNA sequences combined with *in vitro* analyses of affinity to nuclear matrix components is currently considered the best way to predict and look for S/MAR elements in the genome (Evans *et al.*, 2007).

The locations and extent of SIDD occurring in linear or circular double-stranded DNA molecules can be efficiently predicted with the web-based tool WebSIDD, a program freely available on the internet whereby the user inputs the DNA sequence and the program outputs the calculated transition probability and destabilisation energy of each base pair in the sequence (Bi *et al.*, 2004). The SIDD profiles are calculated using a previously developed statistical method that takes into account strand separation, superhelicity and twisting within denatured regions (Benham *et al.*, 1997).

Some groups have used WebSIDD to assist in the interpretation of their results concerning S/MAR -containing vectors, and the subsequent improvement of their vector systems (Broll *et al.*, 2010; Giannakopoulos *et al.*, 2009). Since the S/MAR element is strongly destabilised, the addition of such a sequence to a small topologically constrained episome may have a stabilising effect on other regions. This may then lead to the condensation of the promoter or the origin of replication, decreasing transgene expression or interfering with the capacity of the vector to replicate. When constructing S/MAR-containing episomal vectors, WebSIDD can be used as a tool to assess the positional effects of various components of the expression cassette, which enables the construct to be optimised *in silico* prior to subsequent *in vitro* and *in vivo* analysis.

1.6 Aims and objectives of this thesis

This thesis intends to explore methods for the establishment of mitotically stable episomes in mammalian cells. Whereas prokaryotic sequences seem to inhibit establishment of replicating episomes, selection pressure and prolonged exposure to the nucleoplasm appear to favour it. Hence the approaches chosen in this thesis for facilitating episome establishment were designed to follow these conditions; viral vectors were chosen as they do not include prokaryotic sequences, and the experimental protocols were devised to include a prolonged exposure of the vector to the nuclear environment.

Following the design and construction of new gene delivery methods, they need to be assessed for functionality. Usually, this is undertaken *in vitro* prior to moving to *in vivo,* first in model systems and then in clinical trials. Each step introduces more variables into the test system, as causes are harder to control and effects are harder to identify *in vivo* than *in vitro*. For this reason, it's worthwhile to test new systems *in vitro* before proceeding further.

1.6.1 Gene therapy for stem cells

Stem cells are defined as cells that retain an unlimited capacity for self-renewal, and can divide asymmetrically to produce more specialised cell types. Only embryonic stem cells are pluripotent, which means they can divide and become any other cell type found in the organism. Multipotent stem cells are also found in the adult organism and can form any cell type in a given tissue; examples include haematopoietic stem cells, which can divide into all cell types in the haematopoietic system, or neural stem cells, which can divide into neurons and glia.

The attraction for using stem cells in gene therapy lies within the possibility for *ex vivo* manipulation: the cells can be harvested from the patient, treated *in vitro*, and returned to the patient to populate the tissue system to be treated. This method for treatment overcomes both concerns regarding rejection of foreign tissue, as the cellular material originates from the patient himself or herself, and ethical problems regarding the source of the tissue material. Further advantages of using viral vectors for *ex vivo* transduction

instead of directly introducing the vector into the patient include an improved safety profile, as the likelihood of virus dissemination into non-targeted tissues is minimal. It also enables the assessment of expression levels prior to transplantation and the selection and expansion of highly expressing clones.

Another approach is to use *ex vivo* –modified stem cells for the *in vivo* production of a therapeutic proteins. Autologous or allogeneic transplantation of genetically modified stem cells has recently appeared as a credible form of therapy for many disorders usually treated by the administration of recombinant protein, such as anaemia and the haemophilias as well as cancer (Sanz *et al.*, 2011). Stem cells of choice for this type of therapy include adult human mesenchymal stem cells (MSCs), which are favoured due to availability, ease of transduction, high *in vitro* proliferation capacity and immunological characteristics (Aboody *et al.*, 2008).

Following transduction, expansion and selection of highly expressing clones, the cells can be administered either i) systemically as free cells, ii) they can be embedded into an artificial matrix scaffold or iii) they can be encapsidated into particles surrounded by a semipermeable membrane to protect them from the immune system (Sanz *et al.*, 2011). For MSCs, systemic administration does not result in a high number of introduced cells reaching most target organs. The exception is lungs, making the delivery method a viable option for treatment of lung metastases (Studeny *et al.*, 2004). Interestingly, an implant containing scaffold-embedded MSCs provided long-term expression of erythropoietin where systemic administration of cells resulted only in transient expression (Eliopoulos *et al.*, 2003). Another strategy involves grafting scaffold structures containing stem cells for bone repair. MSCs have been successfully used for bone tissue engineering, which has recently been gaining in popularity as traditional bone grafts are associated with donor site morbidity (Khaled *et al.*, 2011). An additional advantage provided by implant-based protein expression is reversibility, as gene expression can be discontinued by the removal of the implant. Encapsidation of donor cells is mainly used in allotransplantation to avoid an immune reaction towards the non-host cells, and has helped to generate some encouraging results in the treatment of tumours by direct injection (Goren *et al.*, 2010).

1.6.2 Approaches towards enhancing establishment of replicating episomes

From previously published work, it was inferred that an initial period of residence in the nucleus may be required or at least beneficial for the establishment of replicating episomes (Broll *et al.*, 2010; Piechaczek *et al.*, 1999). Although antibiotic selection can and has been used to achieve this effect, it precludes the transfer of the protocol to the clinic, and therefore an alternative method avoiding the use of antibiotics is preferable. Another way of prolonging the exposure of the vector genome to the nuclear environment is to slow down cell division. This would need to be executed in a way that does not cause DNA damage, which may increase the rate of integration, and the process needs to be reversible.

Many chemical agents used in cancer therapies to induce a mitotic arrest act by preventing the assembly of mitotic spindles (Gascoigne *et al.*, 2009). However, many of these drugs result in metabolic perturbations and toxicity, and as such are not suitable in our protocol (Davis *et al.*, 2001). Other drugs, such as aphidicolin and butyrolactone, as well as serum deprivation can also be used to induce a reversible cell cycle arrest, but concerns remain about the potential generation of double-strand breaks. Prolonged and severe serum deprivation (0.2% serum in media) has been shown to result in DNA fragmentation in fibroblasts (Kues *et al.*, 2000), and thus should be avoided.

Cell cycle block in tumour cells by methionine restriction was originally devised as a method for the treatment of cancer, as cancer cells were found to be particularly vulnerable to methionine deprivation (Guo *et al.*, 1993). Methionine dependence is a unique feature to cancer cells, and restricting access to it causes cell cycle arrest and apoptosis. Methionine restriction has been used as a method to induce a reversible arrest in the G_0/G_1 phase of the cell cycle in several different cell lines, including tumour cells (Jiang *et al.*, 2007).

Unlike plants, mammals cannot synthesise methionine and therefore it's classed as an essential amino acid. Normal human cells, however, are able to synthesise methionine from homocysteine. There is evidence to suggest that this pathway is blocked in cancer cells, which is why they differ in their reaction to methionine depletion if homocysteine is provided in the media (Guo *et al.*, 1993). As several cell types are used in the protocols in this thesis, homocysteine was not added to differentiate between normal

and cancer cells. Furthermore, methionine depletion should result in reduced protein synthesis in all cells, as methionine is the amino acid corresponding to the start codon of all proteins, and is therefore the first amino acid to be incorporated during the synthesis of every protein.

Methionine depletion has also been used to induce cell cycle arrest in muscle cell cultures to induce differentiation (Kitzmann *et al.*, 1998). A combination of low serum and methionine depletion was able to block cell proliferation of myoblasts without differentiation into myotubes, and the cells were able to re-enter proliferation upon reintroduction of methionine and higher serum concentration indicating reversibility of the cell cycle block (Kitzmann *et al.*, 1998).

1.6.3 Model cellular systems: Chinese hamster ovary (CHO) cells

The Chinese hamster ovary cell line was originally derived from an inbred female laboratory animal in 1957 (Tjio *et al.*, 1958). The cells descend from a spontaneously immortalised population of cultured fibroblast cells. It has been hypothesised that the cell line is of clonal origin, as the first cells and all subsequently derived populations are deficient in proline synthesis (Wurm *et al.*, 2011). Cells require proline due to the absence of the gene for proline synthesis, the block in the biosynthetic chain lies in the step converting glutamic acid to glutamine gamma serialdehyde (Valle *et al.*, 1973).

The chromosome number in wild type Chinese hamsters is very low at 22, which is why their cells were originally indicated as good candidates for cytogenetic research. However, in culture the cells were noted to have a labile karyotype right from the beginning, with variations in chromosome number as well as structural irregularities (Tjio *et al.*, 1958).

The cells were also found to proliferate rapidly and produce high amounts of protein, which is why they have since become popular for both *in vitro* research and engineered protein production. Currently, CHO cells are the cell line of choice for the production of many therapeutic biopharmaceuticals, as they are capable of performing human-compatible post-translational modifications as well as yielding several grams of protein per litre (Wurm, 2004). Another feature increasing the attractiveness of CHO cells for human protein production is their resistance to infection by many human viruses, including HSV and HIV-1, due to lack of viral entry receptors (Conner *et al.*, 2005).

Regulation of cell cycle in CHO cells has been done simply by allowing them to reach a stationary growth phase, which can be used to induce an arrest in the G_1 phase when prolonged for up to 80 hours (Tobey *et al.*, 1970). The cells resume an exponential growth phase following dilution. In a follow-up study, the cells could be induced to re-enter cycling by the introduction of isoleucine and glutamine to the medium, which were then deemed as essential for growth of the cells (Ley *et al.*, 1970).

The cell line used in this thesis, CHO-K1, is a subclone of the parental CHO cell line created by single cell cloning in 1957 (Puck *et al.*, 1958). The karyotype of this cell line is very different to the wild-type Chinese hamster, and only 8 of the 22 chromosomes in the cells are equivalent to wt chromosomes (Wurm *et al.*, 2011). The modal chromosome number of this cell line is 20, although again this varies between subclones (Kao *et al.*, 1970).

The CHO-K1 cells were selected for use in this thesis due to previously reported success in establishing S/MAR-based episomes using this cell line (Broll *et al.*, 2010; Piechaczek *et al.*, 1999). It was considered important to first reproduce the results achieved by other groups, before expanding the research into new areas.

1.6.4 Specific aims & objectives of current study

The work in this thesis aims to explore the potential for generating mitotically stable episomes for safe, long-term gene therapy. To this end, 4 different potentially influencing factors are investigated: i) S/MAR elements, ii) transient cell cycle block using methionine depletion, iii) vector type, and iv) cell type.

The effect of S/MAR elements on episome retention is explored in chapters 3-6. There are two different S/MAR elements used throughout the thesis, the full-length β-IFN S/MAR and its truncated daughter element miniMAR. These elements are included in different vector systems and their ability to establish stable episomes is assessed.

The effect of inducing a transient mitotic block after introducing episomal elements into the cell is also analysed in each Results chapter. It was postulated that interactions between the episomes and the nuclear matrix need to occur, and therefore the episome needs to remain in the nuclear environment for an extended period of time prior to being dispersed by cell division. To achieve this, the cells were induced to undergo a period of enforced cell cycle arrest following introduction of the vectors to the cells.

Chapters 3-5 use different vectors on the same CHO-K1 cell line, and therefore the cell cycle block is very similar in each chapter. Conversely, Chapter 6 explores the possibility for using methionine depletion as a method for achieving cell cycle block in other cell types.

Three different classes of vectors are analysed in this thesis: AAV vectors, IDLVs, and non-viral plasmid and minicircle vectors. Non-viral vectors, including pEPI plasmids, viral transfer plasmids and a minicircle, are analysed in Chapter 3. Chapter 4 explores the episomal retention of both single-stranded and self-complementary AAV vectors. Chapter 5 investigates episomes created by non-integrating lentiviral vectors, as well as the integration status of the assumed episomes.

Finally, in Chapter 6, possibilities for applying the method for stable episome establishment to more clinically applicable cell types are explored. Non-integrating lentiviral vectors are used to develop transduction protocols using human cell lines and murine haematopoietic stem cells.

Chapter 2: Materials & Methods

2.1 General laboratory reagents

All general laboratory reagents were AnalaR grade and supplied by BDH, VWR, Invitrogen or Sigma. Details of more common reagents are listed below, those used only in specialised protocols are listed in the materials of the relevant section. Unless otherwise stated, all solutions were made up using double distilled water (ddH_2O). Solutions used for tissue culture applications or solutions containing protein, detergent or glucose were sterilised by passing through a 0.22 µm filter (Falcon), all other solutions were sterilised by autoclaving at 121^{o}C for 15 min. All solutions were stores at room temperature unless stated otherwise.

Acetic acid (CH_3COOH)	VWR
Agarose	Invitrogen
Ammonium acetate $AcO(NH_4)_2$	BDH
Bovine Serum Albumin (BSA)	Sigma
Boric acid	Sigma
$CaCl_2$	Sigma
Citric acid	Sigma
Dimethyl sulphoxide (DMSO)	Sigma
EDTA	Sigma
Ethanol (EtOH)	VWR
Glucose	Sigma
Hydrochloric acid (HCl)	BDH
Liquid broth (LB)	Sigma
Methanol	VWR
Magnesium chloride ($MgCl_2$)	Sigma
Magnesium sulphate ($MgSO_4$)	BDH
NaOH	Sigma
Na_2HPO_4	Sigma
NP40	Fluka Biochemica

Paraformaldehyde (PFA)	Sigma
Potassium chloride (KCl)	Sigma
Phosphate buffered saline (PBS) pH 7.3	Oxoid Ltd
Sodium chloride (NaCl)	Sigma
Sodium dodecylsulphate (SDS)	Sigma
$NaHCO_3$	Sigma
Sucrose	Sigma
Tri-sodium citrate	Sigma
Trizma Base	Sigma
Trizma hydrochloride (Tris-HCl)	Sigma

2.2 Culture and storage of bacteria

2.2.1 Materials

Methicillin / Ampicillin stock. 1000 x stock prepared as 80 mg/ml of methicillin (Sigma) and 20 mg/ml of ampicillin (Sigma) in water and stored at -20 ^{o}C. Diluted 1000 x for use in LB media and LB agar plates.

Agar. Fisher Scientific.

Glycerol. Sigma.

Lysogeny Broth (LB). Sigma.

LB agar. 1.5 % of agar was added to LB, autoclaved o dissolve agar, and cooled to 45 ^{o}C prior to the addition of antibiotic from stock. 20-30 ml of cooled agar was poured into 100 mm petri dishes close to the Bunsen burner and stored sealed at 4 ^{o}C once set.

SOC medium. 100 µl of 1 M $MgSO_4$ and 20 µl 1 M glucose were added to 10 ml LB medium.

TOP10 competent *E.coli* cells. Invitrogen.

2.2.2 Storage of bacterial stocks

Glycerol stocks were made by combining 300 µl 80 % glycerol with 700 µl bacterial overnight culture in a cryovial and freezing at -80 ^{o}C.

2.3 Plasmid DNA purification methods

2.3.1 Materials

2.3.1.1 Reagents

Plasmid DNA purification kits. Mini, Midi, Maxi, Mega & Giga, Qiagen.

Qiagen kits. Qiaprep spin miniprep kit, and endo-free maxi-, mega- and gigaprep kits.

SOC Medium. 100 µl of 1 M $MgSO_4$ and 20 µl 1 M glucose were added to 10 ml LB medium.

TE. 10 mM Tris pH 8.0, 1 mM EDTA pH 8.0

2.3.1.2. Equipment

Electroporator. Micropulser by Bio-Rad.

Electroporation cuvettes. Invitrogen.

2.3.1.3 Plasmids

pEPI GFP S/MAR.

The pEPI plasmid vector is based on the commercial pGFP-C1 vector (Clontech), with the addition of the S/MAR element from the 5' region of the human β-interferon gene placed downstream of the GFP gene driven by the cytomegalovirus (CMV) immediate early promoter (Piechaczek *et al.,* 1999). The plasmid was kindly donated by Dr Juergen Bode, Helmholz institute, Braunschweig, Germany.

pEPI GFP miniMAR733

The pEPI plasmid with a modified truncated version of the β-IFN S/MAR element (733 bp long) was a naturally occurring mutant discovered in Juergen Bode's lab (Broll *et al.,* 2010). The plasmid was kindly donated by Dr Bode.

pscAAV CMV GFP

The pscAAV CMV GFP plasmid was kindly donated by Dr Amit Nathwani, UCL, London.

pscAAV CMV GFP mMAR

See section 2.5.6.

pscAAV CAG GFP

The pscAAV CMV GFP was engineered by GeneArt to remove the CMV promoter and replace it with a hybrid CMV enhancer / chicken β-actin promoter. The resulting plasmid differs from the parent plasmid by ~300 bp. The sequence was confirmed by sequencing the new section of the plasmid (MWG). Two point mutations were found when comparing the sequencing output and the electronic plasmid map, and as these were both in non-coding regions, and were assumed to be either artefacts of sequencing or, in any case, inconsequential to the function of the vector.

pRRLsc SV40 GFP

Plasmid cloned by Cecilia Goode and kindly donated by Dr Rafael Yáñez-Muñoz.

pRRLsc SV40 GFP miniMAR

Plasmid cloned by Cecilia Goode and kindly donated by Dr Rafael Yáñez-Muñoz.

pRRLsc CMV GFP

Kindly donated by Dr Rafael Yáñez-Muñoz manufactured by PlasmidFactory.

2.3.2 Plasmid DNA transformation of bacteria by electroporation

10 ng of plasmid DNA in a volume of 1 µl of H_20 was added to 25 µl of thawed electrocompetent cells. The cells were electroporated and 300 µl of SOC media was added immediately afterwards. The culture was incubated for 1 h at 37 ^oC and either transferred to 10 ml LB and incubated overnight, or plated out on agar plates.

2.3.3 Qiagen plasmid preparations

All Qiagen plasmid preparations were carried out according to manufacturer's instructions.

2.4 Minicircle

The minicircle and control plasmid pCMV-GFP were manufactured by PlasmidFactoryGmbH (Germany), and contained in their McBox product (PFBox102). Both constructs contain the GFP gene driven by the CMV immediate early promoter and followed by the SV40 polyA signal, and the minicircle is devoid of almost all bacterial sequences (Figure 3.2). The CpG content is also reduced. Further information concerning the exact sequence, or the recombinase enzyme and corresponding recognition site used in the manufacture of these minicircles, is not released by the company.

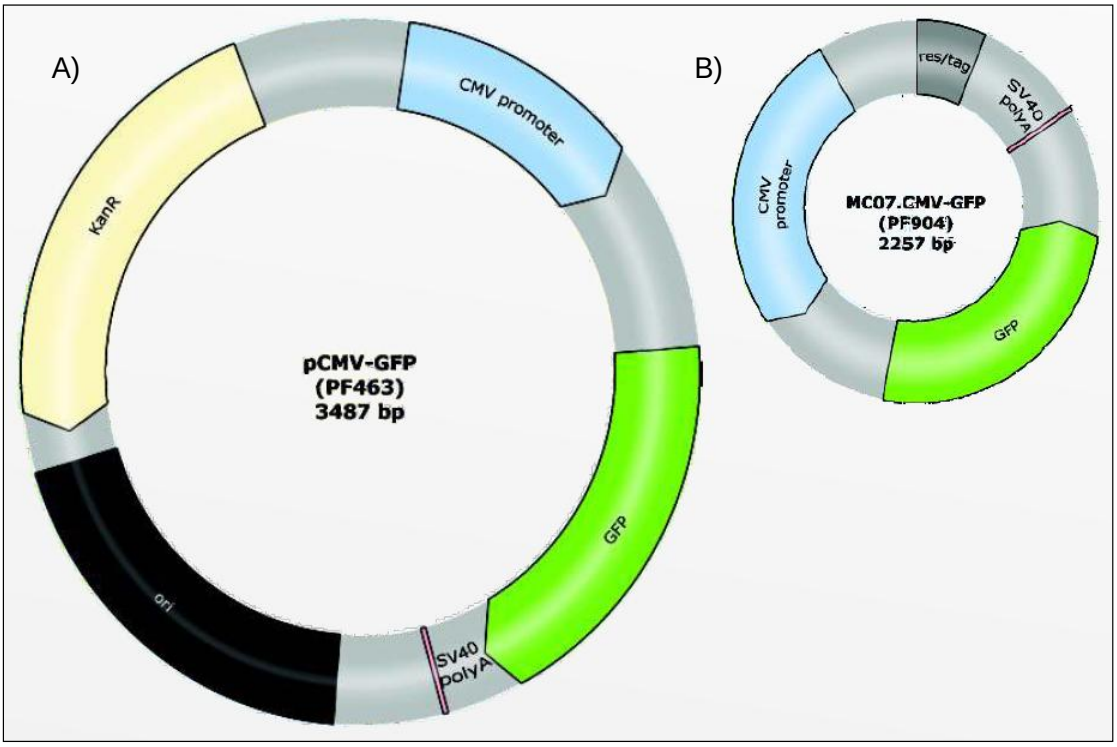


Figure 2.1 Diagrammatic representation of the minicircle and miniplasmid provided by the Plasmid Factory.

Both the minicircle and the pCMV-GFP from Plasmid Factory contain the GFP gene driven by the CMV promoter and followed by the SV40 polyA signal. The minicircle (B) is 1230 bp smaller and contains almost no bacterial sequences; the control plasmid (A) contains a kanamycin resistance gene and a bacterial origin of replication.

2.5 Molecular cloning techniques

2.5.1 Materials

1x TBE buffer. 89 mM Trizma base, 89 mM boric acid, 2 mM EDTA pH 8.0.

5X Loading buffer. Bioline.

DNA molecular weight markers. Bioline. Hyperladders I, IV and V.

DNA polymerase I large fragment (Klenow). New England Biolabs.

Qiaquick Gel extraction kit. Qiagen.

Qiaquick PCR purification kit. Qiagen

Quick-Stick Ligase. Bioline.

Restriction enzymes and buffers. New England Biolabs and Promega.

Shrimp Alkaline Phosphatase. Roche.

SYBR® Safe DNA gel stain. 10 000x concentrate in DMSO. Invitrogen.

UltraPure Agarose. Invitrogen.

2.5.2 Restriction enzyme digestion and gel electrophoresis

To verify the identity of plasmid preparations, a sample (~ 1 µg) was digested with a variety of restriction enzymes containing both single and multiple cutters using the manufacturer's recommended buffers and conditions. The samples were loaded onto an agarose gel using 5x loading buffer and appropriate molecular weight markers. The gel was made using deionised water and agarose at 0.5 – 2 % depending on the expected size of the bands. The samples were then electrophoresed at 60-100V using 1x TBE buffer, with SYBR® Safe DNA stain included in the gel at a final concentration of 0.01% to enable the visualisation of the DNA bands with UV light. The band patterns were compared with expected band sizes to ascertain the correct identity of the plasmid prep.

2.5.3 Klenow blunt ending

If blunt end cloning was required, the termini of restriction digested pDNA fragments were blunt ended by filling out overhangs utilising the large fragment of DNA polymerase I (Klenow fragment). 50-100 µl reaction volumes containing 10-20 µg of restriction digested DNA were prepared with NEBuffer 2 and 33 µM dNTPs and incubated according to manufacturer's recommendations.

2.5.4 SAP dephosphorylation

Shrimp alkaline phosphatase (SAP) was used for the dephosphorylation of 5' phosphates from DNA fragments during molecular cloning. Dephosphorylation was performed after restriction enzyme digestion. First, the restriction enzyme was inactivated at 65^{o}C for 15 min. The dephosphorylation reaction was then set up using 0.9 µl SAP per 50 ng DNA, following manufacturer's recommendations. The reactions were incubated for 1 h at 37 ^{o}C and SAP was inactivated by incubating at 65^{o}C for 15 min.

2.5.5 DNA ligation

DNA ligations were performed utilising the Quick-Stick ligase according to manufacturer's instructions. The reactions were allowed to proceed for 15 min.

2.5.6 Cloning of pscAAV CMV GFP mMAR

The pscAAV CMV GFP plasmid backbone (see section 2.3.1.3) was linearised with SbfI. The pEPI GFP miniMAR733 plasmid (as above) was linearised with KpnI. Both linearised fragments were run on 1% agarose gel and gel extracted. The digested ends were then blunted by using T4 DNA polymerase. Both fragments were then digested with BsrGI, releasing the ~733 bp miniMAR fragment from the insert plasmid and creating a compatible end on the vector plasmid. The digested DNA was run on gel and the appropriate fragments were gel extracted.

The vector and insert fragments were ligated in 3 different proportions (1:1, 1:3 and 1:6 vector to insert) using Quick T4 DNA Ligase. The ligation mixture was purified using

Qiagen PCR Purification Kit and used to transform TOP10 competent E.coli cells by electroporation. Of the resulting colonies, 12 were picked for overnight cultures, and plasmid DNA was purified by Qiagen Miniprep Kit. The purified plasmid DNA was then digested with BglII to determine the presence of the insert fragment. Two of the colonies were deemed to have the correct digest pattern and were made into glycerol stocks and used for a Megaprep kit. The mMAR region of the resulting plasmid was also sequenced by MWG and only one nucleotide substitution or inaccuracy in our original sequence was found in a non-coding region.

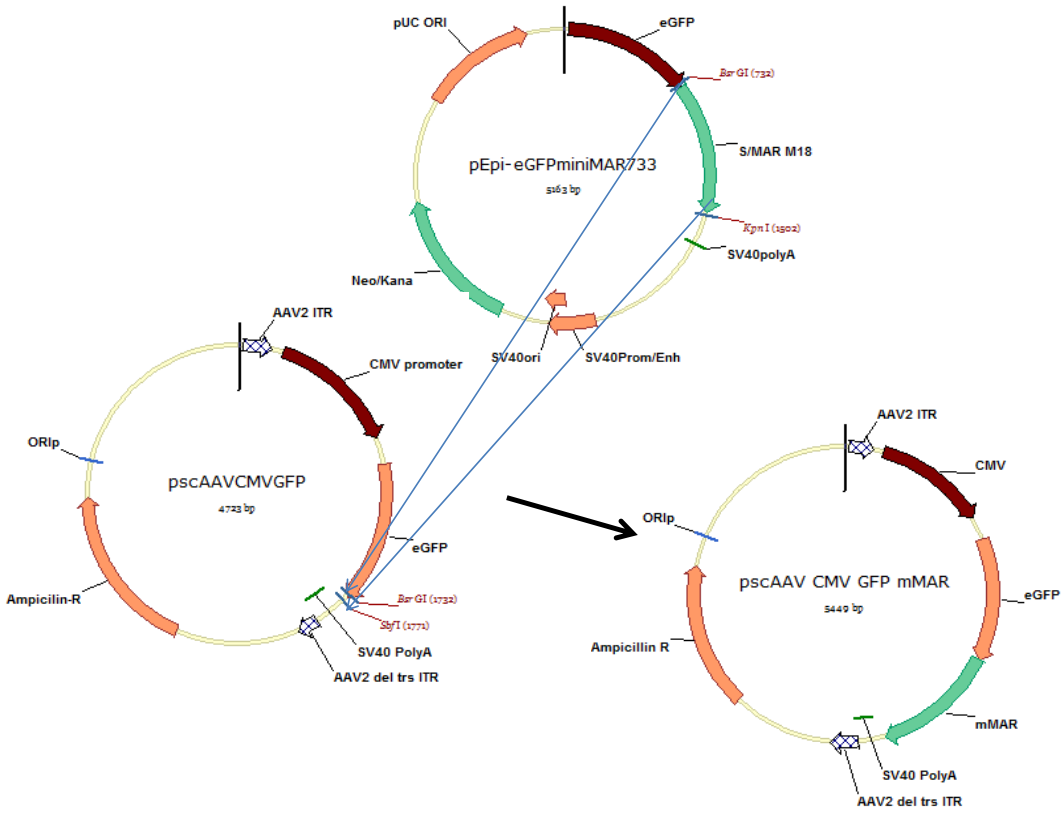


Figure 2.2 A diagram showing the cloning of the miniMAR element into the self-complementary AAV plasmid.

The miniMAR element was excised from a pEPI plasmid and inserted into the pscAAV-CMV-GFP plasmid downstream of the GFP gene.

2.6 Plasmid DNA transfection

2.6.1 Materials

Lipofectamine™ 2000. Invitrogen.

2.6.2 Plasmid transfection

$2x10^5$ CHO cells per well were seeded on 6-well plates. After 24 hours, the cells were transfected using 2 µg DNA/well and LipofectamineTM 2000 (Invitrogen) with a Lipofectamine to DNA ratio 1:2. The transfection media was removed and replaced with normal media after 6 hours. 24 hours post-transfection, the cells from one well transfected with each plasmid and a mock transfected control were harvested and analysed for GFP expression by flow cytometry and stained with PI for cell cycle analysis.

For the cells destined to undergo cell cycle arrest, the media in the remaining wells was changed for DMEM /-Met to induce cell cycle arrest (see section 2.2.7.3.2). The media was replaced with fresh each day for the next 3 days. At the end of the quiescent period, the cells from one well for each plasmid and a control were used for GFP expression analysis and cell cycle analysis by flow cytometry. The remaining cells were split 1:5 and seeded in full DMEM.

Following a 5-day recovery period, cells were analysed for GFP expression and cell cycle phase by PI to confirm return to their normally propagating state. After that cells were kept in continuous proliferating culture and split 3 times a week, and GFP expression was measured once a week by flow cytometry.

The cells in continuous culture were analysed for GFP expression and cell cycle phase 24 hours and 5 days after transfection and at weekly intervals thereafter. The culture was discontinued when the GFP expression was deemed to have reached a stable level.

The cells transfected with the pEPI plasmids and not undergoing cell cycle arrest were frozen down in liquid N_2 after 10 days in culture. After thawing, culture was continued as before, splitting cells 3 times a week and measuring GFP expression by flow cytometry once a week.

2.6.3 Minicircle transfection

Transfection of the minicircle and control plasmid DNA was performed using Lipofectamine^{TM} as detailed in section 2.6.2 for plasmid vectors. The culture conditions for an initial period of cell cycle arrest and a control population in proliferating culture were also done as in section 2.6.2, with the modification that the cells were held in cell cycle arrest for 4 days instead of 3 to ensure efficient cell cycle block.

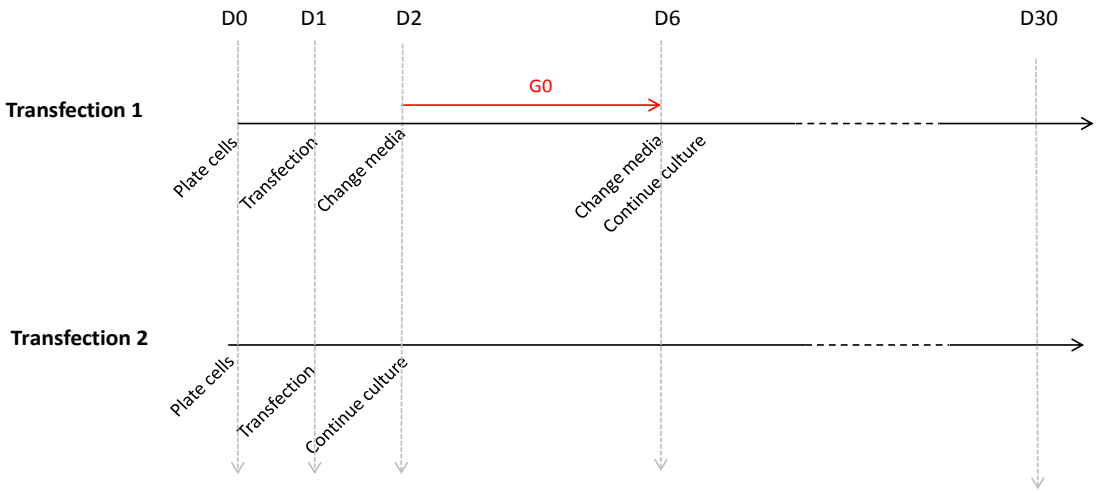


Figure 2.3 Diagrammatic representation of transfection protocol including and omitting a quiescent period.

Transfection 1 includes a transient cell cycle arrest for 4 or 5 days depending on the experiment. Transfection 2 does not include a period of cell cycle arrest and serves as a control population.

2.7 Tissue culture techniques

2.7.1 Materials

Most cell lines used were cultured in a basic medium of DMEM with 10% Fetal Calf Serum and 1 % antibiotics. This basic culture medium will be referred to as full DMEM in later sections.

CellTiter Proliferation assay. Promega.

Dulbecco's Modified Eagle's Medium (DMEM) with 4.5 g/l glucose. Paa.

Foetal Calf Serum. Paa.

L-glutamine. Sigma.

L-methionine. Sigma.

L-proline. Sigma.

Penicillin/streptomycin. 100 Units/ml. Sigma.

RPMI-1640. Sigma Aldrich.

Trypan blue. Gibco.

Trypsin-EDTA (10x). 0.5 % trypsin, 0.2% EDTA in PBS. Sigma.

2.7.2 Cell lines

CHO K1

The Chinese Hamster Ovary cell line CHO-K1 was kindly donated by Juergen Bode.

HeLa

The HeLa cells used in transduction experiments and lentivector titration are a standard cell line used in the lab. Initial passage number used varied between 20 and 24.

293T

293T cells were used in vector manufacture and were confirmed to be mycoplasma free. Initial passage numbers varied between 18 and 31.

2.7.3 Cell viability assessment

Cell viability was assessed by Trypan blue staining, which is based on the assumption that dead cells have a damaged cell membrane and so cannot exclude the dye. Viable cells maintain a clear cytoplasm when mixed with Trypan blue whereas dead or damaged cells are stained blue. 10 µl cell suspension was mixed with 10 µl Trypan blue (0.4%) and left for 1 minute before counting live and dead cells using a haemocytometer on a light microscope.

2.7.4 Culture of CHO- K1 cells

The CHO-K1 cell line has been demonstrated to require proline in the media (Kao *et al.*, 1967), and therefore all media not already containing L-proline was supplemented with 0.02g/l L-proline for use with these cells.

The standard media used for CHO cell culture was DMEM supplemented with 10% FBS, 0.02 g/l L-proline, 100 units/ml penicillin, and 100 g/ml streptomycin.
The methionine-free media for inducing cell cycle arrest was methionine-free DMEM with 2% FBS, 0.02 g/l L-proline, 100 units/ml penicillin, and 100 g/ml streptomycin. It has been estimated that L-methionine concentration in bovine serum is around 15 M, so the final concentration of L-methionine in our experiments would be around 0.75 M (Martinez-Chantar *et al.*, 2003).

2.7.4.1 Growth assay

The experimental plan includes a period in which the cells are held in reversible cell cycle arrest at G_1/G_0, and a potential method for achieving this in CHO cells is culturing the cells in methionine-free, serum-depleted medium (O'Neill *et al.*, 1978). The standard culture medium for CHO cells is Ham's-F12, however, methionine-depleted Ham's-F12 is not available and so methods for culturing CHO cells normally and for arresting the cell cycle using DMEM-based media needed to be optimised.

The ability of CHO cells to survive in DMEM instead of complex media was tested by plating cells from the same CHO line and culturing them in the 2 different media: i) RPMI-1640, supplemented with 10% FBS, 1 mg/ml L-glutamine and 1% PenStrep and ii) DMEM, supplemented with 10% FBS, 0.02g/l L-proline, 1 mg/ml L-glutamine and PenStrep.

CHO cells were seeded at $1x10^4$ cells per well on 6-well plates in 2ml media /well, 2 plates /medium. 2 wells per medium were trypsinised and cells counted on days 2,3,4,5 and 8 after plating. The DMEM supplemented with proline was initially found to support faster cell growth, but after a week of culture the cells in RPMI had grown more. However, the cell morphology and growth rate in the supplemented DMEM was judged to be sufficiently healthy to continue culture in this media (Figure 2.2)

2.7.4.2 Optimisation of CHO cell reversible cell cycle arrest

To assess the capacity of different media to hold cells in G_0, CHO cells were cultured in 4 different media: DMEM with above supplements and 2% FBS without methionine, 10% FBS without methionine, 2% FBS with methionine and as a normal control 10% FBS with methionine. Cells were plated onto 96-well plates in 3 different densities, 15,000 cells/well, 7,500 cells/well and 3,000 cells/well, 5 replicates of each density per media. The effect of different media and different seeding densities on cell proliferation was measured colorimetrically using the Promega CellTiter Proliferation Assay Kit.

At 2 and 4 days after seeding the cells according to each different condition, 20 µl of the MTS tetrazolium compound was added directly onto the cells and incubated for 1.5 hours at 37^oC, during which time it was bioreduced by cells into a coloured formazan product, reflecting the rate of metabolism in each well (Bernabei *et al.*, 1989). The absorbance was recorded at 490nm with a 96-well plate reader.

The cells seeded at the medium density survived best and also exhibited morphologically the most normal phenotype, although only the difference between the lowest and highest cell density was statistically significant (One-way ANOVA followed by Tukey's post-test, $P < 0.05$).

Cells cultured in DMEM with methionine became confluent by day 2 and subsequently cell death was observed due to overconfluence. Statistically, both methionine-free conditions were significantly different from methionine-containing conditions (One-way ANOVA followed by Tukey's post-test, $P < 0.05$).

Although the two methionine-free conditions were not statistically significantly different, all the cell densities grew to become overconfluent in 3 days in 10% FCS, and so the medium cell density of 2 x 10^4 cells/cm^2 and DMEM with 2% FBS without methionine was chosen for future experiments. This media is hereafter referred to as DMEM -Met.

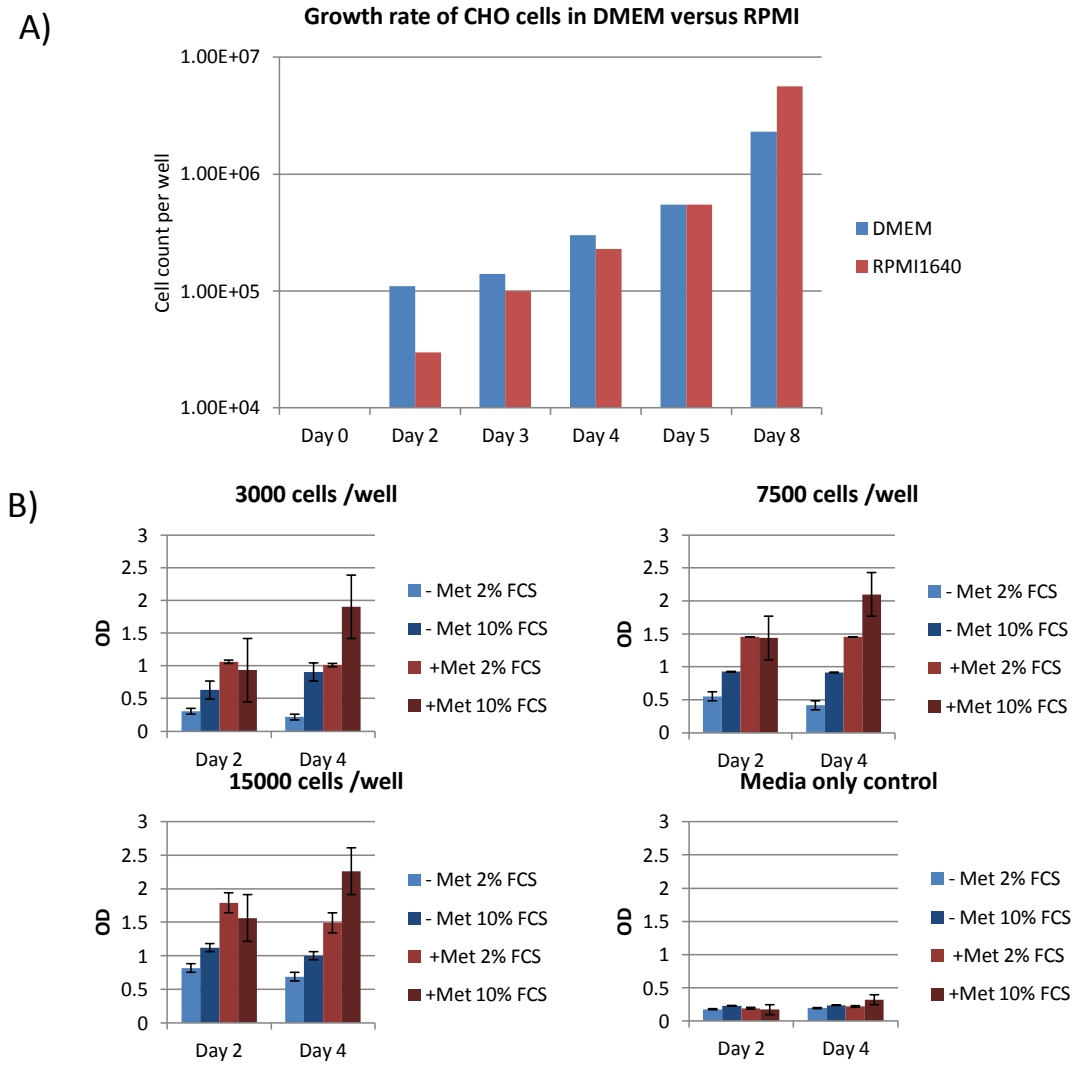


Figure 2.4 Optimisation of CHO cell culture and induction of cell cycle arrest.

The growth rate of CHO cells is similar in supplemented DMEM and RPMI. The values represent the average number of cells counted from 2 plates per condition, Day 0 refers to time of seeding. (B) Cells seeded at 7,500 cells/well on a 96-well plate and cultured in 2% FCS methionine-free DMEM exhibit the least proliferation whilst retaining a sound morphology. Y axis, optical density corresponding to metabolic activity of cells. Data shown is mean +/- SEM, n=6 per condition.

2.7.5 Proliferation assay of CHO and HeLa cells

The effect of different medias on the proliferation of different cell lines was investigated using CellTiter Proliferation assay, according to manufacturer's instructions. Briefly, CHO cells and HeLa cells were seeded at 2 different densities on 96-well plates, corresponding to $1x10^5$ and $2x10^5$ cells /well on a 6-well plate. The cells were seeded in 100 µl methionine-containing 10 % FCS DMEM, and 24 hours afterwards the media was changed to 100 µl methionine-free 2% FCS. The cells were kept in this media for 3, 4 and 5 days, and the media was changed daily. The rate of metabolism in each culture condition was then measured using the proliferation assay kit followed by recording the absorbance at 490nm (see section 2.7.4.2).

2.8 Murine Haematopoietic Stem Cells

2.8.1 Materials

50:50 DMEM/Ham's F12 (powder). Sigma, cat no D9785.

Depletion buffer. PBS pH 7.2, 0.5% BSA, 2 mM EDTA, sterile filtered.

Human Flt3-ligand. Miltenyi Biotech, cat no 130-093-853.

Lineage Cell Depletion kit. Miltenyi Biotec, cat no 130-090-858.

L-methionine. Sigma.

MACS LS Column. Miltenyi Biotec.

Recombinant human IL-6. PeproTech, cat no 200-6.

Recombinant Mouse SCF/c-kit ligand. R&D Systems, Cat no 455-MC.

Trypan blue (0.4%). Gibco.

2.8.2 Harvest & lineage depletion of murine HSCs

Murine bone marrow was collected from the femurs and tibias of male and female Balb/c mice by flushing the shaft with depletion buffer using a syringe and a 26G needle. Cells were kept on ice during all the following steps. Aggregates were removed

by pipetting up and down 10 times with a syringe and a 26G needle, and passing the suspension through at 40µm nylon mesh. Cells were washed by adding 10 ml buffer and centrifuging at 300 *g* at 4^{o}C for 10 min. Supernatant was removed, cells were resuspended in 5 ml buffer and counted using a haemocytometer and Trypan blue. Centrifugation was repeated and cells resuspended at 40 µl of buffer per 10^7 total cells.

Magnetic labelling of cells expressing lineage markers was done using the Lineage Cell Depletion Kit. First, 10 µl of Biotin-Antibody Cocktail was added per 10^7 total cells and the mixture was incubated for 20 minutes on ice. 30 µl of buffer and 20 µl of Anti-Biotin MicroBeads was then added per 10^7 total cells, followed by incubating for 30 min on ice. Cells were then washed by adding 10 ml buffer and centrifuging at 300 *g* at 4^{o}C for 10 min. Supernatant was removed and cells were resuspended at up to10^8 cells in 500 µl buffer.

Magnetic separation was performed using a MACS LS column. The cell suspension was passed through the column under a magnetic field, and the effluent fraction containing unlabelled lineage negative cells was collected. The cells were counted and seeded on non-tissue culture treated 24-well plates at 6 x10^4 -10^5 cells per well.

2.8.3 *In vitro* maintenance of murine haematopoietic stem cells

The culture medium for murine HSCs was made using the 50:50 DMEM/Ham's F12 powder. Whilst stirring at RT, 14.8 g of powdered medium was dissolved in 450 ml tissue culture grade water, 0.6 g $NaHCO_3$, was added, and stirred until dissolved. pH was adjusted to 7.1 using 1 N HCL or 1 N NaOH (to achieve desired pH of 7.2 after filtration). The following amino acids were then added: 0.07725 g $CaCl_2$, 0.029525 g L-Leucine, 0.045625 g L-Lysine, 0.0306 g $MgCl_2$ and 0.02442 g $MgSO_4$. For the methionine-containing media, 0.344 g L-Methionine was dissolved in 10 ml tissue culture water to make stock, filter sterilised, and 500 µl added per 500 ml media.

5 ml PenStrep (100 Units/ml, Sigma) was added for each 500 ml aliquot, and the media made up to 500 ml using tissue culture grade water. The medias were filtered using a Vacupump, and frozen at -20^{o}C in 50 ml aliquots.

Immediately prior to use, a cytokine cocktail was added to the media in order to increase transduction efficiency and promote cell survival (Luskey *et al.*, 1992). The cytokines used were 100 ng/ml Murine Stem Cell Factor, 10 ng/ml Human recombinant Flt-3 and 20 ng/ml murine IL-6, as per protocol provided by S.J. Howe, Institute of Child Health, University College London.

2.9 Fluorescence Assisted Cell Sorting (FACS) Analysis

2.9.1 Materials

2.9.1.1 Solutions

Paraformaldehyde (PFA). 4 % Solution. Sigma.

2.9.1.2 Equipment

FACSCantoII. BD Biosciences.

FACSDiva Software. BS Biosciences.

5 ml snap-cap round bottom tubes. BD Falcon.

2.9.2 Analysis of GFP expression by flow cytometry

Flow cytometry was used as the standard method for assessing the proportion of cells expressing GFP throughout this thesis. To perform FACS analysis, cells were trypsinised and pelleted by centrifuging at 1000 rpm for 5 min in 5 ml snap-cap tubes. The supernatant was removed and the cells were resuspended in 300 µl 4 % PFA. A mock transduced control population was also included in every experiment measuring GFP expression.

The FSC/SSC plot was used to separate the debris from the live cell population by gating. The live cell population was then displayed in a histogram plot using the FITC channel to analyse GFP expression. The mock population was always analysed first to set the gate to be 0.5 % GFP positive. At least 10 000 cells were analysed for each condition or cell population.

2.10 Cell cycle analysis

2.10.1 Materials

Propidium iodide. 1 mg/ml in water, Sigma.

RNase. Sigma.

2.10.1.1 Solutions

Solution I: 10 mM NaCl, 1 mg/ml tri-sodium citrate, 0.06 % NP40, and freshly added 25 µg/ml propidium iodide and 10 µg/ml RNase.

Solution II: 15 mg/ml citric acid, 85.6 mg/ml sucrose and freshly added 40 µg/ml propidium iodide.

2.10.1.2 Equipment

FACSCantoII flow cytometer.

FlowJo Software. TreeStar Inc.

2.10.2 Propidium Iodide Stain

CHO-K1 and HeLa cells were prepared for cell cycle analysis by staining with PI, as follows: approximately $1x10^6$ cells were trypsinised and spun down at 800 rpm for 4 min. The supernatant was discarded and the cells resuspended in the remaining drops of medium by gently flicking the tube. 1 ml of Solution I was added into each tube while shaking and cells incubated at room temperature in the dark for 30 min. 1 ml of Solution II was then added into each tube whilst shaking and incubated at 4 ^oC for 1-2 h. The stained nuclei were analysed with FACSCantoII flow cytometer to determine the amount of DNA in each nucleus.

2.10.3 Cell cycle phase analysis

The flow cytometry data from the PI stained nuclei was analysed to ascertain that the cells designed to enter cell cycle arrest were indeed being held back from normal cycling whilst still remaining viable. The data generated by the FACSCanto flow cytometer was processed using FlowJo cell cycle analysis software to determine the proportion of cells in each cell cycle phase during the experiments. After gating the live cell population, the "cell cycle analysis" function of the software was employed to designate G_1 and G_2 peaks and assess the percentage of cells in G1, S and G2 phases. The remaining nuclei which contain an irregular amount of DNA and do not fit within the curves of any of the 3 classed cell cycle phases were classified as "non-categorised". They may be artefacts of the protocol preparing the nuclei, extracellular debris, or potentially apoptotic cells; it was not possible to distinguish between these with the methods to hand.

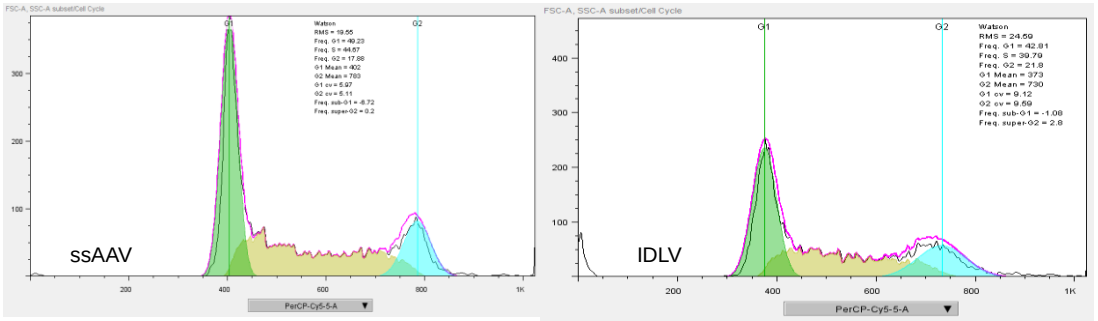
The success of the cell cycle block was assesses separately for each experiment. The percentage of cells in each cell cycle phase at the end of the quiescent period was compared to the equivalent percentages in the same cells prior to and after the cell cycle arrest using One-Way ANOVA.

2.10.4 Exclusion of non-viable mHSCs from FACScan using PI

When using flow cytometry to analyse GFP expression in murine haematopoietic stem cells, the non-viable population was excluded from analysis using propidium iodide (PI). As HSCs are non-adherent, it is not possible to remove dead cells or debris simply by washing prior to trypsinisation, and therefore more care needs to be taken to exclude necrotic or apoptotic populations from GFP analysis to avoid counting autofluorescence as transgene expression. PI enters non-viable cells through the compromised membrane and enables easy exclusion of the population.

The PI stain was performed by adding PI to a final concentration of 5 µg/ml and then excluding the positive population from further flow cytometric analysis.

A) Proliferating CHO cells



B) CHO cells in cell cycle arrest

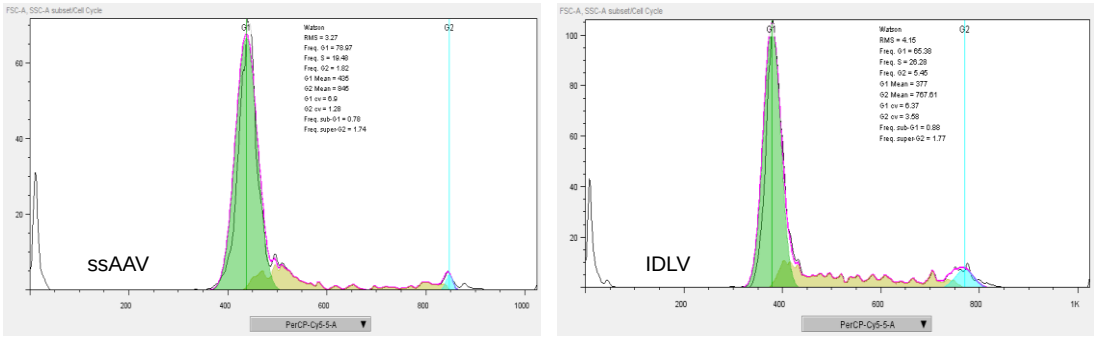


Figure 2.5 Typical FlowJo graphs for CHO cells in cycling and in induced cell cycle arrest.

A) Shows the cell cycle graphs of proliferating CHO cells in 2 independent experiments, where distinct populations in each of the phases of the cell cycle can be seen. B) Shows typical samples of cells in induced cell cycle arrest from 2 independent experiments, where a marked reduction in the S- phase and G_2/M areas of the graph can be seen.

2.11 DNA extraction methods

2.11.1 Isolation of High Molecular Weight DNA from Mammalian Cells

This protocol is designed to minimise the shearing of the genomic DNA and produce genomic DNA fragments larger than 100 kb. High molecular weight DNA from transgene - containing transduced cells can be used for episomal status analysis as breakage of genomic DNA is minimal and the episomes are the smallest DNA molecules present.

2.11.1.1 Materials

Ammonium acetate ($AcO(NH_4)_2$). Sigma.
Extraction buffer. 10 mM TrisCl (pH8.0), 0.1 M EDTA (pH8.0), 0.5% SDS, 20 µg/ml RNAse.
RNase. Sigma.
Proteinase K. Sigma.
TBS. 137 mM $NaCl_2$, 2.7 mM KCl, 25 mM Tris HCl pH 7.4, filter sterilised.

2.11.1.2 Protocol

Approximately $5x10^7$ – $1x10^8$ cells were washed twice with PBS, and the cells were detached using trypsin, which was then neutralised with media. The cells were centrifuged at 1000 rpm for 5 minutes, the supernatant was removed, and the cells were resuspended in 5 ml TBS. The cells were centrifuged at 1500 rpm for 10 mins at 4^{o}C, the supernatant was removed, and the wash repeated once more. The cells were resuspended in TE (pH 8.0) at a concentration of $5x10^7$ cells/ml. The solution was transferred to a conical flask, 10 ml of extraction buffer was added for each ml of cell suspension, and incubated for 1 h at 37^{o}C. Proteinase K was then added to a final concentration of 100 µg/ml, and the suspension of lysed cells incubated for 3 hours at 50^{o}C. The solution was cooled and poured into a centrifuge tube, an equal volume of phenol equilibrated with 0.5 M TrisCl (pH8.0) was added and the two phases gently mixed by turning the tube end over end for 10 minutes. The two phases were separated by centrifugation at 5000 g for 15 mins at RT. The viscous aqueous phase was transferred to a clean centrifuge tube and the extraction with phenol was repeated twice. After the third extraction with phenol, the aqueous phases were transferred to a fresh centrifuge tube and 0.2 volumes of 10 M ammonium acetate was added to precipitate

and remove DNA-bound proteins. 2 volumes of 100% ethanol were added at RT and the tube swirled until the solution was thoroughly mixed and the DNA precipitated. Since we were interested in retaining also the smaller DNA molecules such as episomes, the DNA was collected by centrifugation at 5000 g for 5 minutes at RT in a swinging-bucket rotor. The precipitate was washed twice with 70% ethanol and the DNA collected by centrifugation as described above. The pellet was then dried in an open tube at RT until the last visible traces of ethanol evaporated. The pellet was resuspended in 1 ml TE buffer and allowed to dissolve at 4^{o}C. The absorbance of the DNA was measured at 260nm and 280 nm to ensure that the ratio of A260 to A280 is greater than 1.75. A lower ratio is an indication that significant amounts of protein remain in the preparation. The DNA was stored at 4^{o}C.

2.11.2 DNA extraction from transduced cells

For applications which did not require the DNA to be preserved in high molecular weight fragments, Qiagen DNEasy DNA extraction kit was used according to manufacturer's instructions.

2.12 Southern Blotting

Southern blotting is a method for detecting specific DNA sequences present within a DNA sample. The DNA samples are usually digested using restriction enzymes, and the fragments are separated by size by agarose gel electrophoresis. The DNA fragments are transferred onto a membrane using capillary transfer, and the membrane is then exposed to a hybridisation probe which is labelled to enable detection. The probe is designed to be complimentary to the DNA sequence of interest (Southern, 1975).

2.12.1 Materials

[α-32P] dATP.

Dithiothreitol (DTT). Sigma.

DNA polymerase I Large (Klenow) fragment. 5U/µl, Promega.

Generuler 1 kb Plus DNA ladder. Bioline.

GeneScreen Plus membrane. NEN, Cat NEF9883.

Pharmacia Microspin S-300HR column. GE Healthcare.

Random hexanucleotides, dN6. 60 ng/µl. Sigma.

Spermidine. Sigma.

2.12.1.1 Solutions

20x SSC: 3 M NaCl, 0.3 M tri-sodium citrate dehydrate.

CTG Mixture: 0.5 mM each dCTP, dGTP and dTTP in 10 mM Tris HCl pH7.5.

Denaturing solution: 0.4 M NaOH, 0.6 M NaCl.

Neutralising solution: 1.5 M NaCl, 0.5 M Tris HCl pH 7.5.

Nick Translation Buffer (NTB), 10x: 0.5 M Tris ClH pH 7.5, 0.1 M MgSO4, 1 mM DTT, 500 µg/ml BSA.

Random Priming buffer (RPB): 0.5 M Tris HCl pH 6.9, 0.1 M $MgSO_4$, 1 mM DTT.

2.12.1.1 Equipment

Hybridisation oven.

Geiger counter.

Large electrophoresis gel tank.

X-ray film cassette.

2.12.2 Digestion of mammalian genomic DNA

5 µg of high molecular weight DNA (see section 2.2.11.1) was diluted with water and 10x enzyme buffer for a final digestion volume of 60 µl (ie. 6 µl of 10x buffer + water + DNA = 56 µl) and allowed to stand at RT for 30 min. Spermidine was added to a final concentration of 2 mM (1.2 µl of 100 mM) to increase enzyme activity and the mixture incubated at RT for 1 h. 3 µl of an appropriate restriction enzyme was added and the digest incubated overnight at the optimum temperature for the enzyme. After 12 h, 1 µl of enzyme was added, and the incubation continued for further 12 h or until viscosity

was gone. A 0.7 % agarose gel was run with 1 µl of digests to estimate amounts to be loaded in Southern Blot gel.

2.12.3 Preparation of hot ladder

A mixture was made up of 1 µg DNA ladder, 2 µl 10x NTB, 4 µl CTG, 1 µl (10 µCi) [α-32P] dATP, 1 µl Klenow, water was added to 20 µl, and the mix incubated at RT for 90 min. 40 µl of TE was then added to stop reaction. A Microspin S-300HR column was packed by spinning at 3003g for 1 min. Sample was loaded on column and centrifuged at 3000 g for 2 min.
Approximately 15 Geiger counts of ladder were loaded per lane. Remaining hot ladder was stored at -20^{o}C.

2.12.4 Agarose gel for Southern Blot

A 0.7% agarose gel was cast in 1x TAE with ethidium bromide. The gel was loaded with 10 µg of digested genomic DNA per lane and hot DNA ladder. The gel was run in TAE with ethidium bromide overnight in at 35V, constant voltage.

2.12.5 Salt transfer and UV fixation protocol

Gel was agitated in 0.25 M HCl for 10 min and washed twice with water. It was then agitated in denaturing solution for 30 min and washed twice with water. Lastly it was agitated in neutralising solution for at least 30 min.

A GeneScreen Plus membrane was cut to the size of the gel and pre-wet in water, followed by equilibrating in 10xSSC for 15 min. A capillary blot was then set up using 10xSSC. Order from bottom: Inverted gel, GeneScreen Plus membrane, two 10xSSC – soaked Whatman 3MM paper pieces, stack of blotting paper about 10 cm high, glass plate, book weighing about 500 g. The blot was left to transfer overnight.
The following day, the membrane was retrieved and agitated in 0.4 M NaOH for 1 min followed by 0.2 M Tris-ClH pH 7.5, 1 x SSC for 1 min. It was washed with 2xSSC and DNA crosslinked to membrane using Stratagene UV crosslinker.

2.12.6 PCR for GFP Probe

PCR was performed to create a 717 bp fragment of the GFP gene. Plasmid pRRLsc SV40 GFP was used as template. Forward primer: 3'ATG GTG AGC AAG GGC GAG, reverse primer: 3'CTT GTA CAG CTC GTC CAT. The reaction was set up as in section 2.2.5.6.2 with an annealing temperature of 56 ^{o}C and a total of 35 cycles.

2.12.7 Probe labelling by random priming

100 ng of DNA fragment to be labelled was made up to 4 µl with TE, 1 µl of 60 ng/µl dN6 was added, the mixture boiled for 5 min, and cooled on ice for 2 min.
12 µl of the following mixture was then added to the denatured DNA: 12 µl dH_2O
3 µl 10x random priming buffer, 3 µl CTG. Finally 2 µl (20 µCi) [α-32P] dATP and 1 µl Klenow were added and the reaction incubated at RT for 1 h.
After incubation, 40 µl TE buffer was added, sample was loaded onto a pre-packed Microspin column at 3000 rpm for 2 min.

2.12.8 Hybridisation

Membrane was pre-hybridised with 15 ml Church mix at 68 ^{o}C for at least 1 h.
Labelled probe was boiled for 5 min, transferred to ice for 2 min, and added to hybridisation tube. The reaction was left to hybridise overnight.
The membrane was washed with 2xSSC, 0.5% SDS at 65^{o}C 3 times quickly and 3 x 10 min washes. Finally it was washed with 2xSSC and exposed to Phosphoimager screen.

Church mix: 1% BSA, 7% SDS, 0.5 M Phosphate buffer (made with 28.4 g Na_2HPO_4, 6.9 g Na_2HPO_4•H2O dissolved in water, made up to 125 ml, and added to a total of 500 ml Church mix).

2.13 Polymerase chain reaction –based techniques

2.13.1 Polymerase chain reaction (PCR)

2.13.1.1 Materials & Equipment

2x PCR Mastermix. Genesys Ltd.

Peltier Thermal Cycler PCT-200. MJ Research.

2.13.1.2 Standard Protocol

PCR reactions were performed at various stages throughout the thesis. The general materials & methods are outlined here, and the primers and PCR program used for each reaction are detailed in the appropriate sections.

Template was diluted to 33 ng/µl where possible. Primers were diluted to 1 µM. Master solution was prepared by adding 12.5 µl Mastermix, 1 µl of each forward and reverse primer, and 7.5 µl ddH_2O for each reaction, allowing 10 % extra in total. In each PCR tube were added 3 µl template and 22 µl master solution.

The standard PCR protocol was 92 ^oC for 2 min, 92 ^oC for 30 s, annealing temp for 30 s, 68 ^oC for 30s (or 1 min per 1000 bp), last 3 steps repeated for 30-45 times, 68 ^oC for 10 min and 4 ^oC for storage or product. The annealing temperature was calculated separately for each primer pair.

2.13.2 Quantitative Real-time PCR (qRT-PCR)

Whereas standard PCR will give essentially identical numbers of an amplified sequence regardless of the amount of template sequences, Quantitative Real-Time Polymerase Chain Reaction, or qrt-PCR, can be used to simultaneously amplify and quantify a target DNA sequence (Higuchi *et al.*, 1992).

Quantification can be done as absolute copy numbers or relative amounts when normalised to standard amounts of a “housekeeping gene” known to be present within

the template DNA. In this thesis it was used for vector titrations and the determination of vector genome copy numbers in transduced cells. The general materials and method are listed here, any modifications are noted in the appropriate sections.

2.13.2.1 Materials & Equipment

Corbett Rotor-Gene Q. Qiagen.

Corbett Rotor-Gene software. Qiagen.

SYBR No-Rox Mastermix. Quantace.

2.13.2.2 Standard protocol

The primer sequences used for each qrt-PCR reaction are specified in the appropriate sections. The standard cycling program was 95^{o}C for 10 min, 40 x 60 ^{o}C for 30s and 95 ^{o}C for 1 min. The fluorescence was read at 95 ^{o}C and analysed using Corbett Rotor-Gene software

2.13.3 qrt-PCR of Lentiviral episomes in CHO cells

The quantification of episomes in transduced CHO cells was carried out by Real-Time PCR using the Roche LightCycler 480 system. 100 ng of high molecular weight genomic DNA (see section 2.2.11.1) in 2 µl was used as a template in each reaction. 10 µl of reaction mix contained 300 nM of each primer and 5 µl Absolute Blue qPCR SYBR Green 2X Buffer (Thermo Scientific). Primers were as described in Table 2.1 below.

Standards for the 1-LTR, 2-LTR and LRT reactions were made by diluting a plasmid containing each target sequence so the copy number ranged from 10^2 to 10^9 per reaction. The standard for the β-actin reaction was based on DNA concentration, and was calculated as follows: $1x10^7$ cells were counted and the DNA was extracted using the high molecular weight protocol. On average, this resulted in 300 ng of DNA in a volume of 200 µl. Hence, the amount of DNA expected to be obtained was 3 pg/ cell.

Each sample was analysed twice in triplicate, and any outliers or replicates producing an abnormal melting curve were excluded. Averages from all experiments and duplicates were used in further analyses.

The reaction conditions were standard qRT-PCR cycling: 50^{o}C 2 min, 95^{o}C 10 min, 50x (95^{o}C 15 sec, 60^{o}C 1 min).

Primer name	**Experiment(s)**	**Reaction(s)**	**Primer sequence**
4nr	qrt-PCR	1-LTR	5’ GTCTGAGGGATCTCTAGTTAC
2nr	qrt-PCR Inverse PCR	1-LTR 1st PCR	5’ TCTCTGGTTAGACCAGATCTG
U3 lenti 8	qrt-PCR Inverse PCR	2-LTR 2nd PCR	5’ GTTGGGAGTGAATTAGCCCT
MH535	qrt-PCR	2-LTR LRT	5’ AACTAGGGAACCCACTGCTTAAG
MH532	qrt-PCR	LRT	5’ GAGTCCTGCGTCGAGAGAGC
Actin F	qrt-PCR	B-actin	5’ TGGCATCCACGAAACTACAT
Actin R	qrt-PCR	B-actin	5’ TGGTACCACCAGACAGCACT
HIV 543 bio	LAM-PCR	Linear PCR	5’ AGTGCTTCAAGTAGTGTGTGCC
SK LTR 1	Inverse PCR	1st PCR 2nd PCR	5’GAGCTCTCTGGCTAACTAGG
SK4 LTR bio	LAM-PCR	1st exp. PCR	5’ AGTAGTGTGTGCCCGTCTGT
LC I	LAM-PCR	1st exp. PCR	5’ GATCTGAATTCAGTGGCACAG
HIV LTR 566	LAM-PCR	2nd exp. PCR	5’ GTCTGTTGTGTGACTCTGGTAAC
LC II	LAM-PCR	2nd exp. PCR	5’ AGTGGCACAGCAGTTAGG

Table 2.1. Primer sequences used in qRT-PCR and LAM-PCR for lentivectors.

2.13.4 Linear amplification –mediated PCR (LAM-PCR)

Linear amplification-mediated PCR (LAM-PCR) is a highly sensitive method that can be used to identify unknown sequences flanking known sequences (Schmidt *et al.*, 2007). The protocol consists of a linear PCR with a biotinylated primer, and the resulting product is purified using streptavidin-coated magnetic beads. A second strand is synthesised using random hexanucleotide primers and Klenow fragment of DNA polymerase, and the double-stranded product is digested with a suitable restriction enzyme that cuts within ~200 bp from the 1st primer into the vector sequence and at unknown sites into the genomic DNA. Next, a unidirectional double-stranded linker is ligated to the sticky end created by the restriction enzyme. The DNA is denatured using NaOH and subsequently subjected to two exponential nested PCR reactions achieve maximum amounts of specific product. The first set of primers bind to the linker cassette and 5'biotinylated vector, the second set bind to linker cassette and vector sequence (LTR). The products can be visualised on an agarose gel and compared to expected fragment lengths, and sequenced either by extracting the bands from the gel or by conventional shotgun cloning and sequencing (Gabriel *et al.*, 2009; Schmidt *et al.*, 2007).

Here, LAM-PCR was performed on all clonal cell line hmw DNA, with integrating clone 6 as a positive control and DNA from CHO cells transduced with IDLV SV40 GFP and harvested 24 h post-transduction as a negative control. The protocol was performed as previously published (Schmidt *et al.*, 2007).

The LAM-PCR reactions were carried out by the author in Christof Von Kalle's laboratory at the National Centre for Tumour Diseases at the German Cancer Research Institute in Heidelberg, Germany.

The subsequent Deep sequencing was carried out in the same laboratory using the Roche 454 sequencing system. The bioinformatic analysis of the results was performed by Dr Uwe Appelt.

2.13.4.1 Linear PCR

Each reaction contained 2 µl of hmw DNA ([c] 250 ng/µl), 2.5 µl 10X Taq Buffer (Qiagen), 1 µl dNTPs (10 µM each ,Fermentas), 0.25 µl Taq (Qiagen), 0.5 µl Primer LRT 543 bio (0.5 pmol/µl), and water to a total volume of 25 µl. The PCR reaction was performed as follows: 95^{o}C for 2 min, 95^{o}C for 45 s, 58^{o}C for 45s, 72^{o}C for 1 min, steps 2-4 repeated 50 times, 0.25 µl Taq added, Steps 2-4 repeated 50 times,72^{o}C for 5 min.

2.13.4.2 Restriction enzyme digest

The product was digested with Tsp509I, which cuts ~210 bp into the vector sequence from the 1st primer in a LTR-backbone or LTR-genome junction, and ~105 bp into the second LTR sequence in a LTR-LTR junction.

2.13.4.3 1st & 2nd Exponential PCR

Each reaction contained 2 µl of product from the linear PCR reaction, together with 2.5 µl 10X Taq Buffer (Qiagen), 0.5 µl dNTPs (10 µM each), 0.25 µl each primer (SK4 LTR bio and LC-1), 0.25 Taq (Qiagen) and water to 25 µl. The reaction conditions were the same as above, but repeated only for 35 cycles. For the 2nd nested PCR, 1 µl product from previous reaction was used as a template for primers HIV LTR 566 and LC II.

The products of the reaction were visualised on a Spreadex gel (Elchrom Scientific). They were also subjected to sequencing using the Roche 454 system, creating an average of ~1200 independent sequence reads per sample.

2.14 Adeno-associated virus manufacture, purification and titration

2.14.1 Materials

2.14.1.1 Plasmids

pAd DeltaF6. PlasmidFactory.

p5E18-V2/9. PlasmidFactory.

2.14.1.2 Reagents & solutions

20x SSC (1 L): 1.5 M NaCl, 0.3 M Na_3 citrate.

25:24:1 phenol/chloroform/isoamyl alcohol. Sigma.

2x Proteinase K buffer: 20mM Tris.Cl (pH8.0), 20mM EDTA (pH8.0), 1% (w/v) SDS

3M sodium acetate. pH to 5.2 with acetic acid. Sigma.

70mM Na_2HPO_4. Autoclaved.

Benzonase (E8263 Ultrapure). Sigma.

DNase 1. 10 mg/ml in H_20. Fermentas.

Glycogen. Invitrogen.

HEPES buffer. 0.3M NaCl, 40mM HEPES, pH 7.05. Autoclaved.

HEPES. H7523, Sigma.

HEPES/Na_2HPO_4 buffer. 250 µl 70 mM Na_2HPO_4 and 12.25 ml HEPES, pH 7.05.

Hybridisation buffer (using Amersham RPN3000 kit). 34 ml ECL gold hybridisation buffer to cover blot, 1 g NaCl, 1.7 g blocking agent).

Lysis buffer. 0.15M NaCl, 50mM Tris-HCl, pH 8.5.

Optiprep density gradient medium 60% (w/v) solution of iodixanol in water. Sigma.

PBS-MK solution. 1xPBS with 5 mM $MgCl_2$ and 12.5 mM KCl added.

Phenol red solution. 0.5% in PBS, Sigma.

Polyethylene glycol PEG-8000. Promega.

Primary wash buffer (1 L): 6 M urea, 0.4% SDS, 0.5x SSC.

Proteinase K. Sigma.

Resuspension buffer. 0.4 M NaOH/10 mM EDTA pH 8.0.

2.14.1.3 Equipment

Amicon Ultra -15 filters. PL100, Millipore.

Autoradiography film. GE Healthcare.

Beckman L7-65 Ultracentrifuge.

Beckman Quick-Seal Ultra-Clear 25 x 89 mm tubes

ECL chemiluminescence kit. Amersham.

ECL direct nucleic acid system. GE Healthcare.

Hybridisation membrane. Hybond N+, GE Healthcare.

Nanodrop ND-1000 spectrophotometer. Labtech.

Type 70Ti rotor. Beckman.

2.14.2 Single-stranded AAV preparations

ssAAV8 CAG GFP, ssAAV CAG GFP miniMAR and ssAAV CAG GFP β-Ifn S/MAR (from here on referred to as ssAAV CAG GFP S/MAR) vectors were kindly donated by Dr's Keith and Helen Foster. They were manufactured using the same methods detailed below and titrated with the dot blot method, also as detailed below (section 4.2.1.4.2). The titres of vectors used in the experiment were $3.31x10^{12}$ for ssAAV8 CMV GFP, $4.89x10^{12}$ for ssAAV8 GFP miniMAR and $1.63x10^{11}$ for ssAAV8 GFP S/MAR.

2.14.3 Manufacture of AAV by transient transfection

Recombinant AAV was made by transient transfection of 293T cells. For each prep, 10^7 cells were seeded on 10x 185 cm^2 flasks in full DMEM and transfected 24 hours later when the cells had reached 80-90 % confluency. A total of 500 µg of plasmids were used for each transfection, which for the scAAV preps consisted of 86 µg of either pscAAV CMV GFP or pscAAV CMV GFP mMAR, 280 µg of pAdDeltaF6, and 133 µg of p5E18-V2. The plasmids were diluted in a total of 12.5 ml of 250 mM $CaCl_2$. $CaCl_2$/plasmid was then added dropwise to HEPES/Na_2HPO_4 buffer, whilst constantly vortexing the solution. The solution was left for 15 min at RT for precipitate to form. 200 ml of 2% FCS DMEM was then added to neutralise, media was removed from the flasks and replaced with 22 ml of the transfection media.

Cells were harvested by shaking and scraping 2 days post-transfection and the resulting supernatant centrifuged at 1000 rpm for 10 min to pellet. The supernatant was decanted into a sterile flask and the pellet resuspended in 15 ml lysis buffer and stored at -80 ^oC. To the 200 ml of reserved transfection supernatant, 50 ml of 40 % PEG8000 / 2.5 M NaCl was added, the solution mixed well by inverting 10-20 times, and incubated at 4^oC overnight. The supernatant / PEG mixture was then centrifuged at 4^oC for 30 min at 2500 *g*, the supernatant was discarded and the cell pellet resuspended in a 1-2 ml lysis buffer taken from the cell lysate. Resuspended supernatant / PEG mixture was then added into the cell lysate, and stored at -80^oC.

2.14.4 AAV purification.

To release virus, the lysate was frozen and thawed 3 times, followed by 30 minutes incubation at 37 ^oC with 50 U/ml Benzonase in order to remove any contaminating plasmid from the viral prep.

The lysate was clarified by centrifugation at 3700 *g* at RT for 20 min and passed though a 0.45 µm filter. The iodixanol solutions were prepared using 60 % iodixanol, phenol red solution, and 5x PBS-MK solution.

Solution	Iodixanol	5 M NaCl	5x PBS-MK	H_2O	Phenol red
15 %	12.5 ml	10 ml	10 ml	17.5 ml	-
25 %	20.8 ml	-	10 ml	19.2 ml	100 µl
40 %	33.3 ml	-	10 ml	6.7 ml	-
60 %	50 ml	-	-	-	100 µl

Table 2.2 Iodixanol fractions for AAV purification.
Volumes stated for making 50 ml each.

The solutions were layered very carefully into a Beckman 25 x 89 mm tube, using a length of sterile HPLC tubing attached to a 20 ml syringe, in the following order: 9 ml of 15% solution, 6 ml of 25% solution, 5 ml of 40% solution and 5 ml of 60% solution. The crude lysate was layered very carefully over the top of the gradient. The tube was filled up to the bottom of the neck of the tube with lysis buffer.
Samples were centrifuged at 350 000 *g* for 1 h at 18 °C. After centrifugation, a needle was inserted into the interface between the 60 % and 40 % iodixanol steps (clear step) and 4 ml of this fraction was removed.

2.14.5 Desalting and concentration.

Desalting and concentration of purified AAV samples was carried out using Amicon filters. The filter was pre-rinsed by adding 5 ml 1x PBS-MK and centrifuging for 15 min at 4000 *g*. 5 ml of 1x PBS-MK was added to the AAV fraction, the total volume added to filter device and centrifuged at 4000 *g* for 15-20 min until volume was reduced to ~0.5 ml. 15 ml of 1x PBS-MK was added to filter and centrifugation repeated until volume was reduced again. This was done 3 times to desalt the viral preparation. During the final spin the volume was reduced to 800 – 1000 µl. This was aliquoted in 50-200 µl volumes and stored at -80 ^{o}C.

2.14.6 Quantification of AAV by Dot Blot.

2.14.6.1 Preparation of samples.

To remove any contaminating plasmid DNA, 2 µl and 10 µl of each AAV prep was digested with 5 U of DNase I in 200 µl DMEM for 1 h at 37 °C. An equal volume of 100 µg Proteinase K in 2 x Proteinase K buffer was added and the mixture was incubated for a further 1 h at 37 °C.

A double extraction/purification of viral DNA was performed by adding 400 µl of 25:24:1 phenol/chloroform/isoamyl alcohol, mixing thoroughly and spinning at 13 000 rpm for 10 min at 4 ^{o}C. The DNA was precipitated by adding 1/10 volume (40µl) of 3 M sodium acetate, 40 µg glycogen and 2.5 volumes of 100% ethanol, and incubated at -80 °C overnight. The precipitate was centrifuged at 13 000 rpm for 20 min at 4 °C, the pellet washed with 70% ethanol and left to dry. The DNA was finally resuspended in 400 µl resuspension buffer.

For the standard, a 2-fold serial dilution of the pscAAV CMV GFP plasmid was prepared, ranging from 80 ng to 0.3125 ng. The dilutions were mixed with 400 µl of resuspension buffer and the viral and plasmid samples were heated for 5 min at 100 °C and cooled on ice for 2 min.

2.14.6.2 Preparation of Dot Blots

All denatured DNAs were applied onto a hybridisation membrane, secured onto dot-blot apparatus, vacuum was applied, wells rinsed with 400 µl resuspension buffer and vacuum applied to dry. The apparatus was disassembled and the membrane rinsed in 2 x SSC.

2.14.6.3 Preparation of Probe

A fragment of the GFP in pscAAV CMV GFP was amplified by PCR. The primers used were 5’ TAG GCG TGT ACG GTG GGA GGT C (forward primer) and 5’GAA CTT CAG GGT CAG CTT GC (reverse primer), in a reaction containing 100 ng plasmid DNA, 40 nM each primer, and 12.5 µl 2X PCR MasterMix (Genesys Ltd) in a total reaction volume of 25 µl.

The cycling conditions used were 92 ^{o}C for 2 min, 98 ^{o}C for 1 min, 52 ^{o}C for 30 s, 72 ^{o}C for 1 min, and steps 2-4 were repeated 30 times. The PCR product was run on 2% agarose gel and the product was gel extracted (Qiagen gel extraction kit). The concentration of purified DNA was determined using a nanodrop and the DNA was diluted to a concentration of 10 ng/µl in sterile water.

The labelling of the DNA was performed using the ECL direct nucleic acid system according to the manufacturer's instructions. Briefly, 100 ng of the DNA sample was denatured by heating for 5 min in a boiling water bath and then cooled immediately on ice for 5 min. An equal volume (10 µl) of DNA labelling reagent containing horseradish peroxidase (HRP) complexed to a positively charged polymer was added to the cooled DNA and mixed gently. 10 µl of glutaraldehyde solution was then added into the tube to covalently link the peroxidase to the DNA, and the mixture was incubated for 10 min at 37 °C to allow for the reaction to completion.

2.14.6.4 Membrane Hybridisation & Signal Detection

Hybridisation and stringent washes were carried out using the ECL system (GE Healthcare) according to manufacturer's instructions and buffers optimised as specified below.

To 34 ml ECL hybridisation buffer was added 1 g NaCl and 1.7 g blocking agent, and the mixture was preheated to 42°C. The blot was prehybridised in 5 ml of buffer for 1 h. Labelled probe (from section 2.2.14.5.3) was then added to the solution and hybridised overnight at 42°C. The blot was transferred to primary wash buffer and washed for 20 min with gentle agitation at 42 °C. The wash was repeated for 20 min at 42 °C in fresh primary wash buffer. The buffer was removed and the blot washed 2 times in 100 ml 20 x SSC for 5 min each time at RT.

The blot was developed using ECL chemiluminescence kit (Amersham). The detection agent containing luminol was added directly onto the blot in the darkroom and incubated for 1 min, to allow for the HRP bound to the probe to catalyse the oxidation of luminol resulting in a chemiluminescent signal. The blot was wrapped in Saranwrap, placed in a film cassette, and exposed to autoradiography film (GE Healthcare).

2.14.6.5 Calculations to determine AAV particle number

The MW in daltons of the control double stranded DNA molecule was calculated by multiplying the size of molecule in bp by 650 (average MW in daltons of 1 bp). A weight in grams equal to the molecule's weight in Daltons has a number of molecules in it equal to Avogadro's number, and this is further divided by 10^9 given the number of molecules in a nanogram. The intensity of the control and sample dots was then compared, and based on this an estimate was made of the amount of DNA in ng in the sample dots. This value was then multiplied by the calculated number of molecules/ng control DNA and then by 2 (as AAV is single stranded and the control plasmid is double stranded) to give an estimated number of particles in the sample AAV. For scAAV9 CMV GFP MAR, the difference in size was adjusted for by dividing the control AAV MW by the test AAV MW and then multiplying the calculated amount of particles by this amount.

Based on the dot blot hybridisation analysis, titres of the vectors were estimated to be at 3.92×10^{11} for scAAV9 CMV GP and 3.83×10^{11} for scAAV9 CMV GFP mMAR.

2.14.7 Quantification of AAV by Real-time Quantitative PCR

2.14.7.1 Vector sample preparation

DNA from 10 µl of each virus prep was prepared as detailed in section 4.2.1.4.1. After the final wash with 70% ethanol, the DNA pellets were left to air-dry, and resuspended in 100 µl of TE buffer. Depending on expected vector prep titre based on dot blot, this viral DNA was then either used undiluted or diluted from 10^{-2} to
10^{-6} , diluting the prep 1:10 and using every second one for assay.

2.14.7.2 Standard preparation

Standards were made up in 500µl TE buffer, ranging from ^{109} to 10^3 copies/ 5µl.
Calculation of M (molecular weight for dsDNA, assuming 1 Mol bp = 650 g)
M = (650 g/bp x length [bp]) / mol [g/mol]

Calculation of n (moles): n = m[g DNA] / M
Conversion into number of molecules N: (where Avogadro's number A = 6.022×10^{23} molecules/mol) N = n x A (molecules = mol x molecules/mol)

For pscAAV CMV GFP: length = 4723bp → M = 3.06995×10^{6} g/mol
For 10^{10} copies in 5µl: n = N/A = $10^{10}/6.022 \times 10^{23} = 1.66 \times 10^{-14}$ mol
For 1.66×10^{-14} mol in 5µl: m = n x M = 1.66×10^{-14} x $3.06995 \times 10^{6} = 5.096117 \times 10^{-8}$ g
= 50.96 ng in 5µl.

2.14.7.3 qPCR

Real-time quantitative PCR was performed with standards in triplicate and sample dilutions in duplicate. Each reaction contained 5 µl sample and 15 µl Mastermix (4.88 µl dH_2O, 10 µl 2X Quantace SYBR No-Rox MasterMix containing SYBR green and polymerase, and 0.06 µl each forward and reverse primer [100µM]). The primers used for the reaction were:
Forward primer AAV GFP F: 5'GAC GGC AAC ATC CTG GGG CAC AAG, and
Reverse primer AAV GFP R: 5'CGG CGG CGG TCA CGA ACT C.
The reaction was run using the standard qPCR cycling program (see section 2.2.13.2.2). The titres calculated from the qPCR results were approximately 10-fold lower than those calculated from dot blot.

2.15 *In vitro* evaluation of AAV

For the transduction experiment with cell cycle arrest, transduction efficiency was measured at the time of induction of cell cycle arrest, 24 h after transduction. It was deemed necessary to induce cell cycle arrest 24 h post-transduction in order to avoid dilution of vector genomes by excessive cell division prior to cell cycle arrest.

Vector	Titre
ssAAV8 CAG GFP	$3.31x10^{12}$
ssAAV8 CAG GFP mMAR	$4.89x10^{12}$
ssAAV8 CAG GFP S/MAR	$1.63x10^{11}$
scAAV9 CMV GFP	3.92×10^{11}
scAAV9 CMV GFP mMAR	3.83×10^{11}

Table 2.3 Titres of AAV vectors used in transduction experiments.
All titres were calculated from the dot blot.

1 x 10^5 CHO cells per well were seeded on 6-well plates. After 24 h, the media was removed and replaced with 500 µl serum-free media containing AAV at a MOI of 10^5. The cells were incubated at 37^{o}C for 5 h and then 500 µl of media with 20% FBS was added per well.

24 h post-transduction, a fraction of cells from one well transduced with each vector and a mock transduced control was analysed for GFP expression by flow cytometry, and a PI stain was done on another fraction of the same cells for cell cycle analysis. The media in the remaining wells was changed for DMEM without methionine and with proline and 2% FBS. The media was replaced with fresh each day for the next 4 days.

At the end of the quiescent period, one well for each vector and a control was used for GFP expression analysis and cell cycle analysis. The remaining cells were split 1:5 and seeded in full DMEM with 10% FBS. Cell cycle analysis was repeated 3 days later. After this, cells were split 3 times a week and GFP expression was measured by flow cytometry at least once a week. Culture was discontinued after ~ 40 days when GFP expression was deemed to be at stable levels.

For the cells not undergoing cell cycle arrest, transduction of CHO cells with AAV vectors was performed as above. 24 hours post-transduction, one well transduced with each virus and a control were analysed for GFP expression by flow cytometry, and a PI stain was done on the same cells for cell cycle analysis. Cell cycle and GFP expression analyses were repeated 4 days later, and following this the cells were cultured normally by splitting 3 times a week and measuring GFP expression by flow cytometry at least once a week.

Culture was discontinued after ~ 40 days when GFP expression was deemed to have reached stable levels.

2.16 Lentivector manufacture, concentration and titration

2.16.1 Materials

2.16.1.1 Plasmids

pMDLg/pRRE

pMDLg/pRREintD64V

pRSV-REV

pMD2.VSV-G

pHR'SIN-cPPT-CEW

2.16.1.2 Reagents & solutions

ABI universal master mix. Applied Biosystems.

2x HBS. 100 mM HEPES, 281 mM NaCl, 1.5 mM Na_2HPO_4, pH to 7.12 (7.11-7.13) with 5 and 1 M NaOH.

$CaCl_2$. 2.5 M, Sigma.
DNAse I. Promega.

HEPES. Sigma.

Paraformaldehyde. Sigma.

Polybrene: SIGMA H-9268. Make up 8 mg/ml (1000X) in water. Store at 4^o C.

TE Buffer. Qiagen.

2.16.1.3 Equipment

SW32.1Ti rotor. Beckman.

Polyallomer centrifuge tubes. Beckman.

2.16.2 Manufacture of lentivectors by transient transfection

293T cells were seeded at 3 x 10^6 cells per plate on 15 cm^2 plates, and were transfected 2 days later having reached ~ 90% confluency. The media was replaced with 20 ml fresh full media 2 h before transfection. The DNA mix was prepared in a molar ratio of 1:1:1:2 using 12.5 µg pMDLg/pRRE or pMDLg/pRREintD64V,
6.25 µg pRSV-REV, 7 µg pMD2.VSV-G and 32 µg of either pRRLsc SV40 GFP, pRRLsc SV40 GFP mMAR or pRRLsin PPT CMV eGPF WPRE (lentiviral plasmids were a gift of R. Yañez, RHUL). This was made up to a total volume of 112.5 µl with TE buffer and then 1012.5 µl water was added to make 0.1 x TE. 125 µl 2.5 M $CaCl_2$ was added, the solution vortexed and incubated for 5 min at room temperature. Then, 1250 µl 2 x HBS was added dropwise while vortexing the DNA/$CaCl_2$ mix at full speed. This mixture was then added to cells and the cells returned to incubators set at 37^{o}C, with 5% CO_2. 16 h after transfection the media was removed and replaced with 18 ml fresh media per plate. The supernatant containing lentivector particles was harvested on days 2 and 3 post-transfection.

Harvested medium at was centrifuged at 2500 rpm for 10 min at RT and the supernatants filtered through a 0.22 µm filter. Filtered medium was transferred to high speed polyallomer centrifuge tubes and spun at 50 000 x g for 2 h at 4^{o}C, using SW32.1Ti rotor (Beckman). Supernatant was completely removed. 50 µl of DMEM (without supplements) was added per tube, pipetted up and down for 15 times to resuspend the concentrated pellet. The virus preparation was transferred to an eppendorf tube and centrifuged for 10 min at 4000 rpm, RT to remove debris. The supernatant was then transferred to a new eppendorf tube and adjusted to 10 mM $MgCl_2$. 5 U/ml DNase I was added to the preps to remove plasmid contamination and incubated at 37^{o}C for 30 min. The vector stocks were then frozen in aliquots at -80 ^{o}C until use.

2.16.3 Lentivector titration by GFP flow cytometry

HeLa cells were chosen for the IDLV titration by flow cytometry for GFP expression, as the expression kinetics from IDLV episomes is slower than that from ICLVs and a fast-growing cell line may result in excessive dilution of the gene product (Wanisch *et al.*, 2009). HeLa cells are further slowed down by the polybrene used to aid transduction, therefore they were used for all GFP titrations of IDLVs in this chapter.

The cells were seeded at 10^5 cells per well in 2 ml medium on 6-well plates. The next day, 10-fold vector dilutions were made in full DMEM,
ranging from 10^{-3} to 10^{-6}. Medium was removed from wells and 1 ml of DMEM with 16 µg/ml polybrene was added per well, as it has been shown to increase transduction efficiency by neutralising the charge repulsion between virions and the cell surface (Davis *et al.*, 2004). The cells were then transduced with the addition of 1 ml of each virus dilution /well, or 1 ml DMEM for mock.72 h after transduction cells were harvested, resuspended in 300 µl PBS 2% paraformaldehyde and run through flow cytometer, using the mock transduced cells to set background-control levels of fluorescence. Titre was calculated using dilution with 1-10% green cells, using the formula [GFP transducing units (TU)/ml = % green cells x 10^5 (cells/well on day 0) x 1/vector dilution]. Only vectors with a titre higher than 10^7 TU/ml were used in the *in vitro* transduction experiments.

2.16.4 Lentivector titration by qrt-PCR

6-well plates were seeded with 10^5 HeLa cells per well in 2 ml medium.
The cells were transduced the next day with 1 ml of $5x10^{-4}$ and $5x10^{-5}$ dilutions of vector preps (MOIs 0.5 and 0.05 for a 10^8/ml vector) and 1 ml of DMEM with16 µg/ml polybrene and a mock control. 24 h after transduction, cells were harvested and DNA was prepared with DNeasy tissue kit according to manufacturer's instructions (Qiagen).

Standards of known copy or cell number were measured in triplicate and made in AE buffer. Samples were measured in duplicate. In the PCR negative control (duplicate) Qiagen DNA elution buffer AE was used instead of cell extract.

For human ß-actin qPCR the primers were as follows:
Forward primer: 5'-TCACCCACACTGTGCCCATCTACGA-3',
Reverse primer 5'-CAGCGGAACCGCTCATTGCCAATGG-3',
Probe 5'-(VIC)-ATGCCCTCCCCCATGCCATCCTGCGT-(TAMRA)-3'.
Reactions were set up using ABI universal master mix in a total volume of 20 µl, containing 10 µl master mix, 5 µl sample, primers to a final primer concentration of 300 nM each primer, and probe to a final concentration of concentration 200 nM. Standards used were 10-10^5 copies of pRTBACTSTD or "HT1080 cell equivalents"/reaction.

Standard cycling conditions were: 50˚C 2 min, 95˚C 10 min, 50x (95˚C 15 sec, 60˚C 1 min).

For lentiviral backbone qPCR (Bushman's late reverse transcript reaction (Miller *et al.*, 1997) the primers were:
Forward primer 5'-TGTGTGCCCGTCTGTTGTGT-3',
Reverse primer 5'GAGTCCTGCGTCGAGAGAGC-3'.
Reactions were set up using Quantace SensiMix SYBR No-Rox mastermix in a reaction volume of 20 µl, containing 10 µl master mix, 6.25 µl sample, and primers at a final concentration of 300 nM each. Standards used were 10-10^9 copies of pHR'SIN-cPPT-CEW per reaction. Standard cycling conditions were: 50^{o}C 2 min, 95^{o}C 10 min, 50x (95^{o}C 15 sec, 60^{o}C 1 min).

2.17 In vitro lentivector evaluation

2.17.1 Transduction and cell cycle arrest induction protocol for CHO cells

2 x 10^5 CHO cells per well were seeded on 6-well plates, as this was found to be the optimal cell density for successful induction of cell cycle arrest (see section 2.2.7.3.2).
After 24 h, the media was removed and replaced with 1 ml media containing polybrene at a final concentration of 8 µg/ml to aid transduction by neutralising the charge repulsion between the virions and the cell surface (Davis *et al.,* 2004).The cells were transduced by adding 1 ml media containing vector dilution corresponding to MOI 0.5 by flow cytometry. An MOI of less than 1 was chosen in order to avoid non-specific integration events, which may increase as higher MOIs of vector particles are used.
24 hours post-transduction, one well transduced with each virus and a control were analysed for GFP expression by flow cytometry, and a PI stain was done on another fraction of the same cells for cell cycle analysis. For the cells designated to undergo cell cycle arrest, the media in the remaining wells was changed for DMEM without methionine, with proline and 2% FBS. The media was replaced with fresh each day for the next 4 days. At the end of the quiescent period, one well for each virus and control was used for GFP expression analysis and cell cycle analysis.
The remaining cells were split 1:5 and seeded in DMEM with proline and 10% FBS. Cell cycle analysis was repeated 3 days later. After this, cells were split 3 times a week

and GFP expression was measured by flow cytometry at least once a week. The cells were stored in liquid N_2 after ~ 70 days.

For the control cells not undergoing cell cycle arrest, transduction was done as above. 24 h post-transduction, one well per vector was analysed for GFP expression and cell cycle phase by flow cytometry. Cell cycle and GFP expression analyses were repeated 4 days later, and following this the cells were cultured normally by splitting 3 times a week and measuring GFP expression by flow cytometry at least once a week. Cultures were discontinued after ~70 days and cells stored in liquid N_2.

2.17.2 Dilution cloning of lentivector –infected CHO cells

To obtain clonal populations of CHO cells transduced with IDLV vectors and stably expressing GFP, 1 cell population for IDLV SV40 GFP, IDLV SV40 GFP mMAR and IDLV CMV GFP each was retrieved from liquid N_2 and cultured for ~ 14 days prior to cloning. Cells were seeded on 96-well plates at a density of 0.7 cells/well in 100 µl media containing 50 µl fresh DMEM + Pro +PenStrep +10% FCS, and 50 µl CHO cell conditioned media. Conditioned media was taken from 24-hour old untransduced CHO cell cultures and filtered through 0.2µm syringe filters. After 7 days, 100 µl of the same 50-50 media was added to wells. After further 7 days in culture, wells containing GFP positive clonal populations were identified by FACS, and 12 clones for each vector were trypsinised and moved onto 12-well plates. For IDLV CMV GFP, only 5 GFP positive wells were identified, and at least 2 of these appeared to contain descendants of more than 1 cell. Cells from all 5 wells were trypsinised, mixed together and re-seeded at a density of 0.7 cells/well as previously. This time, 12 positive clones were obtained. When a sufficient amount of cells was available for each clone, clonal GFP expression was confirmed by flow cytometry.

2.17.3 Lentivector transduction and cell cycle arrest induction protocol for HeLa cells

Transduction of HeLa cells was performed as detailed for CHO cells (section 2.17.1), using IDLV SV40 GFP, IDLV SV40 GFP mMAR and IDLV CMV GFP at an MOI of 1. In a departure from the CHO cell protocol, $2x10^5$ HeLa cells were seeded in each well

at the start of the experiment, and the cells were only held in induced cell cycle arrest for 4 days. At the end of the quiescent period, the cells were not trypsinised and re-seeded but instead kept in the same wells and the media was changed every 2 days for 6 days until the cells recovered and resumed normal division.

2.17.4 Lentivector transduction and cell cycle arrest induction protocol for murine HSCs

2.17.4.1 Protocol for optimisation of induced cell cycle arrest in murine HSCs

To test the capacity of the media to hold the cells in cell cycle arrest or to promote survival and proliferation, 2 adult BalbC mice were sacrificed and the bone marrow obtained from the femurs and tibia by flushing with buffer (PBS pH 7.2, 0.5% BSA, 2 mM EDTA). Lineage-negative mHSCs were separated from rest of the bone marrow by performing lineage depletion using a Miltenyi MACS kit according to the manufacturer's instructions. Cells were seeded at a density of $8x10^4$ cells/ well in a 24-well plate (non –tissue culture treated) in 500 µl methionine-containing, 20 % FCS DME/F-12 with all 3 cytokines. Two days after seeding, the media was changed in 3 of the 6 wells for 500 µl methionine-free 2 % FCS media containing all 3 cytokines. The cells were kept in culture and live and dead cells were counted using Trypan blue at days 6, 9, 15 and 20 after seeding.

To count cells, the cells were carefully pipetted up and down for approximately 30 s and 10 µl was removed for counting. When the cell count exceeded $1x10^5$ cells/ well, the cells were split 1:4 prior to re-seeding by spinning down at 300 g for 4 min in an eppendorf, removing supernatant and resuspending in 500 µl fresh media for plating.

2.17.4.2 Protocol for transduction and a period of induced cell cycle arrest in murine HSCs

For the transduction experiment, 4 adult BalbC mice were sacrificed and the bone marrow obtained from the femurs and tibia as above, followed by lineage-negative mHSCs were separation using the Miltenyi MACS kit according to the manufacturer's instructions. Immediately following depletion, the lineage-negative cells were seeded at

a density of $1x10^5$ cells per well on a 24-well plate. The cells were seeded and transduced using 11 different conditions, as described in Table 2.4 below.
Combinations used in the experiment were as follows:

A. Methionine-free 2% FCS medium with no cytokines
B. Methionine-free 2% FCS medium with mSCF only
C. Methionine-free 2% FCS medium with all 3 cytokines
D. Complete media with no cytokines
E. Complete media with mSCF only
F. Complete media with all 3 cytokines

	Initial medium	Transduction Y/N	24 h	90 h
1	Medium A	N	Medium A	Medium F
2	Medium F	N	Medium F	Medium F
3	Medium A	Y	Medium A	Medium F
4	Medium B	Y	Medium A	Medium F
5	Medium B	Y	Medium B	Medium F
6	Medium F	Y	Medium A	Medium F
7	Medium F	Y	Medium B	Medium F
8	Medium F	Y	Medium C	Medium F
9	Medium F	Y	Medium F	Medium F
10	Medium D	Y	Medium D	Medium F
11	Medium E	Y	Medium E	Medium F

Table 2.4 Different combinations of media and cytokines used in HSC transduction experiment.

The cells were seeded in 300 µl initial medium and transduced directly after plating by adding 40 µl virus prep per well (IDLV CMV GFP, titre: $1x10^8$ vp/ml by flow cytometry, corresponding to MOI 40).
24h after plating and transduction, 200 µl media was added to counteract evaporation.
90h after transduction, the cells were pipetted up and down to disperse clumps and detach from well, and counted. $1x10^5$ cells were removed from each well, and centrifuged at 300 g for 4 min at 8^oC. The cells were then resuspended in 400 µl fresh

media and plated into fresh wells. Remaining cells were analysed by flow cytometry for GFP expression, excluding dead cells with a PI gate.

Four days after seeding and transduction, the cells in each well were resuspended and counted. If the cell count exceeded 10^5 cells /well, 10^4 cells were removed for flow cytometric analysis of GFP expression, again excluding dead cells using PI. The remaining cells were replaced in the culture dish resuspended in fresh full media. In wells were the cell count did not reach 10^5 cells /well, the cells were replaced in fresh media without removing any cells.

This procedure was repeated at 7, 11, 14 and 18 days after the start of the experiment.

2.18 Fluorescent *In Situ* Hybridisation

The FISH analysis was carried out by the author at the Cytogenetics Department at King's College, London, with the help of Dr Caroline Ogilvie and Dr Angela Davies.

2.18.1 Materials

2.18.1.1 Reagents

20x SSC. 3 M NaCl and 300 mM Trisodium Citrate, pH 7.0.

DAPI. 160 ng/ml, Cytocell.

Fixative. 3:1 Methanol:Glacial Acetic Acid. Sigma.

KCl, 75 mM. Sigma.

LSI/WCP Hybridisation buffer. Abbott Molecular.

Nick Translation System. Invitrogen.

Rubber solution. Weldtite Ltd.

SpectrumGreen and SpectrumOrange dUTPs. Abbott Molecular.

Tween®20. Sigma.

2.18.1.2 Equipment

Franklin electric pump.

Microscope imaging software. Isis 2, Metasystems.

Microscope. Zeiss Axioplan 2.

Savant Refrigerate vapour trap.

Savant SpeedVac® centrifuge / concentrator.

2.18.2 Cell Fixing Protocol

Approximately 1-5 x 10^6 CHO cells were detached using trypsin/EDTA. Cells were transferred into a 15ml Falcon tube and centrifuged at 1000 rpm for 5 minutes to pellet. The supernatant was decanted and the pellet resuspended in remaining drops of media. 7 ml of prewarmed, hypotonic KCl solution was added, tubes inverted to mix and incubated at RT for 5 min. This was done in order to enlarge the cells by osmosis, so that the nuclei would lie further apart on the resulting slide. Tubes were centrifuged 1200 rpm for 6 min, supernatant decanted, and the pellet again resuspended in the remaining drops by flicking. To fix the cells whilst preventing shrinkage, 5 ml of 5 % acetic acid was added, the tubes were inverted to mix, and centrifuged at 1200 rpm for 6 min. Supernatant was removed and pellets resuspended in 5ml of freshly made fixative. Any clumps were removed with a Pasteur pipette.
Cells were washed with fixative twice more to remove any traces of media. Finally, the fixed preparation was centrifuged at 1200 rpm for 6 min and stored at -20 ^{o}C.

2.18.3 FISH Probe labelling

FISH probes were prepared for each vector by labelling the corresponding transfer plasmid, each containing around least 3 kb complementary DNA, with SpectrumGreen and SpectrumOrange fluorochromes. The plasmid DNA was labelled using Nick Translation System according to the manufacturer's instructions, using 1 µg plasmid and labelling with both SpectrumGreen and SpectrumOrange dUTPs. The reaction was incubated at 15^{o}C for 1.5 h and the product stored in the dark at -20 ^{o}C.

Probe mixes were prepared for FISH by adding Cot-1 DNA to each labelled probe (50x the weight of probe) and then 2x the total volume 100% ethanol. The probes were pelleted and dried using Speedy Vac apparatus and Savant Centrifuge / concentrator. The dry pellets were resuspended in Abbott LSI/WCP Hybridisation buffer at a concentration of 6.7 ng/µl and stored in the dark at -20^{o}C.

2.18.4 FISH Run Procedure

Fixed cells were dropped onto slides pre-soaked in fixative and the area containing cells was marked using a diamond pen. Slides were soaked in PBS pH 7.4 for 5 minutes and then rinsed twice with sterile water. They were then dehydrated by incubating for 2 min each in 70%, 90 & and 100% ethanol. The slides were air-dried, and 15 µl (50ng) probe was applied onto a cover slip which was then placed onto the slide and sealed with rubber solution (Weldtite Ltd). The slides were incubated at 72^{o}C for 6 min and then at 37^{o}C overnight.

The next day, the coverslips were removed by washing in 4x SSC and 0.05% Tween pH 7.0. A stringent wash was performed for 5 min at 72^{o}C in 0.4x SSC pH 7.0. After this, the slides were incubated for 2 min at RT in 4x SSC/0.05% Tween pH 7.0, and then rinsed 3 times with PBS for 2 min each time. After the final PBS wash, the slides were air dried and 15 µl of the nuclear stain DAPI (Cytocell,160 ng/ml) was applied together with a new cover slip, which was then sealed with nail varnish.

2.18.5 Imaging and analysis of FISH signals

Although the cell cultures from which the FISH slides were derived were not blocked in metaphase, at least 50 spontaneous metaphases were present on all the slides. Images were taken for ~20 metaphases per slide using 100x magnification and the Zeiss Axioplan 2 microscope and Metasystems Isis 2 software.

Three slides were designated for detailed analysis: mock transduced control, and clonal cell lines transduced with either integrating or non-integrating SV40 GFP vector. For each slide, 80 fields using 40x magnification lens were processed by counting the number of nuclei in the field, the number of signals outside nuclei, number of signals

inside nuclei, and number of nuclei without signals. The ratio of signals inside nuclei versus outside nuclei was calculated for each field, to account for variation within different regions of the slide. The number of signals per nucleus was calculated as the number of "in" signals per field minus the number of nuclei without signals. Statistical analyses were performed on A) the ratio of signals inside v outside nuclei, using one-way ANOVA and comparing "true" samples to the negative control, and B) the number of signals per nucleus on each slide using one-way ANOVA.

2.19 Software for computational analysis

2.19.1 Software for sequence manipulation

Vector sequences and hypothetical vector structures were constructed using Vector NTI (Invitrogen).

2.19.2 SIDD Analysis

The computational analysis of vector sequences was performed using the WebSIDD service available at http://orange.genomecenter.ucdavis.edu/benham/sidd/

The algorithm considers 3 factors affecting the strand separation potential at each location of the sequence under investigation; the local thermodynamic stability of the base pair, the torsional energy of rotation of the single strands within unpaired regions, and the residual supercoiling free energy. The potential for strand separation is calculated for each base pair, as well as taking into account the competition between all sites in a domain under superhelical stress (Bi *et al.*, 2004).

The user controls some of the settings. The default setting assumes the input DNA sequence to be circular, if a linear sequence is specified, the algorithm adds 50 bp of G-C at the end to create a domain border and then treats the sequence as circular. The user must also choose between 2 types of energetics, near neighbour and co-polymeric. The co-polymeric type assigns a certain free energy for each A-T pair, and a different free energy for each G-C pair. The near neighbour type assigns different entropies and enthalpies to each of the 10 possible combinations of neighbour types.

The results are quite similar in both cases, and for analyses here the near neighbour opening energetic was used as more low-energy states are found, although computational time required is also increased (Bi *et al.*, 2004).

In the default settings, the temperature is assumed to be 37^{o}C, and the default salt concentration is consistent with standard *in vitro* conditions, and the settings were left as such for all sequences submitted for analyses here. The default level of superhelical stress corresponds to the midrange superhelicity of bacterial plasmids. The algorithm also assumes each region of strand separation is no greater than 250bp in size, based on observations made under physiological conditions (Benham *et al.*, 2004).

2.20 Statistical analyses

All statistical analyses were carried out using GraphPad Prism version 4.00 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com.

For the statistical analyses of the cell cycle data, Student's t-test was used to compare the proportions of cells separately for each cell cycle phase between treatments, ie, the proportion of cells in G_1 phase was compared between arrested and control, etc.

For the statistical analyses of data from transduction and transfection experiments, the percentages of cells expressing GFP were taken from the last measurement and One-way ANOVA was used to compare all groups. Tukey's post-hoc test was used to establish whether significant differences existed between specific sets of data.

Chapter 3: Effect of S/MAR elements and induced cell cycle arrest on retention of transfected non-viral vectors in CHO cells

3.1 Introduction

3.1.1. Testing mitotic stability of episomal non-viral vectors

Most successful studies to date reporting stable S/MAR-based episomes have utilised systems based on the pEPI plasmid. The pEPI has been reported to be retained for hundreds of cell generations in the absence of selection, but all previous work has imposed initial selection pressure after transfection in order to de-select cells not successfully transfected (Hagedorn *et al.*, 2011). However, the need for enrichment of stably transfected cells by antibiotic selection renders the protocol less clinically applicable; although the use of antibiotics such as kanamycin has been approved for some clinical trials, it is unlikely that antibiotic selection will become a standard part of gene transfer methods due to the risks, however small, of horizontal gene transfer to endogenous flora or anaphylactic shock upon administration of vector due to antibiotics present in the plasmid preparation (Marie *et al.*, 2010). In addition, selection pressure may result in increased frequency of integration as it favours stably and strongly expressing cells. Therefore, a protocol not relying on antibiotic selection would be a safer and more robust alternative for human gene therapies.

In addition to the pEPI plasmids, another non-viral episomal system which has been reported to confer mitotically stable extrachromosomal transgene expression is the minicircle. A S/MAR –containing minicircle is practically devoid of all bacterial sequences which might result in its elimination from mammalian cells (Broll *et al.*, 2010). The S/MAR element is predicted to provide a replication origin and attachment to the nuclear scaffolding, and thus mitotic retention. As the minicircles contain no bacterial sequences or antibiotic resistance markers, they are a safer alternative for long-term gene transfer.

The study by Broll et al into the long-term episomal retention of minicircles featuring a S/MAR element unearthed a naturally occurring mutant of the original β-interferon S/MAR. This element, termed miniMAR in this thesis, is a result of recombination within the original element and only 733 bp long. Incorporated into a derived minicircle vector instead of the full-length S/MAR, it appeared to confer equally strong episome retention as the original element, without the recombination and integration propensities of its parent element. In the study in question, initial selection pressure by the antibiotic G418 was used to select for the successfully transfected population (Broll *et al.*, 2010). No further studies have been published using the truncated miniMAR element, even though the reduction in size makes is more applicable for a number of vectors with limited capacity.

To date, the only S/MAR -based systems reported to successfully confer episomal retention are the pEPI plasmids and minicircles. This is a likely indication of the lack of success of alternative ventures, as more than a decade has elapsed since the first reports of the success of the S/MAR element by Piechaczek et al. The large size (~1.7kb) of the original S/MAR element excludes it from some viral vectors with limited packaging capacity such as self-complementary AAV; therefore a new and equally functional element with reduced size would be particularly useful for many applications.

The Chinese hamster ovary (CHO) cell line was chosen for the experiments in this chapter to confirm reproducibility. Many of the publications in which the pEPI was found to replicate episomally have used CHO cells and therefore using the same cell line reduces the number of variables affecting the environment of the S/MAR element (Piechaczek *et al.*, 1999).

3.1.2 Aims and objectives of the chapter

In this chapter, the hypothesis that the S/MAR element assists in the episomal establishment of non-viral vectors was investigated. Although this hypothesis has been supported by results published by other groups (see Introduction, section 1.5.3.1), no data has been published on whether selection or sorting is necessary to enrich successfully transfected cells. Therefore, the key objective here was to investigate the behaviour of S/MAR elements without selection pressure. Instead of selection, the cells were subjected to an initial period of induced cell cycle arrest, to find out whether an

extended presence of the plasmid vector in the nuclear environment assists in episome establishment.

The first question addressed in this chapter is whether the S/MAR element contained in the pEPI plasmid vector system is capable of establishing mitotically stable episomes when not subjected to antibiotic selection. To this end, two pEPI vectors reported to have successfully become mitotically stable under initial selection pressure were selected for *in vitro* transfection studies and monitored for transgene expression with no selection applied.

The second subject of investigation centres on the need of the S/MAR element to interact with components of the nuclear matrix in order to confer mitotic stability. It was postulated that in order for the interactions to take place, the element would need some time in the nuclear environment, prior to being dispersed by cell division. To achieve this, the cells were induced to undergo a period of enforced cell cycle arrest following introduction of the vectors to the cells. If the extended time period in the nucleus is beneficial for the establishment of the episomes, the protocol should result in a higher proportion of the vectors becoming mitotically stable.

As a mild and reproducible and simple method to induce cell cycle arrest, the effect of methionine depletion on CHO cells was examined, to see whether withdrawal of the majority of methionine from the culture media would indeed be sufficient to encourage the cells to enter cell cycle arrest.

The next set of experiments was designed to answer the question of whether the same S/MAR element behaves differently when provided on different plasmid backgrounds.
As the miniMAR element has previously only been studied as part of a minicircle vector, it was included in the experiments in the pEPI backbone. If the element retains the replication origin qualities of its parent, it should be able to confer mitotic retention as well as the parent element. Placing it in a different backbone also answers the question of whether the S/MAR elements are affected by neighbouring sequences *in cis* and therefore behave differently depending on the backbone.

To assess the capacity of S/MAR elements to provide episomal retention in vector systems other than the pEPI, their effect on the retention of viral transfer plasmids was also investigated. This was also done to see whether plasmids similar in size and content

to the pEPI plasmids would exhibit similar retention levels as the pEPI; if the S/MAR element is the only sequence necessary for episome retention then this should be the case.

Only the truncated miniMAR element was incorporated into the cassette of self-complementary AAV (scAAV) and lentiviral transfer plasmids, since the packaging capacity of scAAV is very limiting. Although the lentiviral cassette packaging capacity is more permissive, the plasmids were kept as similar as possible for comparison purposes. The transfer plasmids were constructed in view of using them at a later date for making viral vectors, and therefore the lengths were kept under the known packaging capacities. The non-miniMAR containing viral transfer plasmids were included in the study as controls.

Lastly, an experiment was devised to determine whether, in fact, an induced period of cell cycle arrest following transfection may be enough to allow any circular strands of DNA present in the nucleus to become established as stable episomes. To this end, a minicircle containing only the CMV GFP expression cassette in the absence of S/MAR elements was transfected into cells and subjected to an enforced period of cell cycle arrest to accommodate any spontaneous interactions with the nuclear matrix.

3.2 Results

3.2.1 Computational duplex destabilisation analysis of S/MAR-containing plasmid vectors

The pEPI plasmid incorporating the full-length β-IFN S/MAR element has been reported to exhibit mitotic stability and its daughter plasmid containing a naturally occurring truncated mutant miniMAR has also been shown to be retained in replicating cells (Broll *et al.*, 2010). The SIDD profiles of both plasmids show a pattern typical for S/MAR elements, a destabilised region with periodic minima in the energy required for strand separation (Figure 3.1, A and B).

The viral transfer plasmids were analysed for any changes in the destabilisation profiles following the introduction of the miniMAR element. In both the AAV and lentiviral transfer plasmids, the addition of the S/MAR element results in a strongly destabilised S/MAR region at the cost of the rest of the plasmid becoming more stabilised. In the AAV transfer plasmid, the effect is less dramatic but can be observed in the promoter and polyA regions of the antibiotic resistance gene (Figure 3.1, C and D), where the energy required for strand separation (G(x)) is increased by approximately 2 kcal/mol for both elements.

The lentiviral transfer plasmid profile is more prominently affected than the AAV plasmid. The effect is especially strong in the region between the ampicillin resistance gene and the bacterial origin of replication, where the energy required for strand separation is increased from <1 to ~5 kcal/mol. The region containing the Gag gene, primer binding site (PBS) and Rev responsive element can also be seen to become more stabilised, and G(x) at the central polypurine tract (cPPT) is increased from 4 to 6.5 kcal/mol (Figure 3.1, E and F). However, the CMV promoter and the GFP gene are unaffected by the proximity of the miniMAR element, which suggests equal levels of transgene expression can be expected from plasmids containing and omitting the miniMAR.

SIDD analysis of the minicircle could not be conducted as the full sequence information was not disclosed by PlasmidFactory, who supplied the vector. However, as no S/MAR elements are included in the vector, the unavailability of this data was not considered to be crucial.

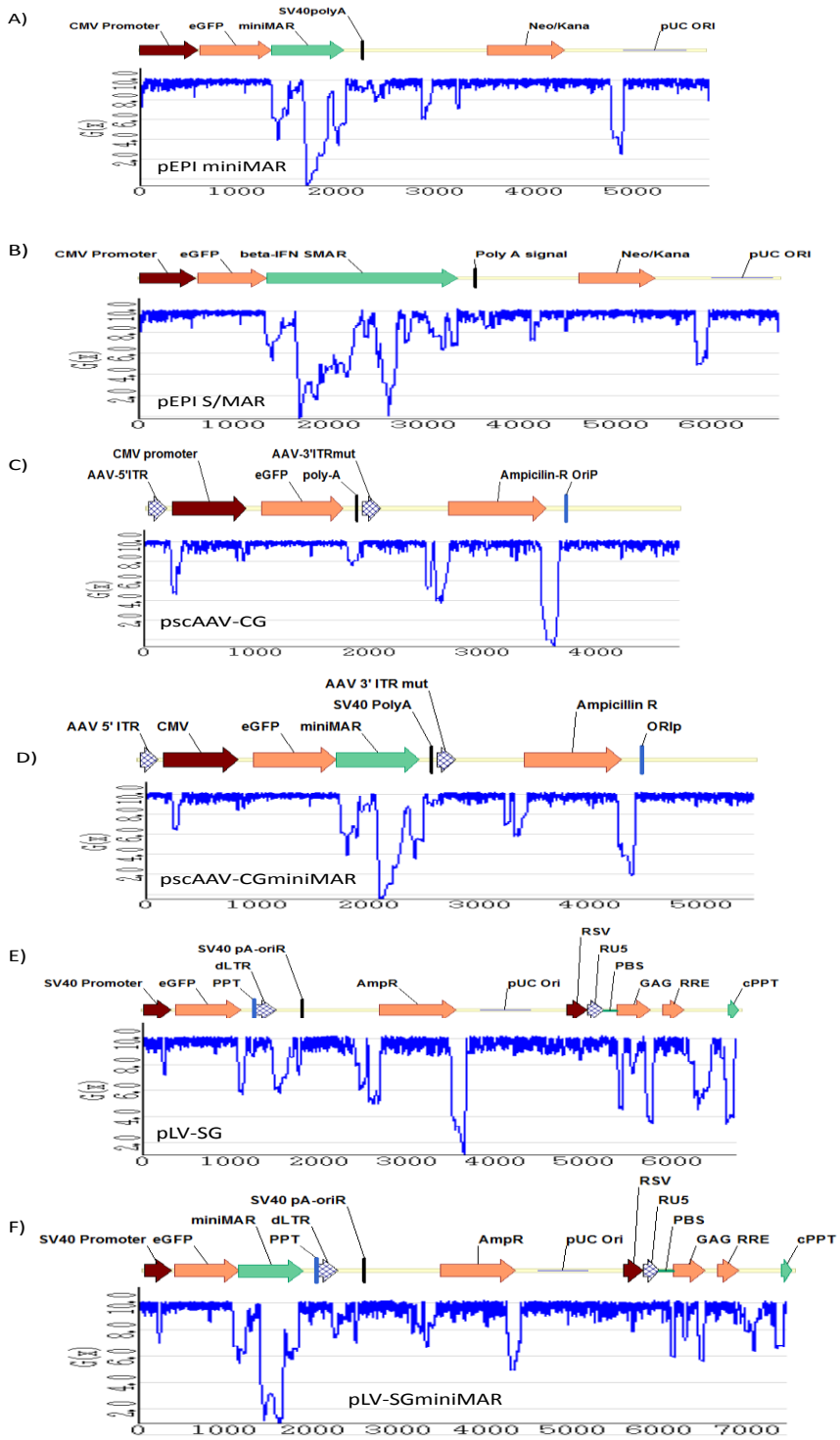


Figure 3.1 SIDD analysis of plasmid vectors.

pEPI plasmid vectors incorporating the miniMAR (A) and full-length S/MAR (B) elements each contain a destabilised region with periodic energy minima typical of S/MAR elements. The addition of the miniMAR element into the AAV transfer plasmid pscAAV-CGm (D) and lentiviral transfer plasmid pLV-SGm (F) result in condensation of destabilised regions when compared with the transfer plasmids without S/MAR elements (pscAAV-CG, C) and (pLV-SG, E). G(x), average free energy of all states x in which the base pair separates. X axis, length of construct in bp.

3.2.2 Optimisation of induced cell cycle arrest for plasmid transfections

All transfection experiments included a set of cells which were induced to undergo a period of cell cycle arrest achieved through methionine and serum depletion, and a control set allowed to proliferate continuously. To ascertain that comparable effects were achieved each time, the cells were analysed for cell cycle phase by a propidium iodide stain followed by FlowJo cell cycle analysis of the data. The data is presented in figure 3.2 as percentages of the cells in each cell cycle phase at the end of the cell cycle arrest. In statistical analyses, Student's t-test was used to compare the proportions of cells in each cell cycle phase between treatments.

In the pEPI plasmid experiment (Fig 3.2 A), all cell cycle phases were significantly different between the two treatments, with $P < 0.001$ for G_1, S and G_2 phases. Similarly, in the viral transfer plasmid experiment (Fig, 3.2 B) all cell cycle phases were significantly different between the quiescent and non-quiescent treatments; G_1 phase, $P = 0.0018$, S phase, $P < 0.001$, and G_2 phase, $P = 0.0002$.

The minicircle experiment encountered high levels of cell death following transfection, and a trend towards cell cycle arrest can be observed from the percentages of cells in each cell cycle phase at the end of cell cycle arrest at Day 6 (Figure 3.2 C). The difference between treatments is statistically significant for the proportion of cells in S-phase, $P < 0.05$, and G_2 phase, $P < 0.001$.

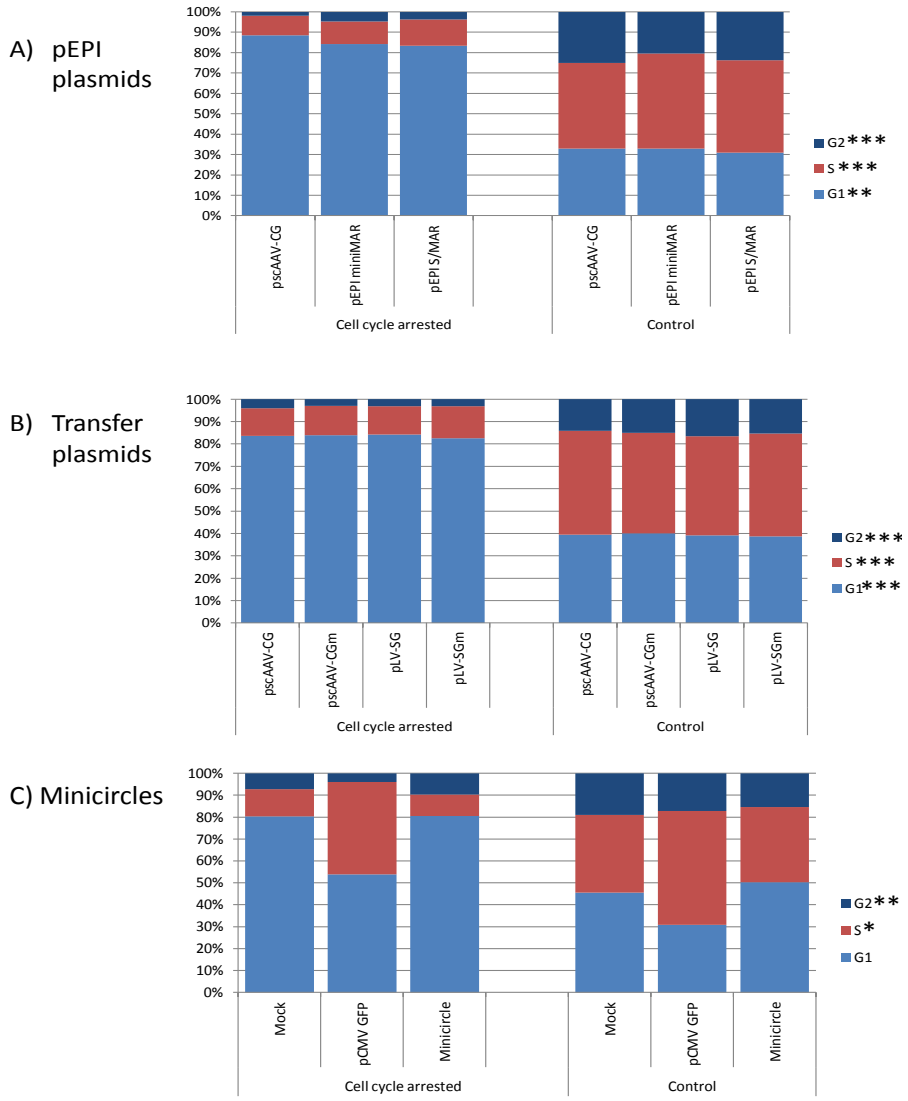


Figure 3.2 Cell cycle analysis by propidium iodide staining and flow cytometry of plasmid and minicircle -transfected CHO cells with and without induction of cell cycle arrest by methionine and serum depletion of culture media.

CHO cells were plated, transfected with plasmids, minicircle or mock control after 1 day, and then treated with methionine-free medium from D2 to D6, over a 4 day period. At D6, cultures were returned to normal growth medium and allowed to proliferate as normal with routine passaging as required. To confirm induction of transient cell cycle arrest, cell cycle phase analyses were performed by propidium iodide staining and flow cytometry at D6, at the end of the methionine depletion period. The coloured sections indicate the percentages of cells in each cell cycle phase as determined by PI staining followed by FlowJo analysis of the data.

*The cell cycle phases determined to differ significantly between treatments are denoted with *, as indicated by Student's t-test comparing each cell cycle phase separately between arrested and control at D6. * = P< 0.05, **= P< 0.01, ***= P< 0.001. n=3.*

3.2.3 Optimisation of plasmid transfection conditions in CHO cells

To conduct the planned plasmid transfection experiments, the optimum cell density and DNA concentration was determined so that maximum transfection efficiency would be achieved.

CHO cells were seeded at 3 different densities from $1x10^5$ to $5x10^5$ and transfected with 3 different amounts of DNA from 2 to 8 µg /well (Figure 3.3). The two lower cell densities produced > 92 % successfully transfected cells with 2 µg plasmid, whereas cells seeded at the higher density achieved only 55 % positive cells with 4 µg plasmid. Transfection with 2µg DNA gave overall the best pattern of GFP expression across the cell density range. Therefore the lowest cell density and lowest DNA concentration was used in the transfection experiments to maximise transfection efficiency whilst conserving materials.

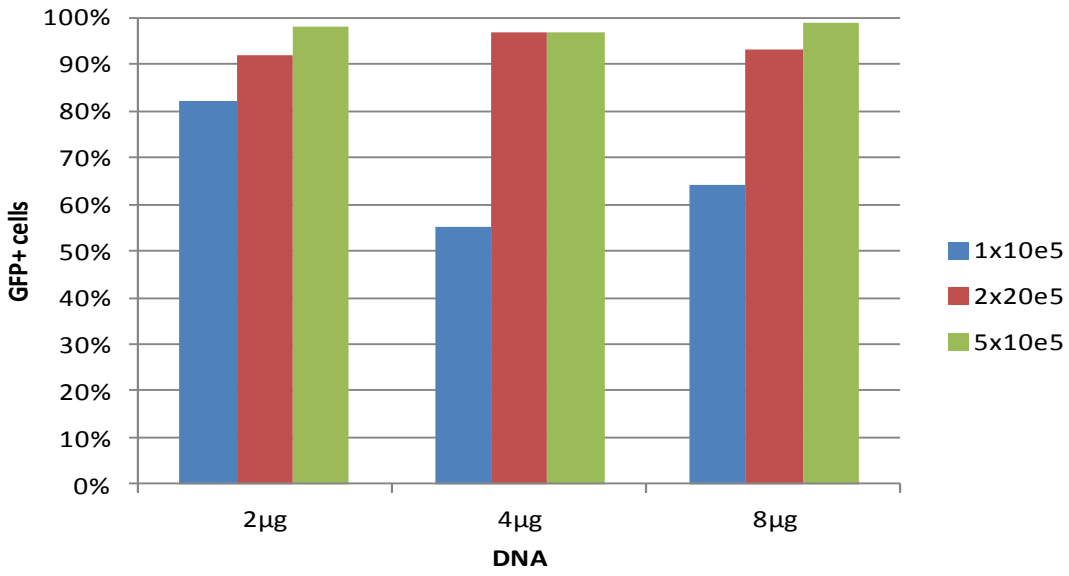


Figure 3.3 Optimisation of CHO cell transfection.

CHO cells were seeded at densities of $1x10^5$, $2x\ 10^5$ and $5x10^5$ CHO cells /well on 6-well plates and transfected with 2, 4 and 8 µg of pscAAV-CG DNA/well using 2:1 DNA: Lipofectamine™ 2000 ratio. The proportion of GFP-expressing cells was determined by flow cytometry 48 hours post-transfection. Transfection of CHO cells with 2 µg of DNA at a density of $2x10^5$ or $5x10^5$ cells/well produced > 92 % GFP+ positive cells. The lower cell density of $1x10^5$ resulted in 55 – 82 % transfection efficiency. FACScan analysis threshold set to 10^4 cells/ condition.

3.2.4 Effect of S/MAR element and induced cell cycle arrest on plasmid stability in transfected CHO cells

3.2.4.1. Comparison of pEPI S/MAR and pEPI miniMAR

To evaluate whether S/MAR and miniMAR plasmids would support episomal replication and stability, CHO cells were transfected with the pEPI plasmids containing the full-length and the truncated version of the S/MAR element. Plasmid pscAAV-CG containing the CMV-GFP expression cassette was included as a control. For each plasmid, two sets of cells were transfected simultaneously; one set was induced to undergo cell cycle arrest by methionine depletion for 4 days post-transfection and the other was kept in continuously proliferating culture. Initial transfection efficiency was determined by flow cytometry 24 h post-transfection, and varied between 55 - 60 % for pEPI-miniMAR, 73 -74 % for pEPI-S/MAR and 52 - 77 % for pscAAV-CG. The difference between the efficiencies is not statistically significant, $P = 0.8378$ (One-way ANOVA, n = 3).

The percentage of GFP+ cells in cultures not subject to induced cell cycle arrest declined steadily over the first 10 days (Figure 3.6 A), and the cells held quiescent showed a similar decline once they were released back into normal cycling (Figure 3.6 B). The proportion of GFP+ cells continued to decline in both populations reaching stable levels by 3 weeks post-transduction. The percentage of GFP+ cells stabilised at low levels in both culture conditions at 0.3 – 2.9 % above background after 30 days.

The percentage of GFP+ cells after 30 days varied depending on the vector used in both treatment groups. The highest level of GFP expression was achieved by pEPI S/MAR; the proportion of stably expressing cells was 2.6 % above background in cells after an initial period of cell cycle arrest and 1.9 % in cells cultured continuously. In comparison, pEPI miniMAR resulted in 1.0 % positive cells with cell cycle arrest and 0.8 % without cell cycle arrest, and the control plasmid pscAAV-CG resulted in 1.3 % and 1.6 % GFP positive cells with and without cell cycle arrest, respectively (Figure 3.4, C and D).

Both pEPI plasmids exhibited a tendency for increased transgene retention following an initial period of cell cycle arrest. In contrast, the proportion of GFP-expressing cells for the control plasmid not containing a S/MAR element was not increased by the cell cycle

arrest and was actually 0.3% higher after being kept in continuously proliferating culture.

Statistical analyses indicate that the differences between vectors are significant. Within the groups of transfected cells subjected to an initial period of cell cycle arrest, %GFP+ in cells transfected with pEPI S/MAR is significantly different to that in cells transduced with the other 2 vectors (one-way ANOVA $P < 0.0001$, Tukey's test for pair-wise comparison, $P < 0.001$ for both comparisons). Similarly, within the groups of transfected cells kept in continuously proliferating culture there is a significant difference between cells transfected with pEPI S/MAR and pEPI miniMAR (One-way ANOVA followed by Tukey's post-test, $P <0.0001$).

The difference between the treatment groups is also statistically significant for cells transfected with pEPI S/MAR (Tukey's test, $P < 0.0001$). The difference between treatments was not significant for the other 2 vectors. For all statistical analyses, the GFP percentages were included from all triplicates from the last 2 measurements at 26 and 30 days post-transfection, n = 6.

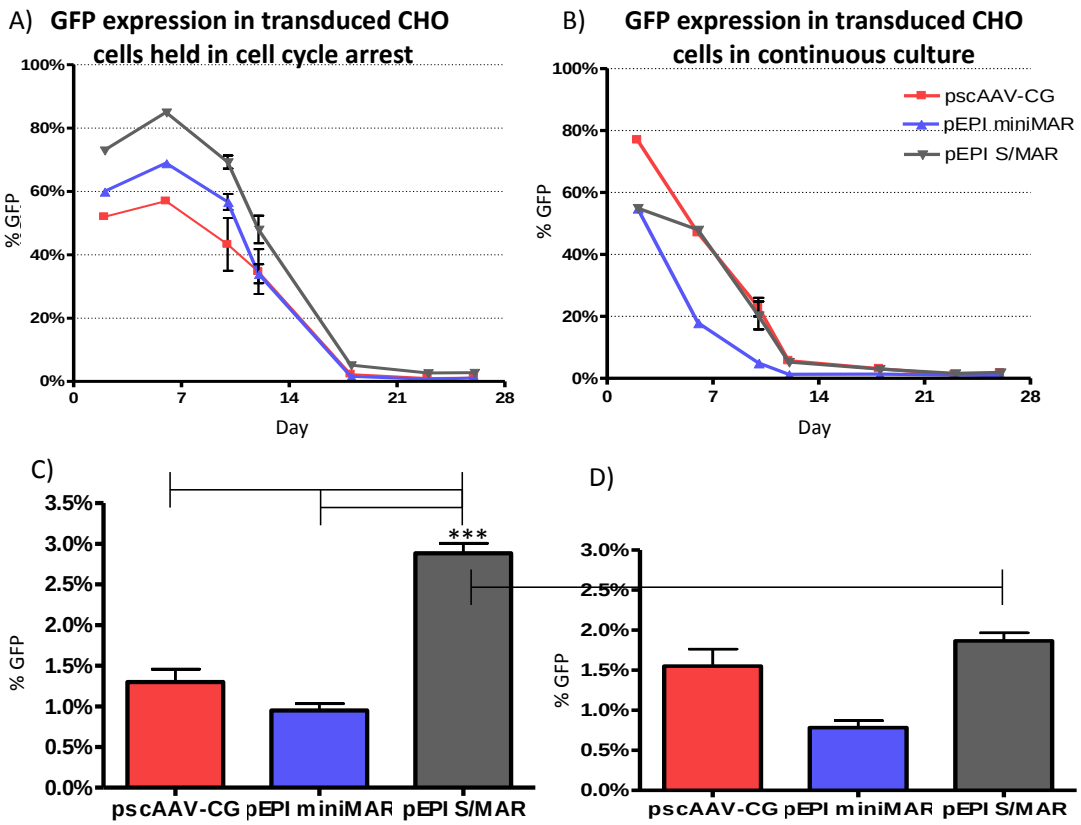


Figure 3.4 The % GFP+ cells is maintained at a higher level in cultured CHO cells after transfection with pEPI plasmids containing the full-length S/MAR.

*CHO cells were plated into culture, and after 1 day transfected with 2 µg pscAAV-CG, pEPI S/MAR or pEPI miniMAR. Cells were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media at D2 (A and C) or allowed to proliferate freely (B and D). From D6, all cells were allowed to proliferate normally with routine passaging for a period up to D28. In (C) and (D) the stabilised transduction levels at the D28 time point are shown. Data are means ± s.e.mean, n=3. The proportion of stably GFP+ cells was significantly different when transfected with pEPI containing the full-length S/MAR element (P <0.001, denoted with ***, as analysed by 1-way ANOVA and Tukey's post-test (n = 6, triplicates at 2 time points /group). The mock population was set at 0.5%.*

3.2.4.2 Viral transfer plasmids

To investigate the episomal retention characteristics of the viral transfer plasmids, 4 plasmids were used to transfect CHO cells *in vitro*: the self-complementary AAV transfer plasmids with and without the miniMAR element, and the lentiviral transfer plasmids with and without the miniMAR element.

- pscAAV-CMV-GFP (pACG)
- pscAAV-CMV-GFP-mMAR (pACGm)
- pLV-SV40-GFP (pLSG)
- pLV-SV40-GFP-mMAR (pLSGm)

All plasmids contain the GFP gene; in the AAV plasmids it is driven by the CMV promoter and in the lentiviral plasmids it is driven by the SV40 promoter (See Materials & methods, section 2.4.1.2).

Two sets of CHO cells were transfected simultaneously, one of which was held in cell cycle arrest for 4 days following transfection whereas the other one was kept in continuously proliferating culture. The expression of GFP was monitored by flow cytometry.

Transfection efficiency was measured by GFP expression levels 24 h after transfection. It was slightly lower for pLSGm (36 - 44 %) and pACGm (41 - 49 %) and higher for the equivalent plasmids without the miniMAR elements, at 48 - 55 % for pLSG and 46 - 59% for pACG. However, the difference is not statistically significant as shown by One-way ANOVA, $P = 0.4721$ (n =3).

The cells not subject to induced cell cycle arrest after transfection displayed a steady decline in GFP expression over 10 days to a low, stable level (Figure 3.5 B). The cells kept quiescent for 4 days exhibited a delayed kinetics (Figure 3.5 A) whereby the GFP expression stayed high during cell cycle arrest but declined within 10 days following return to permissive proliferating culture conditions. In both cases, the expression stabilised at levels between 0.1 % and 1 % above background by day 35.

There were no significant differences in the GFP expression levels between different plasmids or treatments as demonstrated by One-way ANOVA. The inclusion of the miniMAR element did not increase the retention levels of either the AAV or the lentiviral transfer plasmid. In fact, when the cells were held in an initial period of cell cycle arrest, both the miniMAR-containing plasmids showed a tendency towards decreased episomal retention compared to the equivalent plasmid without the miniMAR element. Upon introduction of the miniMAR, the GFP positive percentage decreased from 0.9 % above background to 0.4 % in the AAV transfer plasmid, and from 1.1% to 0.9 % in the lentiviral transfer plasmid.

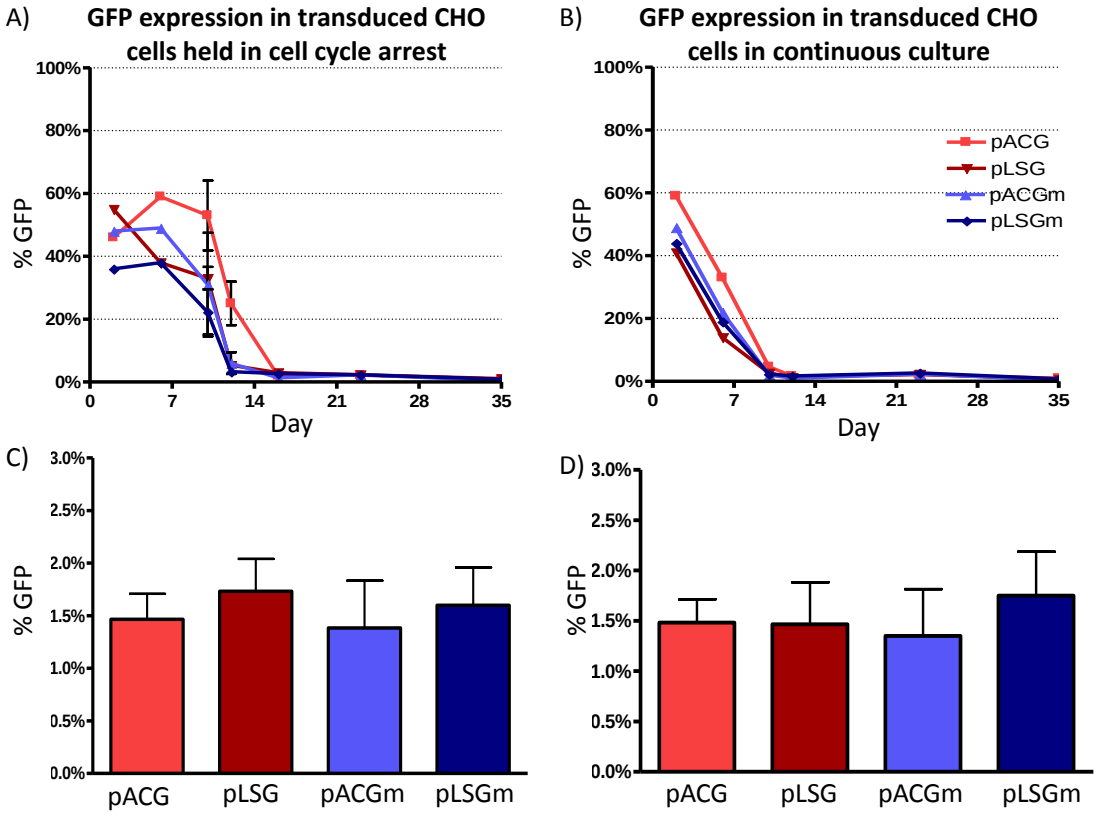


Figure 3.5 The miniMAR element does not increase the retention of GFP expression in cultured CHO cells after transfection with transfer plasmids.

CHO cells were plated into culture, and after 1 day transfected with 2 µg lentiviral (pLV-SG and pLV-SGm) and AAV (pACG and pACGm) transfer plasmids. Cells were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media at D2 (A and C) or allowed to proliferate freely (B and D). From D6, all cells were allowed to proliferate normally with routine passaging for a period up to D35. In (C) and (D) the stabilised transduction levels at the D35 time point are shown. Data are means ± s.e.mean, n=3. The mock population was set at 0.5%. The percentages of GFP expressing cells after 23-32 days in culture do not differ significantly between vectors or treatments (One-way ANOVA).

3.2.4.3 Minicircles

CHO cells were transfected with equimolar concentrations of minicircle and a control CMV-driven GFP -expressing plasmid. The transfection efficiency was measured by flow cytometry for GFP expression 2 days post-transfection and varied between 73-83% for the minicircle and 67-79% for the control plasmid. The difference is not statistically significant as shown by t-test, $P = 0.55$, n = 3.

Again, cell cycle arrest induced by methionine depletion resulted in delayed loss-of-expression kinetics. The proportion of GFP-expressing cells declined steadily following release into proliferating culture and reached levels 0 – 0.9 % above background, measured at 27 and 32 days post-transfection. The cells kept in continuously proliferating culture were slightly less GFP positive by the same time points, containing 0 – 0.2% expressing cells (Figure 3.8, A and B). Comparing the percentages of GFP – expressing cells at the last two time points between treatments by one-way ANOVA, there is a significant difference between the minicircle –transfected cells subjected to an initial period of cell cycle arrest and both vectors in continuously proliferating culture ($P < 0.05$ by Tukey's post-test, n = 6). However, the expressing population of minicircle-transfected cells is still very low at 0.5 % above background (Figure 3.6, C and D).

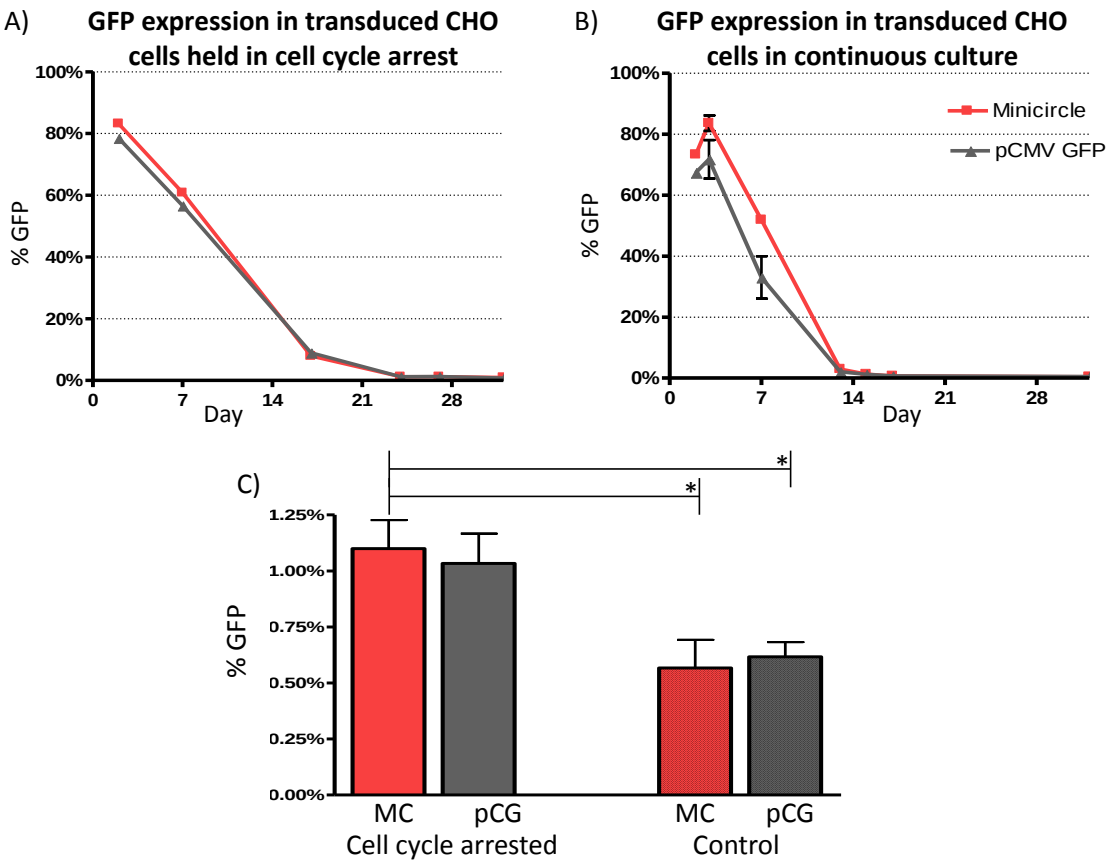


Figure 3.6 Induced cell cycle arrest increases minicircle retention levels in transfected CHO cells.

*CHO cells were plated into culture, and after 1 day transfected with equimolar concentrations of minicircle vector and pCMV-GFP. Cells were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media at D2 (A and C) or allowed to proliferate freely (B and D). From D6, all cells were allowed to proliferate normally with routine passaging for a period up to D32. In (C) and (D) the stabilised transduction levels at the D32 time point are shown. Data are means ± s.e.mean, n=3. The mock population was set at 0.5%. Minicircle transfection followed by an initial period of cell cycle arrest results in a significantly higher level of transgene retention than either minicircle or plasmid transfection without cell cycle arrest (denoted with * in C, P <0.05 by one-way ANOVA followed by Tukey's post-test). MC: minicircle, pCG: pCMV GFP.*

3.3 Discussion

3.3.1 SIDD analysis of S/MAR-containing pEPI plasmids and viral transfer plasmids

Analysis of strand separation potential in the pEPI plasmids was used to confirm the presence and functionality of the full-length and the truncated S/MAR elements. Both were shown to result in periodic minima in the energy required for strand separation across the length of the element, which suggests the S/MAR elements should be functional.

The viral transfer plasmids were analysed for any changes in the destabilisation profiles following the introduction of the miniMAR element. In pscAAV-CG plasmid, a small condensing effect was observed upon introduction of the miniMAR. This was evident mainly in the promoter and polyA signal regions, especially in the plasmid backbone (Figure 3.1 C and D). In the lentiviral pLV-SG plasmid, a similar effect was observed, although in this case the effect was stronger (Figure 3.1 E and F). The condensation of this region can be posited to affect both the polyadenylation site of the ampicillin gene and the replication origin, both of which elements require strand separation to carry out their function. However, as in both cases the effect mainly concerns the ampicillin gene, there should be no effect on the *in vitro* transduction experiments where antibiotic resistance is not required. Similarly, although the bacterial origin of replication is also somewhat condensed, this should not affect the replication capacity of the vectors as the S/MAR element is postulated to function as an origin of replication in mammalian cells.

3.3.2 CHO cells can be induced to undergo cell cycle arrest by methionine and serum depletion

24 h after each transfection, a proportion of the cells were induced to undergo cell cycle arrest to enable putative interactions between the nuclear matrix and S/MAR and other elements in the vectors. The proportion of cells in each cell cycle phase was monitored before, during and after cell cycle arrest as well as in the control population in

continuously proliferating culture to ascertain that similar conditions were achieved in each experiment.

In each experiment, placing CHO cells in serum-depleted, methionine free media caused a significant shift in the cell cycle phase towards fewer cells in S and G_2 phase and more cells in G_1 phase (Figure 3.2), which is suggestive of potential cell cycle arrest at the G_1/S checkpoint. The difference between the two treatments was statistically significant in all transfection experiments, confirming the cells were successfully held back from normal cycling.

3.3.3 Effect of S/MAR elements and induced cell cycle arrest on retention of transfected pEPI plasmid vectors

Although recent publications have indicated that the truncated miniMAR element incorporated into minicircles is able to achieve comparable episome retention to the full-length S/MAR (Broll *et al.*, 2010), in our system this does not appear to be the case. There is a significant difference between the pEPI containing the full-length S/MAR and the truncated miniMAR in both treatment groups; although not significantly different, the pEPI miniMAR also results in fewer stably transduced cells than the control plasmid pscAAV-CG in both treatment groups. This suggests that the miniMAR confers no positive effect on the pEPI plasmid.

In contrast, the proportion of stably GFP-expressing cells achieved by transfection of the pEPI plasmid vector containing the full-length S/MAR element was a remarkable 2.5% - 3.3%., considering no selection was applied at any stage. There is a statistically significant difference in plasmid retention between the pEPI vector incorporating the full-length S/MAR element, compared to cells tranfected with all other plasmids in the two experiments (P <0.001).

It is unlikely that the difference between the populations transfected with the pEPI plasmid and either subjected to cell cycle arrest or kept in continuous culture is simply due to the difference in the number of population doublings; by 30 days post-transduction, the cell cycle arrested population has only undergone 8 fewer population doublings than the control population (38 versus 45 doublings, based on the standard cycling time of 16h for CHO cells). If half of all the episomes were lost at each

doubling, it would only take 6 doublings or 4 days for the percentage of positive cells to drop from 100% to 1.6%, and both cell populations have been kept in continuous culture for at least 3 times longer.

As a control, the viral transfer plasmid pscAAV CMV GFP was included in both experiments, and there was no significant difference in the stably expressing GFP percentages for the same plasmid between the two experiments (P = 0.39), suggesting the results from the two experiments are comparable.

The episomal status of the pEPI S/MAR stably expressing in cells remains to be confirmed. Due to the low percentage of the expressing cells, it may be necessary to obtain clonal populations in order to perform conclusive integration analysis. Since integration can be expected to occur in 0.5 %-1 % of transfected cells in long-term culture (Broll *et al.*, 2010), the 2-fold higher percentage suggests either an increased rate of integration due to the presence of the S/MAR element, or increased episome retention.

Previous work on S/MAR containing plasmids suggests they are retained in very low copy numbers of 2-10 copies per cell (Piechaczek *et al.*, 1999). However, we cannot currently rule out the possibility that some of the non-expressing cells contain transcriptionally silenced plasmid episomes, so that expression levels may not necessarily correlate with copy numbers. Episome copy number analysis and integration status analyses could potentially be carried out by dilution cloning and expansion of GFP-positive cell lines; analysis on a population only 1-3% GFP positive is beyond the detection capacity of most techniques. Especially the pEPI –transduced population expressing at 3% would be of interest to study further.

It is remarkable that with transient cell cycle arrest a significant level of pEPI S/MAR retention was achieved without any selection pressure, unlike previous published experiments with this plasmid where selection was applied at least initially (Piechaczek *et al.*, 1999). Previous reports indicate that in the absence of selection pressure, less than 1% of transfected cells retain transgene expression after 1 month (Papapetrou *et al.*, 2006). The lack of selection pressure could also go some way towards explaining the lack of retention provided by the miniMAR-element, as the study where miniMAR-mediated episome establishment was successful used antibiotic selection and flow cytometry to obtain stably expressing clonal populations (Broll *et al.*, 2010). Although the level of stable GFP expression achieved with pEPI S/MAR was a fairly low ~3%,

the lack of need for antibiotic selection could provide opportunities for the development of clinically significant therapies.

3.3.4 Effect of miniMAR elements and induced cell cycle arrest on retention of transfected viral transfer plasmids

The introduction of the miniMAR element did not increase the transgene retention levels when combined with either lentiviral or AAV transfer plasmids. Indeed, when introduced into the viral vector cassettes, both the transfection efficiencies and levels of stably expressing cells were slightly, although not significantly, lower.

As the pEPI vectors have been designed for gene delivery purposes, it is not surprising they performed better than viral transfer plasmids designed for making viral gene delivery vectors. Although it was interesting to test the effect of the addition of the miniMAR element on retention of the viral transfer plasmids, this was mainly done to test the functionality of the cassette and the lack of retention was not particularly surprising. In both cases, the presence of bacterial sequences in these plasmids is likely to have impaired their ability to become established as episomes, as prokaryotic sequences have been shown to lead to shutdown of transgene expression and vector loss (Chen *et al.*, 2008).

The importance for this data will become apparent in the following chapters as the viral vectors made using these transfer cassettes are tested for episomal retention. Comparing the plasmid vector data with the viral vector data answers the question of whether the transgene cassette and the miniMAR element themselves are responsible for episome retention, or whether the delivery method and the remaining *cis*-acting sequences are also important.

3.3.5 Effect of induced cell cycle arrest on retention of transfected minicircles

The minicircle experiment was done in order to ascertain whether induced cell cycle arrest alone is enough to establish circular DNA in the nucleus as a mitotically stable, replicating episome, without the need for any additional genetic elements. If this is the case, then the minicircles devoid of any bacterial elements should be favoured over the

pCMV-GFP plasmids which do contain a bacterial backbone leading either to loss of episome or loss of expression through heterochromatin formation.

Episomal retention of minicircles has been reported *in vitro* with minicircles containing a S/MAR element (Broll *et al.*, 2010); here, a very small but significant effect on episome retention is achieved through an initial period of cell cycle arrest. This suggests that a period of cell cycle arrest following introduction of exogenous DNA is conducive to the establishment of the episome within the nuclear structure. Although the effect observed here is not significant enough to be of use in the clinic, it presents a possibility to further explore the effect of cell cycle arrest on a minicircle containing an origin of replication region, which would enable mitotic stability. In the study by Broll et al, the minicircle contained the human β-IFN S/MAR element. Combining this element with an initial period of cell cycle arrest or exploring the episome retention potential of other S/MAR or Ori elements may present an interesting subject for further study.

Chapter 4: Effect of S/MAR elements and induced cell cycle arrest on retention of AAV vectors in CHO cells

4.1 Introduction

Although the ability of AAV vectors to form episomal circular structures is well documented, they are stable only in post-mitotic tissues and tend to disappear quickly in dividing cells. Here, we wanted to explore possibilities for inducing interactions between components of the concatenated or circular AAV episomes and the nuclear matrix, resulting in mitotically stable extrachromosomal transgene retention.

The tendency of single-stranded AAV vectors to form large, multimeric concatemers can be combined with the matrix attachment properties of the S/MAR elements to potentially create a circular "mini-chromosome". This would consist of several conjoined genomes, be anchored to the nuclear matrix at regular intervals by the S/MAR element and express the transgene from several loci. The possibility for creating such structures was investigated, as well as their mitotic stability and the strength and duration of transgene expression.

Self-complementary AAV vectors (scAAVs) have an advantage over the single-stranded versions in that there is no need for second strand synthesis, bypassing the main rate-limiting step during transduction as well as preventing degradation of single-stranded DNA by cellular machinery. As scAAVs has been shown to result in the presence of circular monomeric or dimeric structures in the nucleus within 24 h of transduction, hypothetically this would also allow any S/MAR element present to interact with the nuclear matrix soon after viral entry. Here, it was investigated whether the early appearance of dsDNA structures from scAAVs may allow the necessary binding interactions to take place prior to cell division, resulting in the structures to become established in the nuclear environment.

In contrast with the single or double genome circles commonly resulting from scAAV transduction, ssAAVs usually form much larger concatemeric structures containing at

least 4 genomes (Straus *et al.*, 1976). If the number of S/MAR elements per extrachromosomal structure correlates with episome stability, this may confer an advantage to ssAAV since more S/MAR elements per episome would be present. However, no data as of yet is available on whether the retentive effect of the S/MAR elements is dependent on the quantity as well as the quality of the elements.

As well as the number of S/MAR elements present, the types of S/MAR elements incorporated into ssAAV and scAAV differ by necessity. As the packaging capacity of the scAAV is half that of ssAAV, the full-length β-Ifn S/MAR (~ 1.7 kb) could only be incorporated into the single-stranded AAV cassette. Even with ssAAVs the incorporation of the β-Ifn S/MAR would require that the transgene-promoter complex is kept relatively small, and thus restrict the applicability of the vector system somewhat when incorporating therapeutic transgenes. In our hands we thus decided to restrict the S/MAR element added into the scAAV cassette to the ~700 bp miniMAR element, and for comparison purposes this element was also investigated on the ssAAV platform.

The transgene expression and episome retention following AAV transduction was investigated *in vitro*. Although AAV transduction of cell lines in culture is not as efficient as *in vitro* transduction using some other vectors such as lentiviral vectors, it was a necessary step here as the transduction protocol was combined with an induced period of cell cycle arrest following transduction. The potential for such protocols to result in significant episome retention is best initially assessed in easily monitored culture conditions, with the aim of moving into *in vivo* investigations once proof of principle is established.

4.2 Results

4.2.1 Computational duplex destabilisation analysis of predicted AAV structures

To analyse the distribution of duplex destabilisation potential in each vector, computational models were constructed of the structures AAV is expected to form in transduced cells. Due to the 3-dimensional hairpin structures at the ITRs, it was not possible to express the DNA sequence entirely accurately in a conventional circular double stranded sequence format. Therefore, all hypothetical sequences were presented in the statistically most likely conformations and in a way that most closely resembles that of the predicted *in vivo* form.

4.2.1.1 SIDD analysis of wild-type AAV

SIDD analysis was first performed on the wild-type AAV2 genome for comparison purposes. The genome was analysed as a double-stranded, circular construct formed through non-homologous end joining, a form which is commonly found *in vivo* soon after transduction (Duan *et al.*, 1999a). According to the algorithm, two regions of the AAV2 sequence are significantly destabilised: one at 2116-2238 bp spanning the 3' proximal splice acceptor site, and the second corresponding to the polyadenylation site and the 3' ITR at 4374 – 4557 bp. There are 4 other sites where the free energy $G_{(x)}$ required for strand separation is less than 6 kcal/mol indicating a high probability for SIDD, although none of these correspond to known transcriptionally important genomic locations (Figure 4.1 A).

4.2.1.2 Construction & SIDD analysis of ssAAV vector structures

Concatemeric structures from recombinant ssAAV vectors containing the CMV enhancer / chicken β-actin promoter and GFP expression cassette were analysed for strand separation potential. Two different concatenated forms were constructed and analysed for each vector; dimeric, head-to-tail vector containing 2 copies of each ITR at both junctions which would result from non-homologous end joining (NHEJ), and a hybrid sequence that would result from homologous recombination between two ITRs.

The junction between the two genomes was created such that a single copy of each of the component sequences of the ITR hairpin loop (A', B', C' and D', see Chapter 1, Fig 1.2 A) remains, half of them being in the inverted (complementary) orientation. *In vivo*, this sequence would be able to form one complete hairpin structure flanked by AAV sequences on either side.

The destabilisation potential of genetic elements in the concatemer was found to be slightly altered between different potential conformations and rearrangements at the joined ITRs (Table 4.1). In both the form created by NHEJ and the form created by recombination between complementary sequences at the ITR, the promoter and enhancer regions are highly destabilised, with $G_{(x)}$ required for SIDD less than 2 kcal/mol (Table 4.1). The polyadenylation signal derived from SV40 is also destabilised with $G_{(x)}$ less than 6 kcal/mol. The difference between the two conformations lies within the number of potential SIDD sites detected by the algorithms; whereas in the structure with 2 complete ITRs there are 6 potential SIDD sites, in the recombined structure there are 10. However, the locations of SIDD are very similar between both molecules, and the graphical outputs for each molecule are almost indistinguishable.

Potential SIDD sites (cut-off: 6 kcal/mol)		
Dimer with 2 ITRs (NHEJ)	Dimer 1 ITR (HR)	Region
146-389 bp	146-389 bp	CMV enhancer
744-795 bp	743-795 bp	β-actin promoter
	1919-1932 bp	polyA signal
1947-1953 bp	1935-1961 bp	polyA signal
	1965-1986 bp	polyA signal
2409-2652 bp	2339-2582 bp	CMV enhancer
3007-3058 bp	2936-2988 bp	β-actin promoter
	4112-4125 bp	polyA signal
	4128-4154 bp	polyA signal
4210-4216 bp	4158-4179 bp	polyA signal

Table 4.1 Comparison of SIDD locations between two different hypothetical AAV conformations.

Dimeric AAV concatemers were created either by non-homologous end joining or homologous recombination between the ITRs of the standard ssAAV CAG GFP vector without S/MAR elements, and analysed for strand separation potential. The regions indicated for SIDD are highlighted in red in Fig 4.1 B.

The addition of a miniMAR element into the expression cassette drastically changes the SIDD potential of the ssAAV molecules. The central region of the miniMAR is calculated to require very little free energy input for strand separation to occur, with $G_{(x)}$ < 1 kcal/mol (Figure 4.1 C, green bar). Conversely, both the promoter and enhancer regions are condensed; $G_{(x)}$ required for SIDD increases from less than 2 in the concatemer without S/MAR elements, to ≥ 5 kcal/mol in the concatemer with the miniMAR. The polyA site is similarly condensed, with $G_{(x)}$ increasing from < 6 kcal/mol to > 8 kcal/mol.

The addition of the full length β-IFN S/MAR element into the cassette all but disables the capacity for strand separation of the promoter and enhancer elements, further increasing the free energy $G_{(x)}$ to > 7 kcal/mol for both elements. The polyA site becomes undistinguishable from neighbouring regions in the graphical SIDD output (Figure 4.1 D).

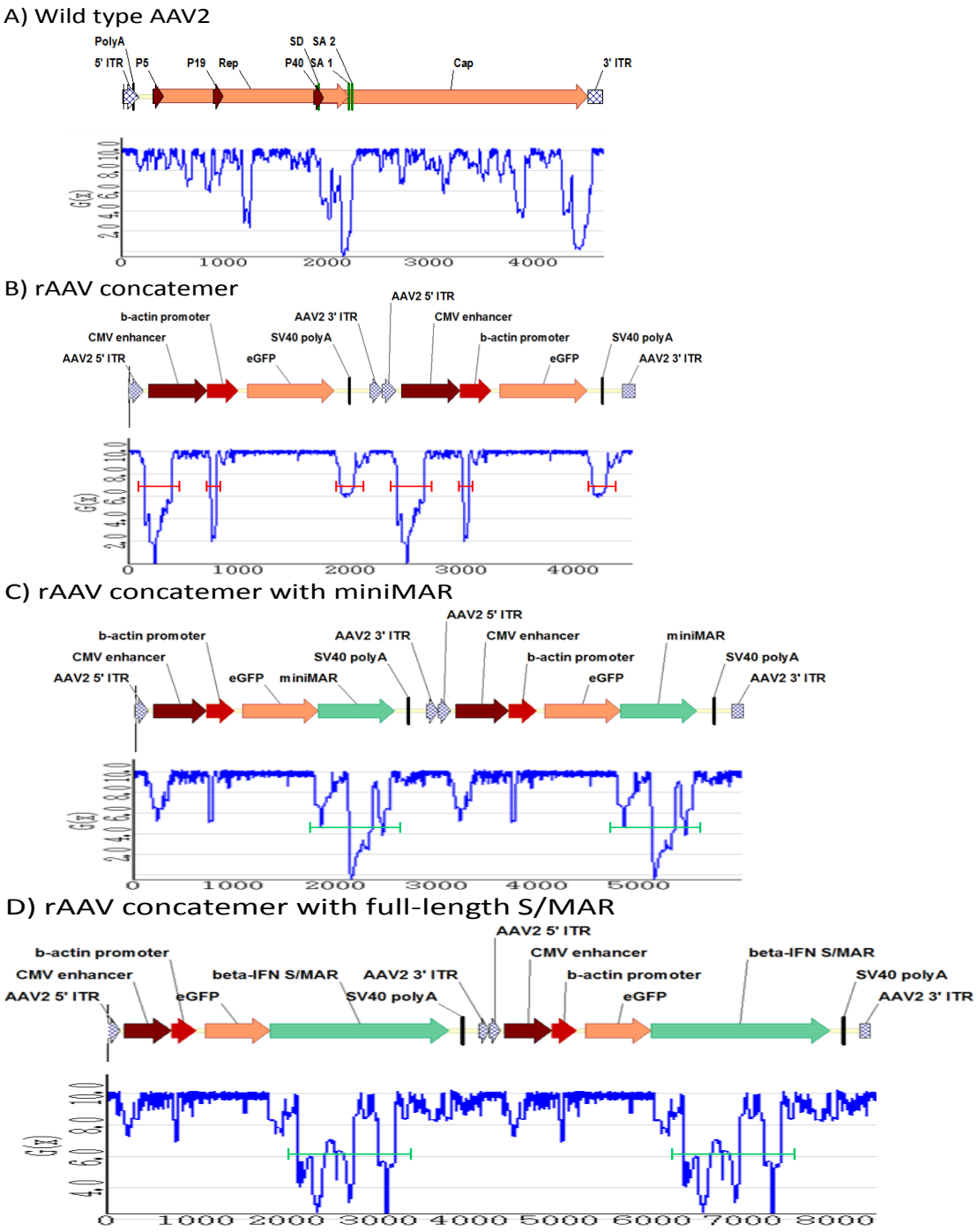


Figure 4.1 SIDD analyses of hypothetical concatenated structures generated by recombinant ssAAV vectors with and without S/MAR elements.

In wild-type AAV, the splice acceptor site and polyA signal are destabilised (A). Here, head-to-tail concatenated structures generated by ssAAVs after transduction were analysed for SIDD potential. In dimeric structures of AAV without S/MARs (B), the promoter, enhancer and polyA are destabilised (highlighted in red). In structures containing the miniMAR (C) and β-IFN S/MAR elements (D), the S/MAR regions are highly destabilised (green) resulting in condensation of elements in the rest of the domain.

4.2.1.3. Construction & SIDD analysis of scAAV vector structures

As the most common forms of scAAV genomes in the cells are predicted to be genomes containing single or double genomes, these configurations were used as a basis for SIDD analysis. Dimeric circles were constructed such that the closed ends (mutated ITRs) and open ends (wtITRs) of each double stranded genome were assumed to recombine with each other; this "head-to head dimer" has been suggested to be the most common *in vivo* conformation (Figure 4.2 A). Only the addition of the miniMAR element to the genomes was considered, as the packaging capacity cannot accommodate the larger full-length element.

In both the monomeric and dimeric circular scAAV molecules, the destabilised regions correspond to the CMV promoter and the SV40 polyA signal, with the promoter requiring less free energy for strand separation (Figure 4.2). The difference between the two different conformations is subtle. For the monomeric form $G_{(x)} < 1$ kcal/mol for the promoter and < 4 kcal/mol for the polyA, whereas in the dimeric form the $G_{(x)} < 2$ kcal/mol for both elements, indicating that the polyA experiences less competition with the promoter in the larger conformation. Again, the duplex stability increases dramatically in both conformations after the introduction of the miniMAR element. In the circular monomer, the energy required to separate the strands at the CMV promoter rises from <1 kcal/mol to >3 kcal/mol, and in the dimer the effect is amplified by the presence of two miniMAR elements such that $G_{(x)} > 6$ kcal/mol at the promoter.

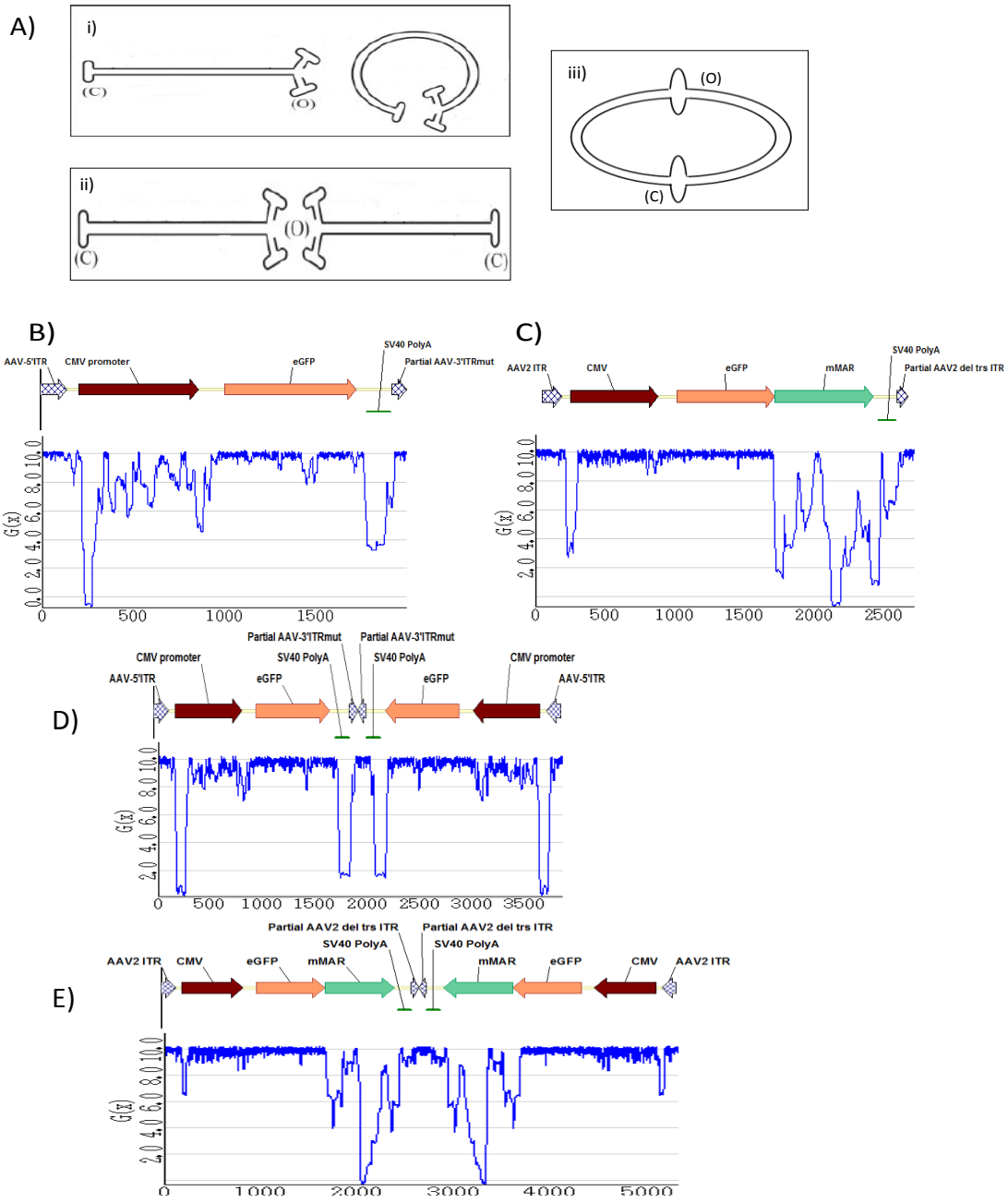


Figure 4.2 The S/MAR element has a profound effect on the destabilisation potential of hypothetical scAAV structures.

The palindromic scAAV genome folds back onto itself after entry into the cells, forming a double-stranded genome which can then combine to form a variety of monomeric and dimeric structures (i, ii and iii) (A). B) and C) represent the monomeric double-stranded genomes generated by the MAR-containing and non-MAR-containing scAAV, and the corresponding SIDD profiles. D) and E) represent the dimeric circular genomes with and without the MAR, and the corresponding SIDD profiles. In both C) and E), the addition of the miniMAR element can be seen to reduce the destabilisation potential of the polyA and promoter regions.

4.2.2 Optimisation of induced cell cycle arrest for ssAAV & scAAV transduction

As in Chapter 3, a proportion of the cells were induced to enter a reversible period of cell cycle arrest following AAV transductions. This was done as previously, using methionine and serum depletion.

The cell cycle phase analysis was performed using propidium iodide staining of the nuclei followed by flow cytometric analysis. The resulting data were analysed for cell cycle phase using the FlowJo software.

In both experiments, the methionine depleted media had a significant effect on the transduced cell population. After 5 days in methionine free serum depleted media, the proportion of cells in G_1 phase was significantly different to the cells in the proliferating control population, and both experiments show a similar change in the proportions of cells in each cell cycle phase (Figure 4.3, A and B). There was slight variation in the significance of the differences between the experiments; the difference between the G_1 phases was significant in the ssAAV experiment (P <0.05). In the scAAV transduction experiment, the difference in the proportion of cells in G_2 phase was significant (P <0.005).

Both data sets were analysed using Student's t-test comparing each cell cycle phase between the two different treatments. The significance of the differences between the treatments at the end of the quiescent period is shown for each phase in Figure 4.3. Cells containing an uncharacterised amount of DNA as defined by the cell cycle analysis software were excluded from the graphical representation.

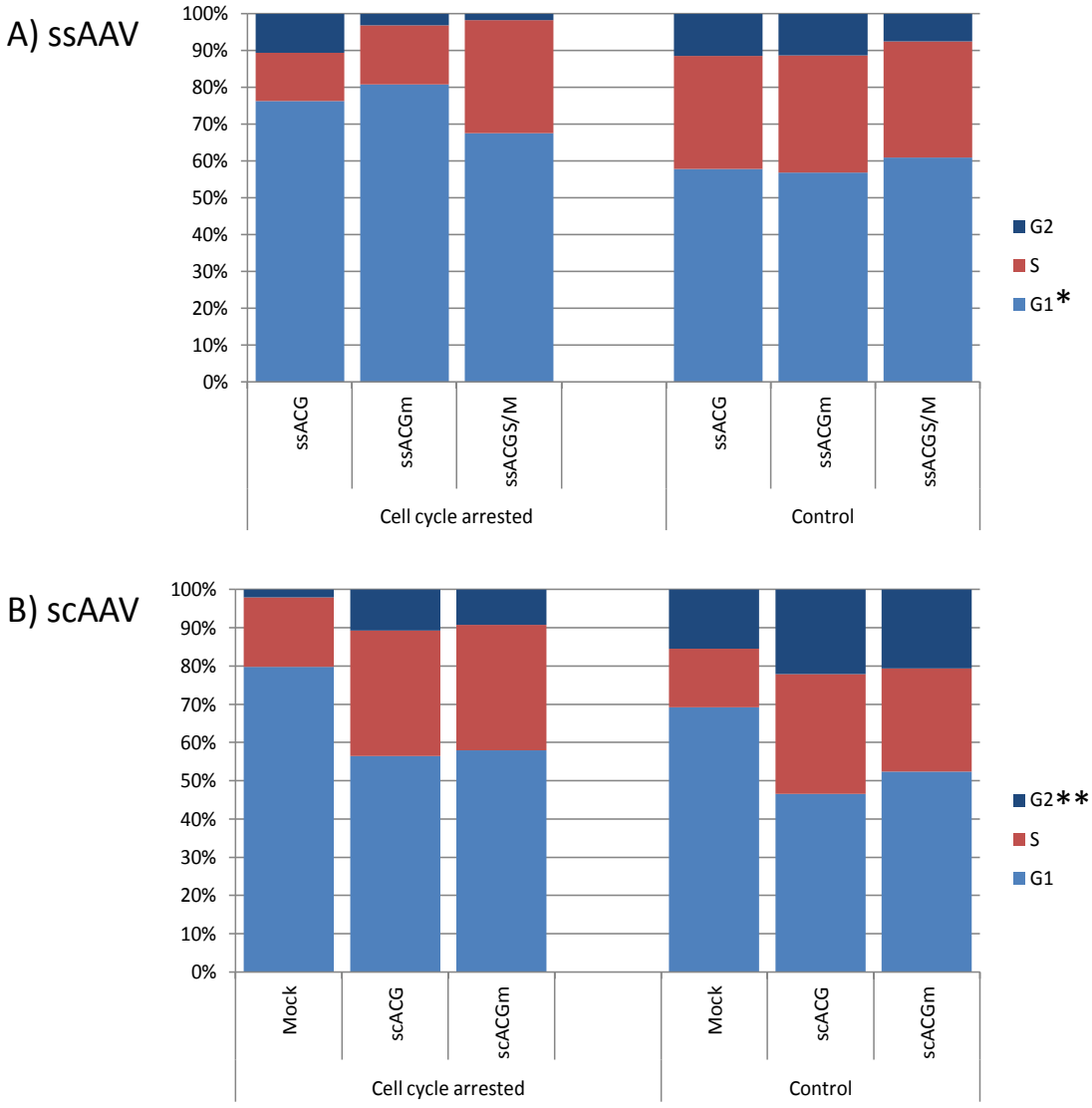


Figure 4.3 Cell cycle analysis by propidium iodide staining and flow cytometry of AAV-transduced CHO cells with and without induction of cell cycle arrest by methionine depletion.

*CHO cells were plated, transduced with ssAAV or scAAV vectors after 1 day with an MOI of 10^5, and then treated with methionine-free medium from D2 to D7, over a 5 day period. At D7, cultures were returned to normal growth medium and allowed to proliferate as normal with routine passaging as required. To confirm induction of transient cell cycle arrest, cell cycle phase analyses were performed by propidium iodide staining and flow cytometry at D7, at the end of the methionine depletion period. The coloured sections indicate the percentages of cells in each cell cycle phase as determined by PI staining followed by FlowJo analysis of the data. The cell cycle phases determined to differ significantly between treatments are denoted with *, as indicated by Student's t-test comparing each cell cycle phase separately between arrested and control at D7. * = P< 0.05, ** = P< 0.005. N=3.*

4.2.3 Optimisation of *in vitro* AAV transduction of CHO cells

CHO cells were transduced with scAAV-CMV-GFP vectors with MOIs ranging from 10 to 10^5. GFP expression in transduced cells was measured 48 h post-transduction by flow cytometry. As a result, MOI 10^5 was chosen for subsequent AAV transduction experiments since it resulted in the highest percentage of GFP+ cells, although even then the maximum achieved expression was 11%. A higher MOI was not possible due to constraints presented by vector availability, as well as the maximum volume of vector that can be added per well without significantly affecting the media concentration and pH.

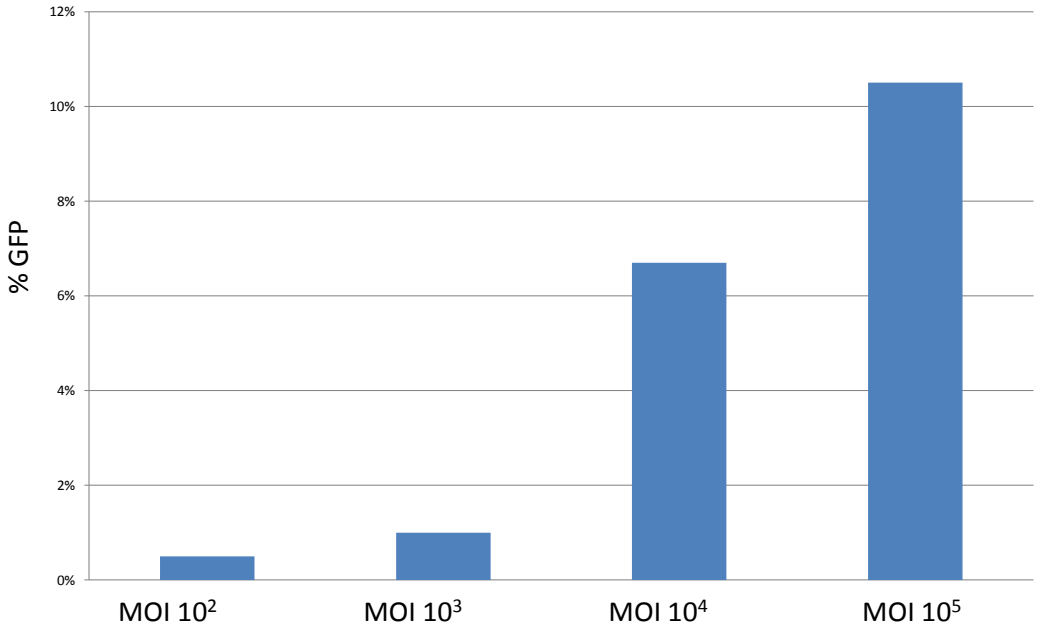


Figure 4.4 Optimisation of MOIs for AAV transduction.

1 x 10^5 CHO cells were seeded on 6-well plates in 2 ml media. 24 hours after seeding, the media was removed and 500 µl fresh serum-free media was added containing 10^2, 10^3, 10^4 or 10^5 scAAV9 CMV GFP virus particles per cell (MOIs 10^2, 10^3, 10^4 and 10^5 as calculated from the dot blot titres). After incubating at 37°C for 5 hours, 500µl of media containing 20% FBS was added. The proportion of GFP-positive cells was analysed 2 days post-transduction by FACScan analysis, with the mock population set to 0.5% GFP. FACScan analysis threshold set to 10^4 cells/ condition.

4.2.4 The effect of S/MAR elements and induced cell cycle arrest on episomal retention of single-stranded AAV vectors in transduced CHO cells

To evaluate the impact of transient cell cycle arrest and the presence of S/MAR elements on stability of ssAAV transduction and potential establishment of stable replicating AAV episomes, *in vitro* transduction experiments were conducted using 3 different single-stranded AAV serotype 8 vectors:

- ssAAV8 CAG GFP (ssACG)
- ssAAV8 CAG GFP miniMAR (ssACGm)
- ssAAV8 CAG GFP S/MAR (ssACG-S/M)

CHO cells were transduced with the ssAAV vectors using MOI of 10^5. At 24 h after transduction, the percentages of GFP+ cells in the cultures were 15, 16 % and 48% respectively. Transduced cultures were then either allowed to proliferate freely, or subjected to methionine depletion to induce cell cycle arrest for 5 days before returning to full growth media and propagated for 40 days with routine passaging.

In cultures not induced to undergo cell cycle arrest, the % GFP+ cells increased slightly up to day 3 post-transduction, reaching 22% GFP+ for both ssACG and ssACGm and 61% for ssACG-S/M. Thereafter the % GFP+ decreased rapidly reaching <1% by 7 days post-transduction.

In cultures induced to undergo transient cell cycle arrest between D2 and D7, a similar but temporally delayed effect was seen. However analysis at a later time point (Figure 4.5 C) shows that cells transduced with the S/MAR-containing vector and subjected to a transient cell cycle arrest retain expression at 3% GFP+ cells, which is significantly higher than all other vectors and culture conditions ($P < 0.001$ by one-way ANOVA and $P < 0.001$ for all pair-wise comparisons using Tukey's post-test, comparing % GFP+ cells at last 3 data points).

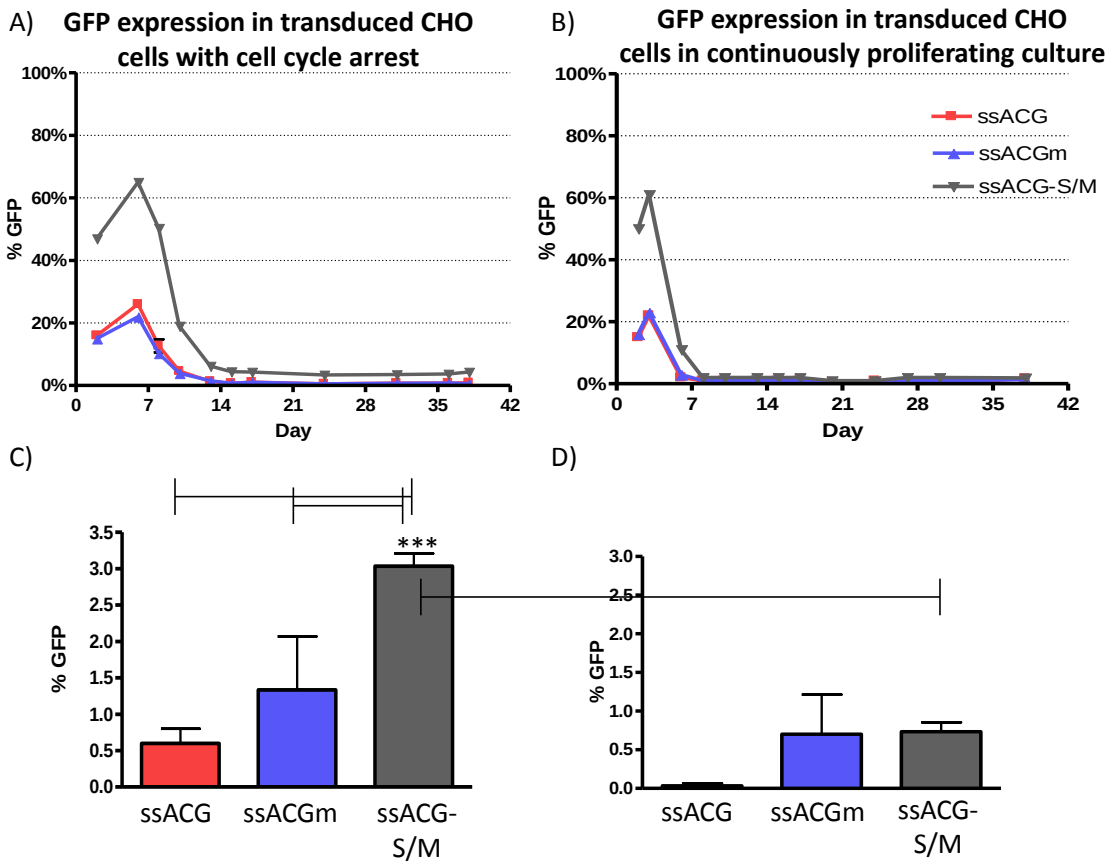


Figure 4.5 The retention of GFP expression in CHO cells transduced with ssAAV vectors is enhanced by S/MAR elements and induced cell cycle arrest.

*CHO cells were plated into culture, and after 1 day transduced with ssACG, ssACGm and ssACG-S/M vectors (MOI=10^5). Cells were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media at D2 (A and C) or allowed to proliferate freely (B and D). From D7, all cells were allowed to proliferate normally with routine passaging for a period up to D40. In (C) and (D) the stabilised transduction levels at the D40 time point are shown. Data are means ± s.e.mean, n=3. At D40 %GFP+ cells in cultures transduced with ssACG-S/M and subjected to an initial cell cycle arrest was significantly higher than for all other cultures (***= $P < 0.001$; one-way ANOVA with Tukey's post-hoc test). ssACG = ssAAV8 CMV GFP, ssACGm = ssAAV8 CMV GFP miniMAR, ssACG-S/M = ssAAV8 CMV GFP S/MAR.*

The intensity of the fluorescence emitted by the transduced cells was monitored by flow cytometry. Figure 4.6 shows the mean fluorescence intensity (MFI) of each transduced cell population, averaged weekly for all test points and triplicates. The cell populations transduced with the full-length S/MAR and held in an initial period of cell cycle arrest (light grey bars) show consistently elevated fluorescence intensity.

One-way ANOVA of the values from the last 3 measurements results in a statistically significant difference between the vectors (P <0.0002), and Tukey's test for multiple pair-wise comparisons states that the difference between the vector with the full-length S/MAR combined with cell cycle arrest is significantly different to the other 2 vectors when not held in cell cycle arrest.

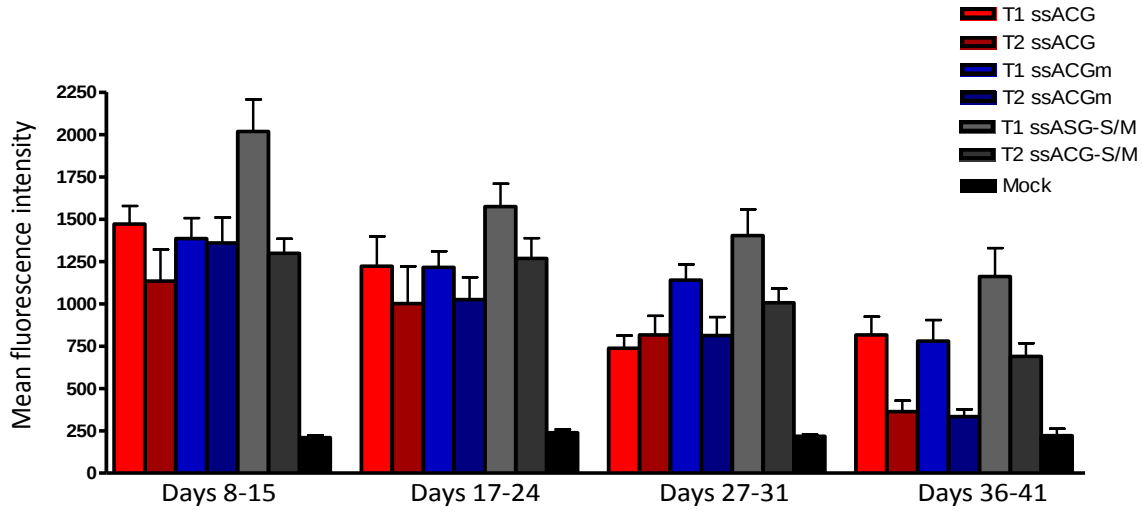


Figure 4.6 ssAAV vectors containing the β-IFN S/MAR and subjected to a period of induced cell cycle arrest show an elevated MFI in transduced cells.

CHO cells were transduced with ssACG, ssACGm and ssACG-S/M vectors (MOI=10^5). Day 1 post-transduction, cells were either induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media from D2 to D7 (T1) or allowed to proliferate freely (T2). From D7, all cells were allowed to proliferate normally with routine passaging for a period up to D41. MFI was measured by flow cytometry every 3-5 days. Each column represents a mean for each vector and treatment averaged over 3 time points, n=9 ± s.e.m. The data shows a trend towards higher expression intensity in cells transduced with the full-length S/MAR –containing vector combined with an induced period of cell cycle arrest (light grey bars). Treatment 1 (T1) = cultures with an initial cell cycle arrest, treatment 2 (T2) = continuously proliferating cultures. n= 6-9. ssACG = ssAAV8 CMV GFP, ssACGm = ssAAV8 CMV GFP miniMAR, ssACG-S/M = ssAAV8 CMV GFP S/MAR.

4.2.5 The effect of S/MAR elements and induced cell cycle arrest on episomal retention of self-complementary AAV vectors in transduced CHO cells

Vector cassettes including and omitting the miniMAR element were used to produce self-complementary AAV serotype 9 vectors to be used for *in vitro* transduction experiments:

- scAAV9 CMV GFP (scACG)
- scAAV9 CMV GFP miniMAR (scACGm)

CHO cells were transduced with the scAAV vectors at an MOI of 10^5 as previously. The transduction efficiencies were measured 24 h after transduction by flow cytometry, and were 10-11% for scACG and 1% for scACGm.

A proportion of the transduced cells were induced to enter cell cycle arrest at day 1 post-transduction. After 5 days in cell cycle arrest, 72% of the cells transduced with scACG and 5% of the cells transduced with scACGm were expressing GFP (Figure 4.7 A).

The remaining transduced cells were allowed to proliferate freely. In these cultures, GFP expression peaked at 3 days post-transduction, reaching very similar expression levels of 72% and 4% positive cells respectively (Figure 4.7 B).

Regardless of whether the cells were subjected to an initial period of cell cycle arrest or not, the %GFP+ cells declined during 10 days in proliferating culture before reaching low but stable levels of 0.9-2.1% (scACG) and 0-0.3% (scACGm) above background (Figure 4.7, C and D). Cells transduced with scACG maintained a higher proportion of GFP+ cells than cells transduced with scACGm, and the difference in stable %GFP+ cells was statistically significant between the vectors in both treatment groups ($P < 0.001$ by two-tailed t-test). For scACG the cell cycle arrest appears to have resulted in lower vector retention levels, and the difference is statistically significant (one-way ANOVA followed by Tukey's post-test for pair-wise comparison, $P < 0.05$).

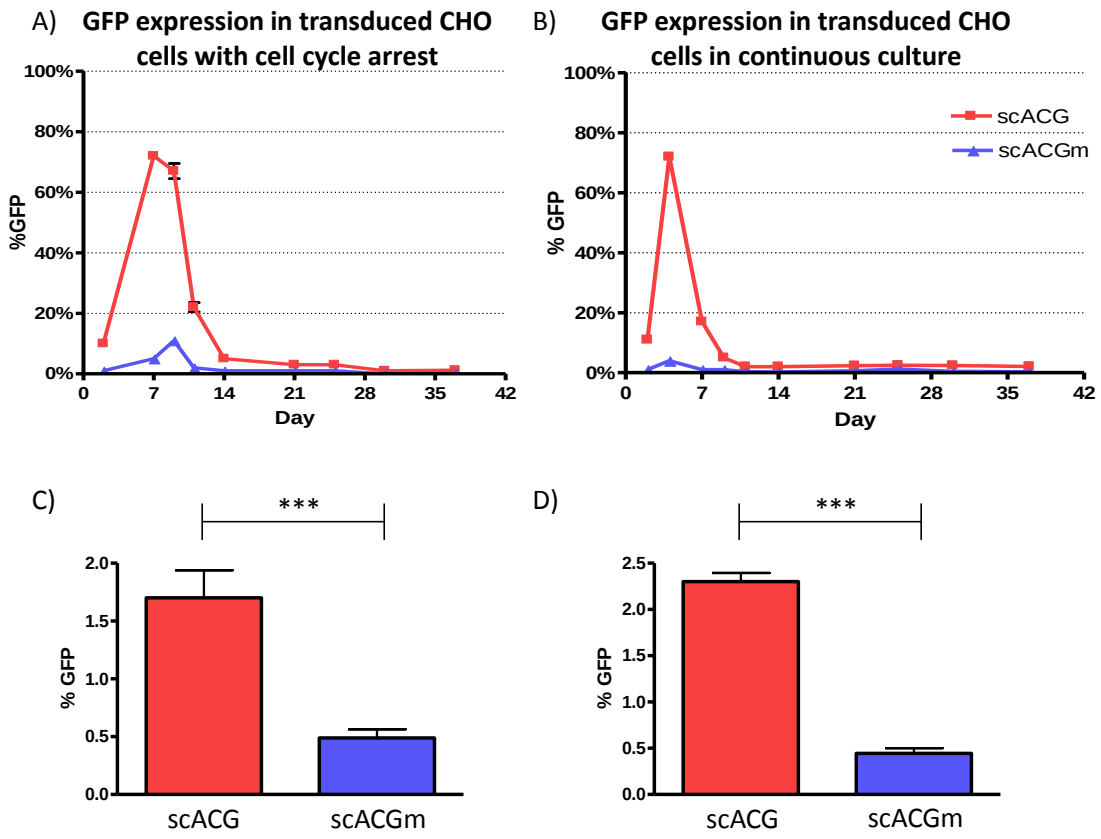


Figure 4.7 Transient GFP expression only is achieved in CHO cells transduced with scAAV vectors.

*CHO cells were plated into culture, and after 1 day transduced with scACG or scACGm vectors (MOI=10^5). Cells were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media at D2 (A and C) or allowed to proliferate freely (B and D). From D7, all cells were allowed to proliferate normally with routine passaging for a period up to D37. In (C) and (D) the stabilised transduction levels at the D37 time point are shown. There is a statistically significant difference (***, P <0.001 by ANOVA and Tukey's post-test) between the vectors regardless of whether the cells were subjected to an initial period of cell cycle arrest (C and D). The quiescent period did not result in a significant increase in retention for either of the vectors. scACG = scAAV9 CMV GFP, scACGm = scAAV9 CMV GFP miniMAR.*

The mean fluorescence intensity of each cell population was monitored at regular intervals, with somewhat differing results to those achieved with the single-stranded AAV vectors (Figure 4.8 B). Here, the MFI of cells transduced with scAAV9 CMV GFP was significantly different to those transduced with the miniMAR –containing vector when compared across all time points in both treatment groups (P <0.0001, paired t-test).

The stable level of fluorescence intensity as determined by averaging the MFI between all measurements from 14 to 37 days post-transduction was higher for both vectors after an initial period of cell cycle arrest. The averages and associated significance as determined a paired t-test are shown in Figure 4.8 A.

A)

Vector	Av. MFI with Cell cycle arrest	Av. MFI without Cell cycle arrest	Significance of difference
scAAV9 CMV GFP	2902	1970	P = 0.024
scAAV9 CMV GFP mMAR	778	412	P = 0.023

B)

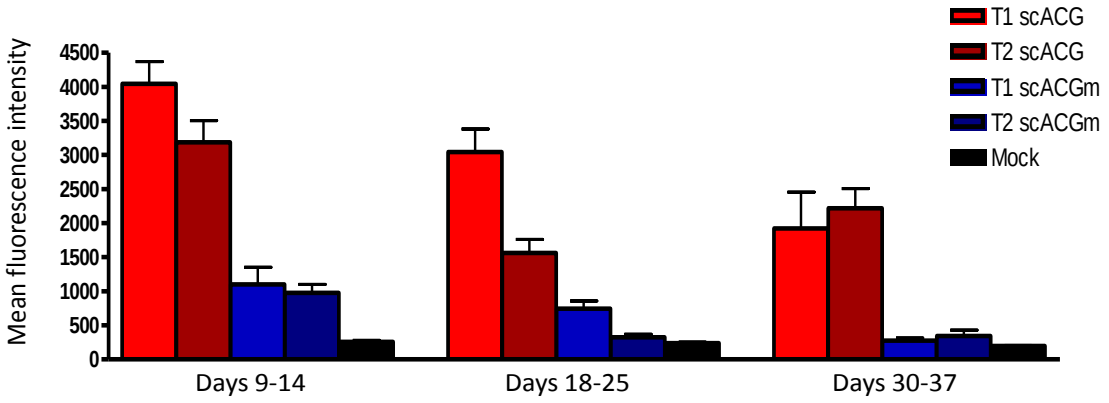


Figure 4.8 Mean fluorescence intensity of GFP expression is higher in cells transduced with scAAV9 CMV GFP without the miniMAR-element.

CHO cells were transduced with scACG and scACG-S/M vectors (MOI=10^5). At day 1 post-transduction, cells were either induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media from D2 to D7 (T1) or allowed to proliferate freely (T2). From D7, all cells were allowed to proliferate normally with routine passaging for a period up to D37. MFI was measured by flow cytometry every 3-5 days.

In (A), the stable level of fluorescence intensity as determined by averaging the MFI between all measurements from 14 to 37 days post-transduction is shown.

In (B), each column represents a mean for each vector and treatment averaged over 3 time points, n=9 ± s.e.m. The MFI of cells transduced with scAAV9 CMV GFP (sc9CG) is significantly different to those transduced with the miniMAR –containing vector sc9CGmM across all time points in both treatment groups (P < 0.0001, paired t-test). Treatment 1 (T1) = cultures with an initial cell cycle arrest, treatment 2 (T2) = continuously proliferating cultures. n= 6-9. scACG = scAAV9 CMV GFP, scACGm = scAAV9 CMV GFP miniMAR.

4.3 Discussion

4.3.1 Induction of a period of cell cycle arrest following AAV transduction

The induction of a reversible period of cell cycle arrest was successful in both ssAAV and scAAV transductions. Following 5 days in methionine and serum depleted media, in the ssAAV transduction experiment the proportion of the cells in the G_1 phase significantly higher than in the proliferating control population. A trend towards an increase in the G_1 population and a significant decrease in the G_2 population was also observed in the scAAV transduction experiment. This would be expected to happen if the cell cycle was arrested at the end of G_1 phase, and so it can be concluded that a substantial amount of cells were arrested at this stage in both experiments.

4.3.1.1 Construction of hypothetical nuclear episome structures of AAV vectors

In order to simulate the DNA structures which the AAV nuclear episomes will form upon transduction of cells, *in silico* models were created that attempt to mimic the types of molecules created by AAV DNA recombination and circularisation events. During transduction, AAV virions enter the nuclei and release the packaged DNA genomes, which then form episomal structures that vary both within and between the vector types used.

When the packaged genome is single-stranded DNA as in many AAV vectors, the next step is either second strand synthesis or annealing of opposite sense strands. Once double-stranded, the AAV genomes form circular episomes by recombination and annealing between ITRs. In self-complementary vectors, the packaged genome is a palindrome and is able to fold back onto itself forming a double-stranded DNA molecule with the mutated ITR structure in the middle and two terminal repeats folded in hairpin structures. This overcomes the need for second-strand synthesis, and also changes the types of episomal genomes present in the transduced cells as the double stranded genomes have different ITRs at each end (McCarty *et al.*, 2001).

A study into the structure of ITR to ITR junctions in ssAAV vectors showed that the most likely model for AAV circularisation involves the initial rearrangement of the ITRs into hairpin structures followed by recombination and/or annealing between the two ITRs at each end of the genome (Duan *et al.*, 1999b). This may be succeeded by intermolecular recombination events, resulting in an array of junctional sequences containing anywhere between one and two complete ITRs depending on where the recombination has occurred. Genomes can recombine in various ways to produce concatemers in a head-to-tail, head-to-head or tail-to-tail orientation. It has been shown that the head-to-tail conformation is more common in persistent AAV genomes than the alternative head-to-head or tail-to-tail structures (Duan *et al.*, 1998).

The genomes of scAAV vectors have also been found to concatemerise *in vivo*, although less efficiently than ssAAV. Most of the vector genomes circularise upon themselves within 24 hours of entering the nucleus. Intermolecular recombination occurs with an efficiency of about 2%, with recombination between the closed hairpin ends (mutated ITRs) being more efficient than that between the open ends containing the wt ITRs (Choi *et al.*, 2005). Therefore, the scAAV dimers were constructed as head-to-tail dimers, with the mutated ITRs and wild type ITRs each assumed to recombine with a homologous ITR.

However, it is important to note that all AAV sequences form secondary hairpin structures *in vivo* and hence the SIDD predictions may not be accurate; the strand separation potential predicted at these sequences by the algorithm may indeed be a driving force behind the formation of the hairpin structures. Whether the 3-dimensional hairpin structures are more stable than the equivalent double-stranded DNA is not taken into account by the algorithm, although it is likely to be the case considering the dominance of the hairpin conformation both *in vitro* and *in vivo* (Berns *et al*, 2007).

4.3.1.2 Implications of SIDD analysis of ssAAV & scAAV structures

The most likely forms of ssAAV-derived genomes more than 24 hours post-transduction are concatemers consisting of 2 or more genomes; however, the maximum length of input sequence for the WebSIDD algorithm is 10 kb, and therefore the analysis was performed on dimeric, concatenated structures as multimeric concatemers would exceed the maximum allowed length. Within each concatemer, there are several possible ways in which the ITRs may join either through homologous recombination between shared regions or non-homologous end joining (NHEJ) and as a result, the ITR structures within concatemers may differ and each contain varying combinations of ITR sequences.

The two extreme alternative forms are one containing 2 copies of all ITR sequences, and one where all duplication of sequence motifs has been eliminated through homologous recombination, resulting in 1 complete ITR structure at the junction. In reality, the sequences of ITR-ITR junctions vary, with some duplication of sequences; hence it was important to analyse both extremes to define the range of possibilities that lie in between.

Although the graphical outputs of the alternative forms appeared similar, the algorithm found 10 suggested SIDD sites in the molecule with completely recombined ITRs, whereas only 6 potential SIDD sites were found in the molecule containing a completely duplicated ITR-ITR junction (see Table 4.1). The likely explanation for the molecule with less ITR sequences to contain more SIDD sites is that each ITR sequence places torsional stress on the molecule and therefore has a modest condensing effect on the neighbouring sequences. An ITR-ITR junction containing only 1 of each ITR sequence component will therefore allow for slightly more destabilised regions to appear. Experimentally, it has been shown that the predominant sequence motif at the ITR-ITR junctions is one complete hairpin structure flanked by two D-motifs (Duan *et al.*, 1999b). As very little difference was found in the location of SIDD sites between the two extreme forms of ssAAV CAG GFP despite some differences in the exact locations of likely strand separation, the more common form containing a recombined ITR junction was chosen for the analysis of the S/MAR containing vectors.

The lack of strand separation potential near any of the wild type AAV promoters is a likely result of the need for the gene expression to be tightly controlled and shut down during latent infection. The promoters are not destabilised until activated by transcription factors, which allows them to remain quiescent during latent infection. In the recombinant AAV genomes, all of the exogenous transgene transcriptional elements (enhancer, promoter, polyA signal) were chosen for the transgene cassette due to their ability to drive strong constitutional gene expression. It is possible that there is a correlation between the high strand separation potential and the strength of gene expression endowed by these elements.

In all of the episomal structures studied, the addition of S/MAR elements causes changes in the destabilisation potential of the rest of the domain. Regions that exhibit high duplex destabilisation potential, such as promoters, polyA signals and origins of replication, become more stabilised when a S/MAR element is added in the same domain. In the case of both ssAAV and scAAV, the addition of S/MAR elements results in severe condensation of the formerly destabilised regions. This may impede the access of the transcription machinery to the transcription start site, and therefore result in a reduction of expression levels.

On the other hand, the strong destabilisation potential conferred by the S/MAR element onto all the vectors studied suggests that the element may be capable of replication initiation and nuclear matrix interactions in the context of these domains, as the function of S/MARs in episomal retention is linked with their capacity for strand separation. The strength of SIDD for the S/MAR element did not change in the alternative episomal forms of vector genomes, suggesting that none of the alternative *in vivo* forms of the vectors should show an advantage over other forms.

4.3.2 Effect of S/MAR elements and induced cell cycle arrest on the retention of AAV genomes in transduced CHO cells

4.3.2.1. Vectors used in the experiments

The reason for including both ssAAV and scAAV vectors in the experiments was that they bestow different potential advantages to the gene delivery system; the ssAAV cassette can accommodate both miniMAR and full-length S/MAR elements, whereas the scAAV is able to deliver DNA that will be double-stranded soon after entry. This could be of importance if the S/MAR element needs to be double-stranded in order to complete the interactions with the nuclear matrix, something which is currently not known.

The ssAAV and scAAV vectors were packaged in capsids belonging to a different serotype. AAV9 exhibits a similarly wide tissue tropism compared to AAV8 with the added advantage of being able to cross the blood-brain barrier and transduce cells of the CNS, which makes it a good prototype vector for a multitude of applications. In *in vitro* applications, the range of transduction efficiencies for CHO cells was similar between the two serotypes, and the same MOI of 10^5 was used for both ssAAV and scAAV. Here, AAV9 was included with the anticipation of further *in vivo* experiments; however these were not carried out due to the poor performance of the miniMAR-containing scAAV9 *in vitro*.

Of the ssAAV vectors, the full-length S/MAR containing vector achieved the highest transduction efficiency, 61% GFP+ cells 3 days post-transduction, compared to the other two ssAAV vectors which both yielded only 22% GFP+ cells at the same time point. This may have in part been influenced by potential inaccuracies in titration resulting in a larger number of ssACG-S/M used in the transduction. The titres of ssACG and ssACGm used in the experiment were very similar at ~$3x10^{12}$ and $5x10^{12}$, whereas the estimated titre of the ssACG-S/M vector was 10-fold lower $1.6x10^{11}$ (see Materials & Methods, section 2.2.11.2). However, it is also possible that the difference in GFP expression is due to the properties of the full-length S/MAR element, which may act as a transcription enhancer or prevent gene silencing.

The transduction efficiencies varied about 10-fold for the two different scAAV vectors, with scACG yielding 72% GFP+ cells at 3 days post-transduction, whereas the same vector containing the miniMAR element yielded only 4% GFP+ cells. This was despite the vectors having very similar titres of $3.92x10^{11}$ and $3.83x10^{11}$ (see M&M, section 2.15, table 2.3) (see also section 4.3.2.2).

4.3.2.2. The effect of S/MAR elements and cell cycle arrest on retention of AAV vectors in transduced CHO cells

The ssAAV8 CAG GFP S/MAR vector showed an increased propensity for long-term gene expression, which was achieved only when the cells were held in cell cycle arrest after transduction. Under these conditions some 3% of the cells stably expressed GFP, in the absence of any selection, for more than one month of continuous growth. Cells transduced with the same vector but not held in cell cycle arrest lost GFP expression completely by D10. These results potentially suggest that the activity of the S/MAR element is enhanced by an extended presence in a quiescent nucleus, which may assist in the establishment of replicating episome status.

The miniMAR-element did not have an effect on long term transduction with either ssAAV or scAAV vectors. In the case of self-complementary vectors, the addition of the miniMAR element to the cassette actually decreased the %GFP+ cells in both treatment groups. Not only that, but inclusion of the miniMAR also significantly decreased the mean fluorescence intensity of the positive cells. The most likely explanation for this is that the presence of the miniMAR element in this configuration inhibits transcription, and support for this hypothesis is offered by the SIDD analysis which shows the CMV promoter condensing from <1 kcal/mol to ~7 kcal/mol upon the addition of the miniMAR in dimeric circles (see Figure 4.2).

When incorporated into the ssAAV cassette, the miniMAR element did not result in any difference in transgene expression or retention, whether combined with an initial period of cell cycle arrest or not. For the ssAAV vectors, the MFI is not decreased by the presence of the miniMAR element as drastically as it is for the scAAV vectors (Figures 4.6 and 4.8). Accordingly, when comparing the outputs of SIDD analysis it can be observed that the promoter regions are not as severely condensed in the ssAAV after the addition of the miniMAR, probably due to the longer length of the concatenated

sequence. In ssAAVs, the energy required for strand separation at the CMV promoter region is increased from <1 kcal/mol to ~5 kcal/mol, which although increased is considerably less so than the equivalent increase in scAAV.

The reason for increased transgene retention following inclusion of the full-length S/MAR element in the cassette could be protection against silencing or insulator effect conferred by the element, as has been shown in other vector systems (Dang *et al.*, 2000). However, it is also possible that during cell cycle arrest, the S/MAR element acts to promote increased integration into the host genome, although this is unlikely as no such effect has been reported in other studies not featuring a period of enforced cell cycle arrest. To investigate this further, dilution cloning could be used to obtain clonal GFP-positive populations, which could then be used for Southern blot and LAM-PCR analysis for the episomal status of the AAV genomes.

4.3.3 Conclusions

When incorporated into an AAV vector cassette, the β-Inf S/MAR produced some intriguing effects. The inclusion of the full-length S/MAR in a ssAAV configuration and combined with a cell cycle arrest resulted in 3% stably expressing population; however, no increase in transgene retention was observed when no initial period of cell cycle arrest was induced. The difference between the full-length S/MAR combined with cell cycle arrest and all other vector configurations was significant. Previous studies of the β-Ifn S/MAR element have reported an increase in episome retention regardless of an initial period of cell cycle arrest, however all studies published to date have used vectors which are imported into the nucleus in double-stranded form (Harraghy *et al.*, 2008; Piechaczek *et al.*, 1999). The results support the hypothesis that to achieve episomal establishment of a single-stranded vector, allowing the vector to remain in the nucleus for an extended period of time following transduction is required so that second strand synthesis can be completed and S/MAR- nuclear matrix interactions can take place prior to cell division.

The naturally occurring truncated variant, miniMAR, produced no effect on transgene expression or retention in ssAAVs. This indicates that the function of the miniMAR element may be dependent on the context in which it is provided, as another group has reported equivalent function between the miniMAR and the full-length S/MAR when

incorporated into a minicircle (Broll *et al.*, 2010). In the scAAV vectors, the miniMAR element was positively detrimental, decreasing both the proportion of expressing cells and the expression levels of positive cells. This is most likely due to the promoter condensing effects of the miniMAR inclusion in such a small, topologically constrained cassette. Interestingly, the β-Ifn S/MAR has previously been shown to result in elimination of function of other elements whose activity depends on strand separation such as origins of replication (Giannakopoulos *et al.*, 2009), supporting the conclusion that in some context the effects of S/MAR elements may be disadvantageous as well as beneficial.

Although the level of stably expressing cells achieved with the combination of the full-length S/MAR and initial cell cycle arrest was fairly low, the proportion could be increased either by placing the population under selection pressure or with cell sorting. Furthermore, such an approach could easily be adopted for *ex vivo* gene therapy, if safe and stable episomal expression can be demonstrated, and therefore continuing the investigations remains a viable option although beyond the scope of the current project.

Chapter 5: The effect of a S/MAR element and induced cell cycle arrest on the retention of integration-deficient lentiviral vectors

5.1. Introduction

Integrating lentiviral vectors are amongst the most widely utilised and effective gene transfer vectors currently under investigation, but integration events may compromise biosafety (see Introduction, section 1.3.2). This chapter aims to investigate the possibility for creating equally efficient and long-lasting transgene expression using safer integration-deficient lentiviral vectors (IDLVs). The studies were designed to investigate the potential of IDLVs to lead to establishment of stable episomes in CHO cells, when combined with S/MAR elements and/or cell culture manipulation to induce a transient period of cell cycle arrest.

IDLVs have a natural tendency towards forming circular episomal structures in the nuclei of transduced cells but these are usually lost during cell division, thus limiting the potential application of such vectors to mainly stable post-mitotic tissues. In this chapter, the potential of S/MAR elements and post-transduction cell cycle arrest to assist in the establishment of mitotically stable replicating episomes was explored. In this case the S/MAR element included in the vectors was the truncated miniMAR, and the transient cell cycle arrest was induced in transduced CHO cells by methionine and serum depletion of the culture media.

The second half of the chapter presents data to evaluate the extra-chromosomal episome or intra-chromosomal integrated status of the IDLV genomes. Polyclonal and clonal cell populations containing putative stable IDLV episomes were analysed by Southern blotting, LAM-PCR, deep sequencing and fluorescent *in situ* hybridisation (FISH) to ascertain genetic configuration, intranuclear location and potential episomal versus integrated status of the vector-derived transgenes.

5.2 Results

5.2.1 Computational duplex destabilisation analysis of the episomal IDLV structures

Both integrating and non-integrating lentiviral genomes give rise to circular products during cell transduction. Circularisation of the linear double-stranded vector DNA can occur by homologous recombination or non-homologous end-joining of the two LTRs. In this way, dsDNA circles containing either one or two LTRs are formed. In cells transduced by LVs and IDLVs, 1-LTR circles are found more commonly than 2-LTR circles, although IDLVs produce comparatively more 2-LTR forms (Bayer *et al.*, 2008).

SIDD evaluations were performed to determine the effect of the presence of one or two LTRs and/or miniMAR or IFN-S/MAR elements on the duplex stability profiles of circular IDLV episomes, and hence their potential ability to support replication.
Firstly, in the context of comparing 1-LTR and 2-LTR circles, SIDD profiles derived from control, miniMAR and IFN-S/MAR vectors showed no significant difference. An example is shown in Figure 5.1 for a control non-S/MAR IDLV vector, and a similar outcome was evident with miniMAR and IFN-S/MAR vectors. This result indicates that at least in terms of duplex destabilisation and the potential for local strand separation and episomal replication initiation, both 1-LTR and 2-LTR circles are indistinguishable.

Subsequent SIDD analyses were then performed on only the predicted 1-LTR forms of circular IDLV nuclear episomes to determine the effect of the presence or absence of miniMAR and IFN-S/MAR elements on the duplex stability profiles. Since S/MAR elements have been shown to need read-through transcription to function effectively, there are various theoretically possible functional locations within the vector cassette; for example immediately after the GFP coding sequence but before the polyA site in the 3'LTR, or after the *gag* gene relying on residual promoter activity from the 5'LTR. In our case only the former variation was evaluated.

In the putative 1-LTR nuclear episome configurations, the miniMAR and IFN-S/MAR elements were found to retain strong periodic duplex instability and propensity for strand separation. The inclusion of either element also generally decreased the duplex instability of other unstable regions throughout the molecule including the SV40 promoter and OriT region of the GFP transgene transcription unit (Figure 5.2). In both cases, the energy required for strand separation at the SV40 promoter increased from ~5.5 kcal/mol to ~7 kcal/mol, and a similar effect was observed in the presence of both S/MAR elements.

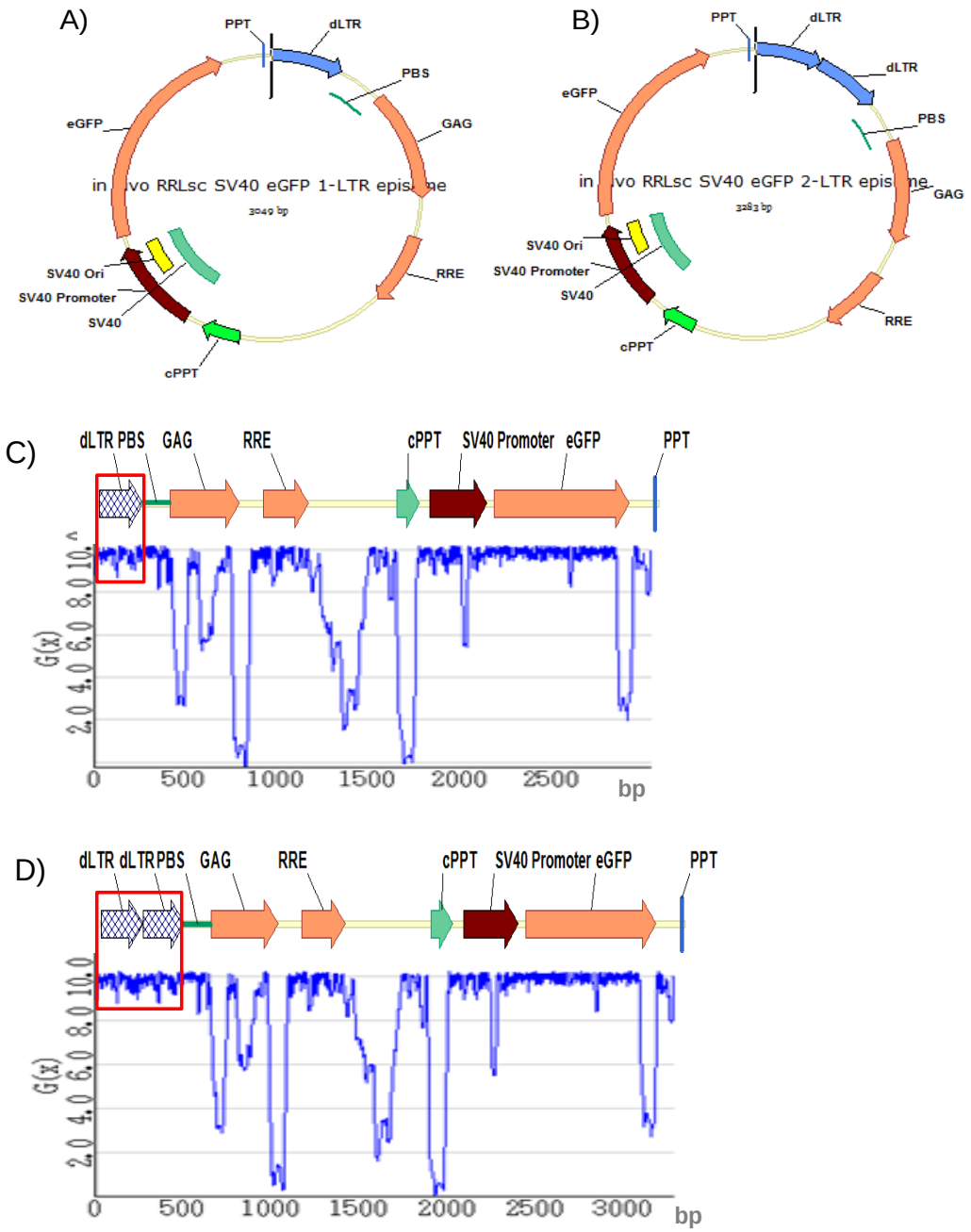


Figure 5.1 SIDD analyses of potential 1-LTR and 2-LTR circular nuclear episomes derived from a standard IDLV do not reveal significant differences.

The strand separation potential profiles from SIDD analyses are shown for 1-LTR (A and C)) and 2-LTR (B and D) circular nuclear episomes thought to arise during cell transduction with the IDLV-SV40-GFP vector. The SIDD plots show G(x),the average free energy of strand separation for each individual base pair in kcal/mol, plotted against the episome base sequence number where position 1 is set at the start of the 1-LTR or 2-LTR regions. Boxed regions in C and D indicate the LTR(s).

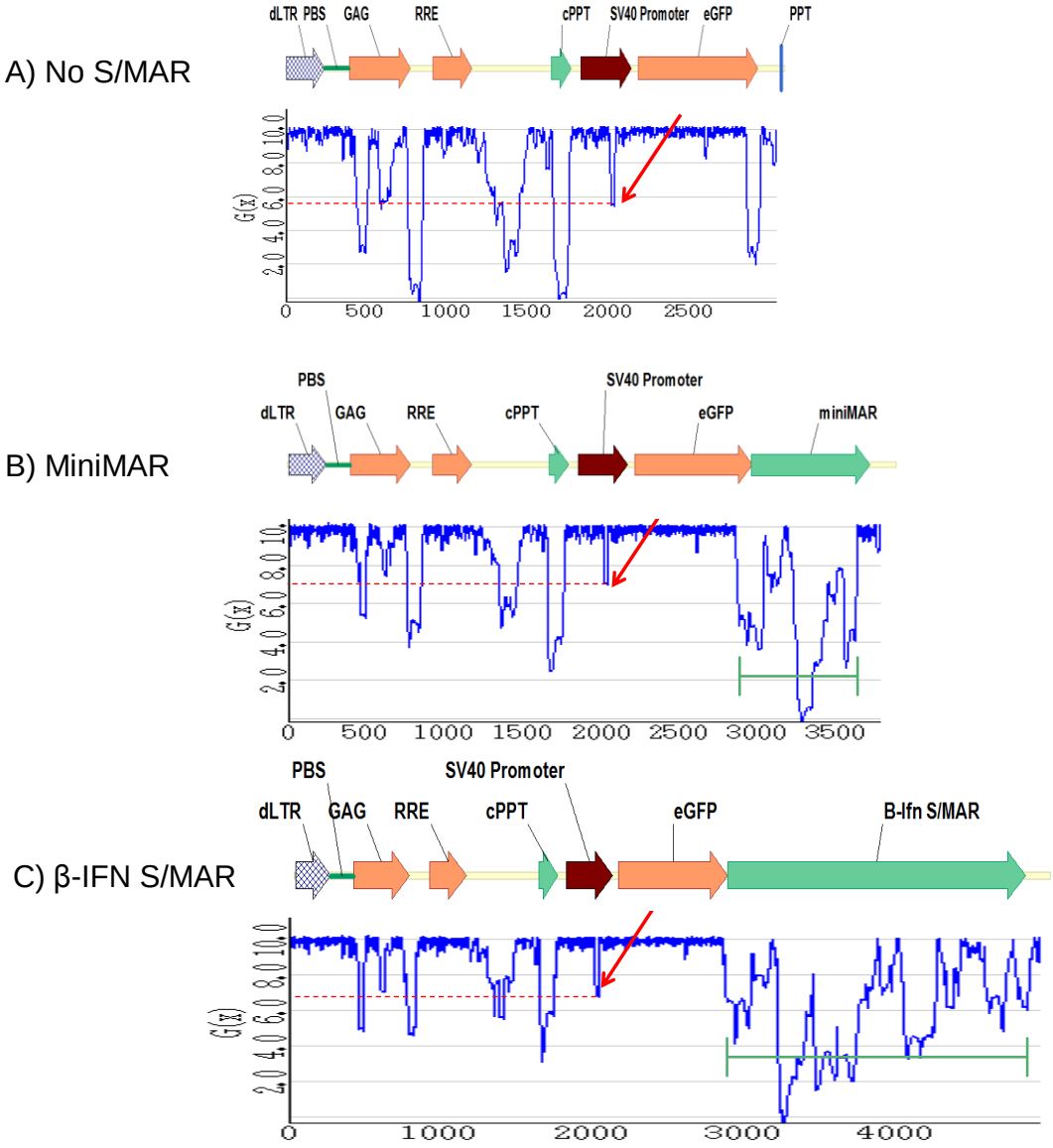


Figure 5.2 SIDD analyses of potential 1-LTR circular nuclear episomes derived from IDLVs containing miniMAR and IFN-S/MAR elements.

The strand separation potential profiles from SIDD analyses are shown for 1-LTR circular nuclear episomes thought to arise during cell transduction with (A) control IDLV-SV-GFP vector, or IDLVs containing (B) miniMAR, or (C) IFN-S/MAR elements located between the GFP coding sequence and the 3'LTR. The SIDD plots show G(x), the average free energy of strand separation for each individual base pair in kcal/mol, plotted against the episome base sequence number where position 1 is set at the start of the 5'LTR. The arrows indicate the destabilisation peak for the SV40 promoter/OriT element, and the green lines the destabilisation regions for the S/MAR elements.

5.2.2 Induction of cell cycle arrest by methionine and serum depletion of culture medium following IDLV transduction

As described in Chapter 3 for cells transfected with plasmids (see section 3.2.2), CHO cell cultures transduced with various IDLVs were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media. Cells were plated, transduced after 1 day, and then treated with methionine-free medium from D2 to D7 over a 5 day period. At D7, cultures were returned to normal growth medium and allowed to proliferate as normal with routine passaging as required. To confirm induction of cell cycle arrest, cell cycle phase analyses were performed by propidium iodide staining and flow cytometry at the D2 and D7 time points, effectively immediately before and at the end of the methionine and serum depletion period.

As shown in Figure 5.3 both for cells transduced with IDLV vectors and mock control, methionine depletion of culture media effectively increased the proportion of cells in G_1 phase (P <0.05), and decreased the proportion of cells in S-phase (P <0.01) and in G_2 phase (P =0.0001). This confirms the effectiveness of methionine depletion for inducing a transient cell cycle arrest, and the independence of the cell cycle arrest of IDLV transduction.

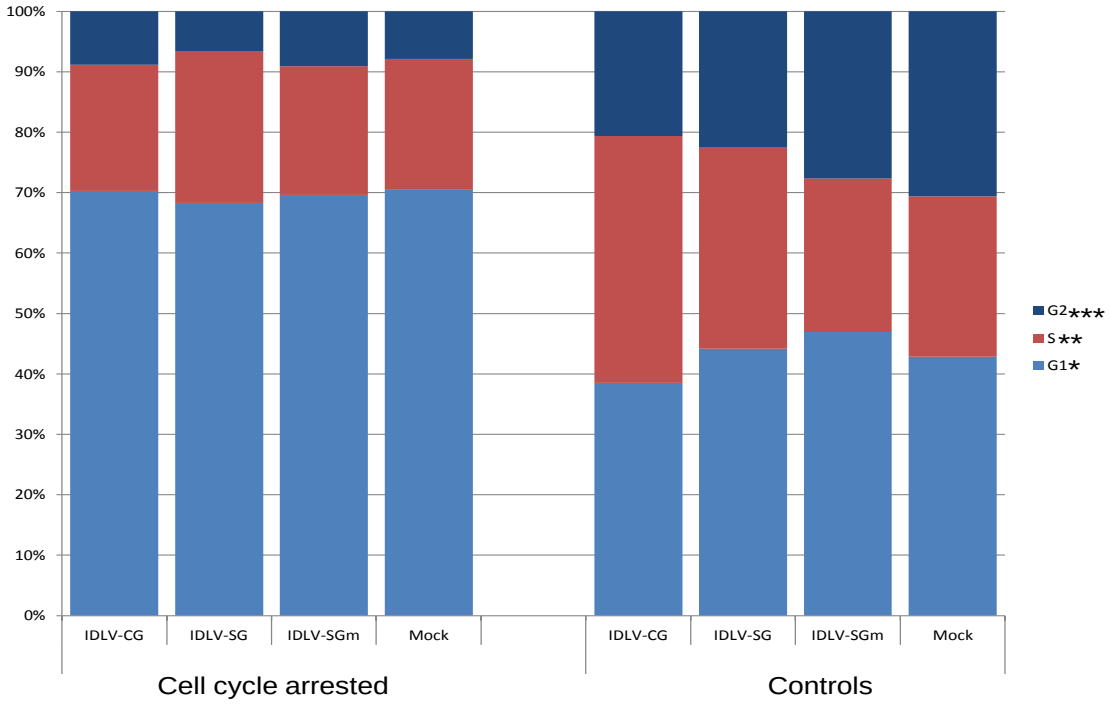


Figure 5.3 Cell cycle phase analysis by propidium iodide staining and flow cytometry of IDLV -transduced CHO cells with and without induction of cell cycle arrest by methionine and serum depletion of culture media.

*CHO cells were plated, transduced with IDLV vectors or mock control after 1 day, and then treated with methionine-free medium from D2 to D7, over a 5 day period. At D7, cultures were returned to normal growth medium and allowed to proliferate as normal with routine passaging as required. To confirm induction of transient cell cycle arrest, cell cycle phase analyses were performed by propidium iodide staining and flow cytometry at D7, at the end of the methionine depletion period. The non-arrested controls were analysed at the same time point. The coloured sections indicate the average percentages of cells in each cell cycle phase as determined by PI staining followed by FlowJo analysis of the data. The cell cycle phases determined to differ significantly between treatments are denoted with *, as indicated by Student's t-test comparing each cell cycle phase separately between arrested and control at D7. * = P <0.05, **= P <0.01, ***= P <0.001. IDLV-SG = IDLV SV40 GFP; IDLV-SGm = IDLV SV40 GFP mMAR; IDLV-CG = IDLV CMV GFP. N=4.*

5.2.3 Transient induction of cell cycle arrest results in significant stable transduction of CHO cells by IDLVs

To evaluate the impact of transient cell cycle arrest and the presence of a S/MAR element on stability of IDLV transduction and potential establishment of stable replicating IDLV episomes, *in vitro* transduction experiments were conducted using 3 different VSVg-pseudotyped integration-deficient LVs:

- IDLV SV40 GFP (IDLV-SG)
- IDLV SV40 GFP mMAR (IDLV-SGm)
- IDLV CMV GFP (IDLV-CG)

After testing at various MOIs it was determined that an MOI of 1 routinely yielded greater than 90% GFP+ cells.

CHO cell cultures transduced with the 3 IDLVs were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media. Cells were plated, transduced after 1 day, and then treated with methionine-free medium from D2 to D7 over a 5-day period. At D7, the cultures were returned to normal growth medium and allowed to proliferate as normal with routine passaging as required for a period of up to 72 days.

As shown in Figure 5.4 (A and B), at D3, IDLV-SG, IDLV-SGm and IDLV-CG transductions yielded > 90% GFP+ cells. At 2 days after transduction of the cells kept in continuous proliferative culture and not subjected to cell cycle arrest, GFP expression began to decline steadily reaching levels of under 3% after 14 days (Figure 5.4, B and D). By 72 days in proliferative culture, only 0.8%, 0.5% and 1.4% GFP+ cells could be detected in cultures transduced with IDLV-SG, IDLV-SGm and IDLV-CG vectors respectively. These levels match what would be expected to result from residual integration (Wanisch *et al.*, 2009). There was no statistically significant difference between the vectors at this stage (One-way ANOVA, n =3).

In the transduced cultures subjected to a transient cell cycle arrest, the GFP expression increased to 100% by the end of the 5-day period of methionine depletion. Thereafter however, the GFP expression pattern in these cultures was distinctly different from those in continuous proliferative culture.

Although the proportion of GFP+ cells did decline from the point when the cultures were released from cell cycle arrest (D7 onwards), the expression level stabilised after ~20 days at relatively high levels: 22%, 17% and 16% GFP+ cells respectively for cultures transduced with IDLV-SG, IDLV-SGm and IDLV-CG vectors (Figure 5.4, A). After 72 days in continuous proliferative culture, these substantial levels of GFP+ cells were maintained at 25%, 15% and 8% respectively for cultures transduced with IDLV-SG, IDLV-SGm and IDLV-CG vectors (Figure 5.4, C).

Thus while there were no differences in GFP expression levels between the cell populations initially, by 72 days in culture the transduction levels in the cell cultures which had been subject to a transient period of cell cycle arrest all remained substantial and very significantly higher than for the cell cultures allowed to proliferate continuously and than what would be expected to result from residual integration. In addition, for the cell cultures which had been subject to a transient period of cell cycle arrest the level of stabilised transduction at D72 was significantly different between the different IDLV vectors. The IDLV-SG vector yielded significantly higher levels of stable transduction compared to IDLV-SGm ($P < 0.01$) and IDLV-CG ($P < 0.001$), by one-way ANOVA followed by Tukey's post-test.

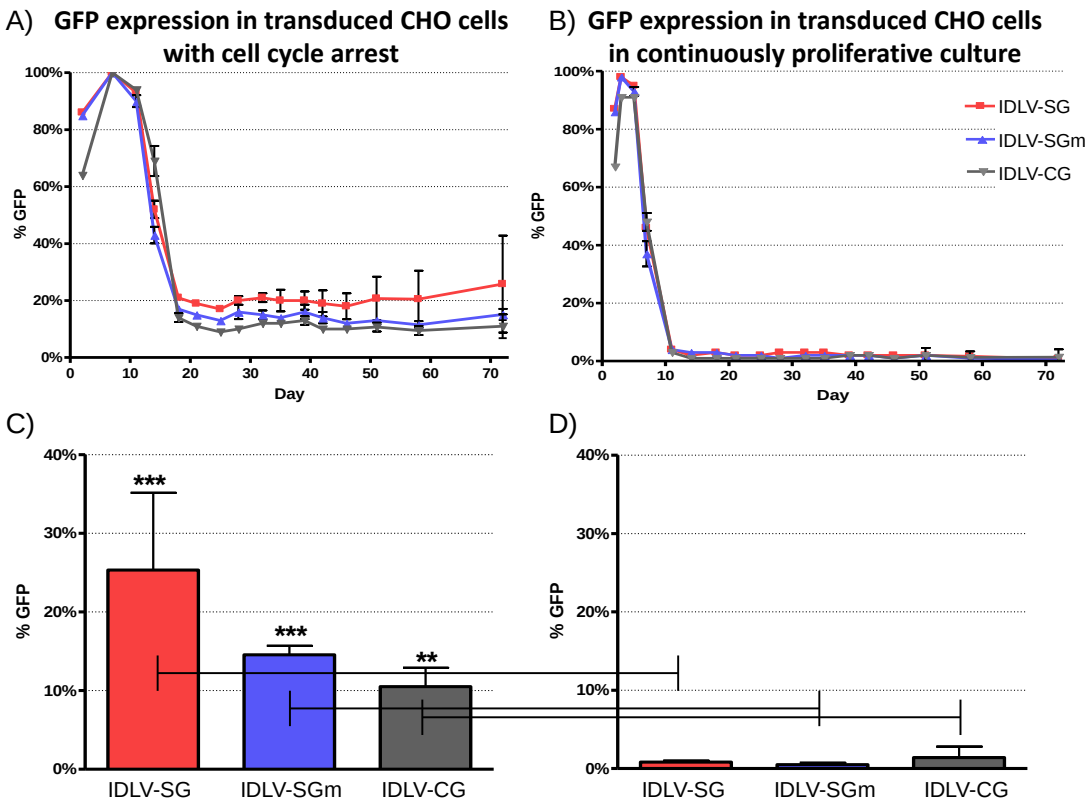


Figure 5.4 Transient induction of cell cycle arrest leads to substantial levels of stable GFP expression in IDLV-transduced CHO cells

*CHO cells were plated into culture, and after 1 day transduced with IDLV-DG, IDLV-SGm and IDLV-CG vectors (MOI=1). Cells were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media at D2 (A and C) or allowed to proliferate freely (B and D). From D7, all cells were allowed to proliferate normally with routine passaging for a period up to D72. In (C) and (D) the stabilised transduction levels at the D72 time point are shown. Data are means ± s.e.mean, n=3. By D72 transduction levels in cultures subject to an initial cell cycle arrest were substantial, and significantly higher than for the cell cultures allowed to proliferate continuously (** P < 0.01 and *** P < 0.001; one-way ANOVA with Tukey's post-hoc test). IDLV-SG = IDLV SV40 GFP; IDLV-SGm = IDLV SV40 GFP mMAR; IDLV-CG = IDLV CMV GFP.*

5.2.4 Long-term transduction of CHO cells by IDLVs following a transient period of cell cycle arrest is a clonally stable phenomenon

To obtain more information about the IDLV-SG, IDLV-SGm and IDLV-CG transduced cell populations which exhibited substantial stable long-term levels of GFP+ cells following the transient induction of cell cycle arrest, a number of clonal cell populations were derived by dilution cloning. CHO cells were transduced as before, subjected to 5 days of methionine and serum depletion, and then allowed to proliferate normally for 100 days. At this point, cells were subjected to dilution cloning procedure, and clones arising after a further 14-day culture period were scored for as either GFP+ or GFP- by fluorescence microscopy. The mixed starting populations were 54%, 12% and 8% GFP+ for the IDLV-SG, IDLV-SGm and IDLV-CG vectors respectively and yielded GFP+ clones at frequencies of 43% (35 out of 80), 14% (17 out of 24) and 6% (5 out of 83). For IDLV-CG, the frequency refers to the first round of dilution cloning (see section 2.17.2).

For each transduced cell population, 12 GFP+ clones were expanded to establish cell lines and allow detailed flow cytometric evaluation. Data for IDLV-SG, IDLV-SGm and IDLV-CG –transduced clones is shown in Figures 5.5, 5.6 and 5.7 respectively.

As seen in Figure 5.5, the parent population of the GFP+ cell transduced with IDLV-SG vector appears slightly bi-phasic and this is reflected in the MFI values for the populations derived by dilution cloning. The majority of the 12 derived clones exhibited a true clonal distribution with close to 100% GFP+ cells and a Gaussian MFI distribution. Clones 4, 7 and 10 were selected for further detailed investigations.

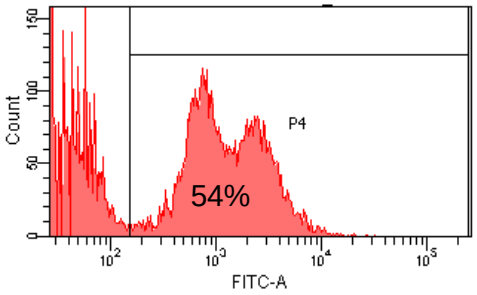
For the IDLV-SGm vector, as seen in Figure 5.6, the original mixed cell population appears more Gaussian but with a wide distribution, which is reflected in the MFI values for the populations derived by dilution cloning. Of these populations, 9 out of 12 exhibited mixed GFP+ and GFP- cell populations suggesting non-clonality. However 3 populations (5, 8 and 11) exhibited a true clonal distribution with close to 100% GFP+ cells and a Gaussian MFI distribution, and were thus selected for further detailed analyses.

For the IDLV-CG vector, as seen in Figure 5.7, the parent population of mixed GFP+ cells did not appear Gaussian but instead exhibited a steadily declining population, which is reflected in the MFI values for the populations derived by dilution cloning.

Of these populations 10 out of 12 exhibited mixed GFP+ and GFP- populations suggesting non-clonality. However, 2 populations (10 and 12) exhibited a true clonal distribution with close to 100% GFP+ cells and a Gaussian MFI distribution, and were thus selected for further detailed evaluations. A third clonal population, clone 2, exhibiting 50/50 pattern of GFP+/GFP- cells was also included for further detailed evaluations.

To evaluate the clonal stability of the IDLV transduction, the 9 selected clonal populations were maintained in continuous proliferating culture for up to 60 days with routine passaging (approximately 3 times weekly in a 1:5 split). The clonal populations were examined every 10 days by flow cytometry to ascertain whether the MFI and %GFP+ cells remained stable. The data is shown in for clones transduced with IDLV-SG, IDLV-SGm and IDLV-CG vectors in Figures 5.8, 5.9 and 5.10, respectively. For the IDLV-SG vector, as seen in Figure 5.8, all clones exhibited consistent %GFP+ cells and MFI levels over 50 days in culture. The %GFP+ cell number varied less than 1% over the period and remained above 95% for all clones. For the IDLV-SGm vector, as seen in Figure 5.9, all 3 clones also exhibited consistent %GFP+ cell numbers over 50 days in culture, although MFI did vary slightly both upwards (clone 8) and downwards (clone 5). The %GFP+ cell number varied less than 3.5% over the period: clones 8 and 11 were consistently >82% GFP+. For the IDLV-CG vector, as seen in Figure 5.11, all 3 clones also exhibited consistent %GFP+ cell numbers over 50 days in culture but did show a tendency towards a decline. For clones 2, 10 and 12, initial GFP+ levels were 63%, 93% and 99% respectively, and by D50 that had changed to 76%, 83% and 83%.

A) CHO IDLV-SG parent population



B) Clones

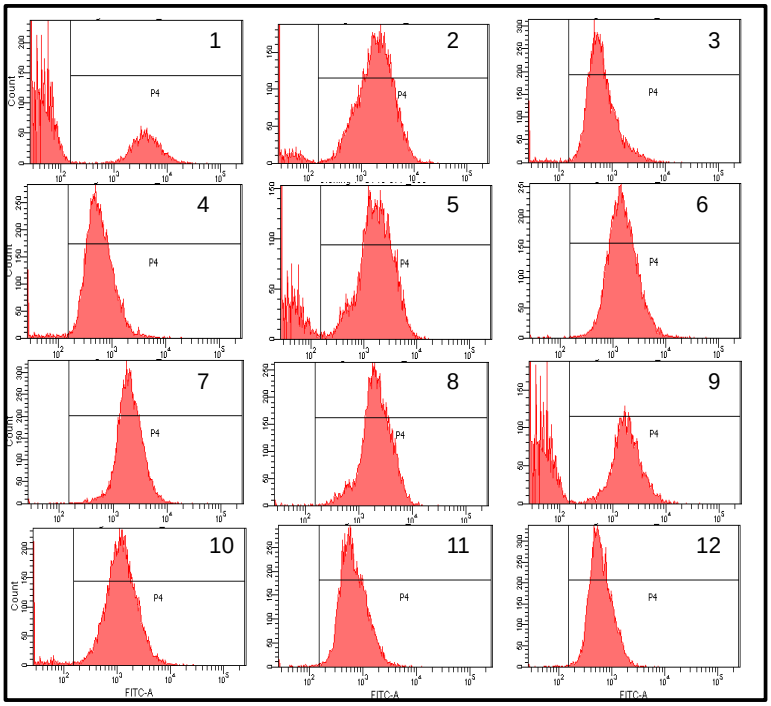
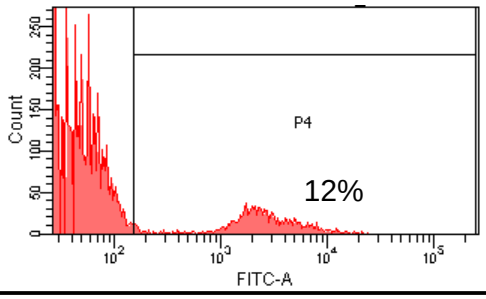


Figure 5.5 Dilution cloning analysis of stable GFP+ CHO cell populations derived from IDLV-SV40-GFP -transduced parent population subjected to a transient cell cycle arrest.

CHO cells were transduced with IDLV-SG vector (MOI=1), subjected to 5 days of methionine and serum depletion, allowed to proliferate for 100 days, and then subjected to dilution cloning in 96-well plates. After 14 days, 12 GFP+ clones were expanded to establish cell lines, and evaluated by flow cytometry. Data is shown for the initial parent population (A) and for the populations derived by dilution cloning (B). The majority of derived populations (9 out of 12) exhibited a true clonal distribution with close to 100% GFP+ cells and a Gaussian MFI distribution. Clones 4, 7 and 10 were selected for further detailed evaluations.

A) CHO IDLV-SGm parent population



B) Clones

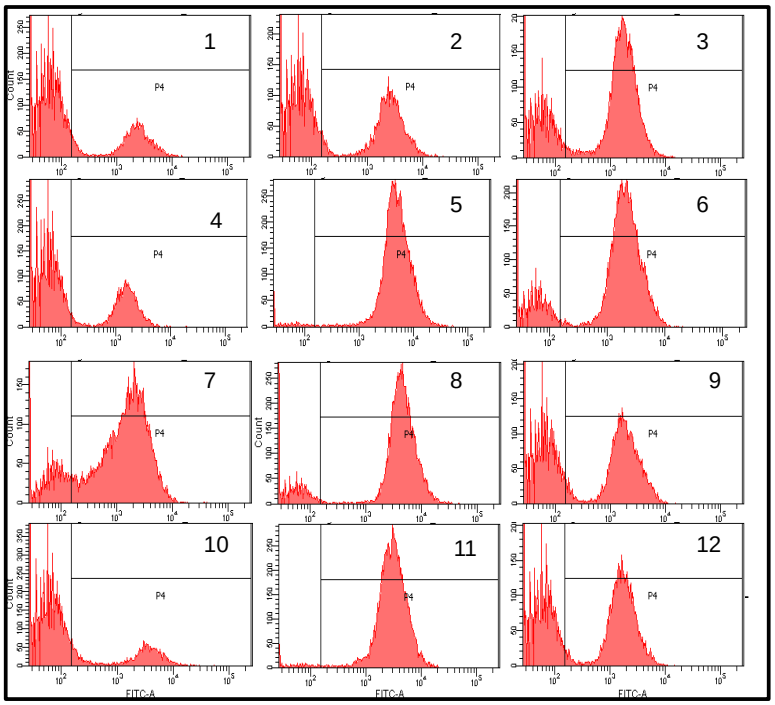
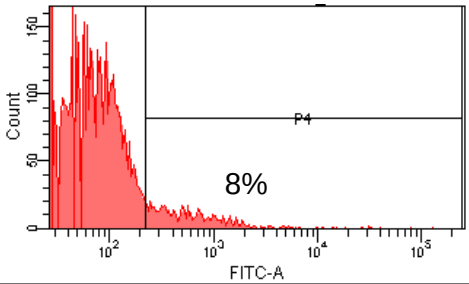


Figure 5.6 Dilution cloning analysis of stable GFP+ CHO cell populations derived from IDLV-SV40-GFP-mMAR -transduced parent population subjected to a transient cell cycle arrest.

CHO cells were transduced with IDLV-SGm vector (MOI=1), subjected to 5 days of methionine and serum depletion, allowed to proliferate for 100 days, and then subjected to dilution cloning in 96-well plates. After 14 days, 12 GFP+ clones were expanded to establish cell lines, and evaluated by flow cytometry. Data is shown for the initial parent population (A) and for the populations derived by dilution cloning (B). Clones 5, 8 and 11 exhibited a true clonal distribution with close to 100% GFP+ cells and a Gaussian MFI distribution, and were selected for further detailed evaluations.

A) CHO IDLV-CG parent population



B) Clones

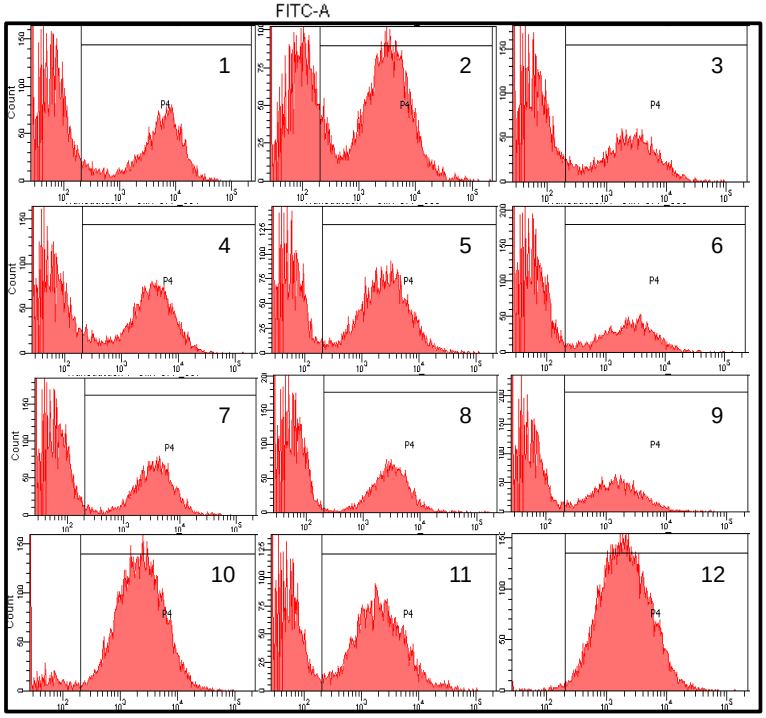


Figure 5.7 Dilution cloning analysis of stable GFP+ CHO cell populations derived from IDLV-CMV-GFP -transduced parent population subjected to a transient cell cycle arrest.

CHO cells were transduced with IDLV-CG vector (MOI=1), subjected to 5 days of methionine and serum depletion, allowed to proliferate for 100 days, and then subjected to dilution cloning in 96-well plates. After 14 days, 5 GFP+ populations were obtained and used for a second round of dilution cloning in 96-well plates. After 14 days, 12 GFP+ clones were expanded to establish cell lines, and evaluated by flow cytometry. Data is shown for the initial parent population (A) and for the populations derived by dilution cloning (B). Clones 10 and 12 exhibited a true clonal distribution with close to 100% GFP+ cells and a Gaussian MFI distribution, and were selected for further detailed evaluations. Clone 2 was additionally chosen for further experiments.

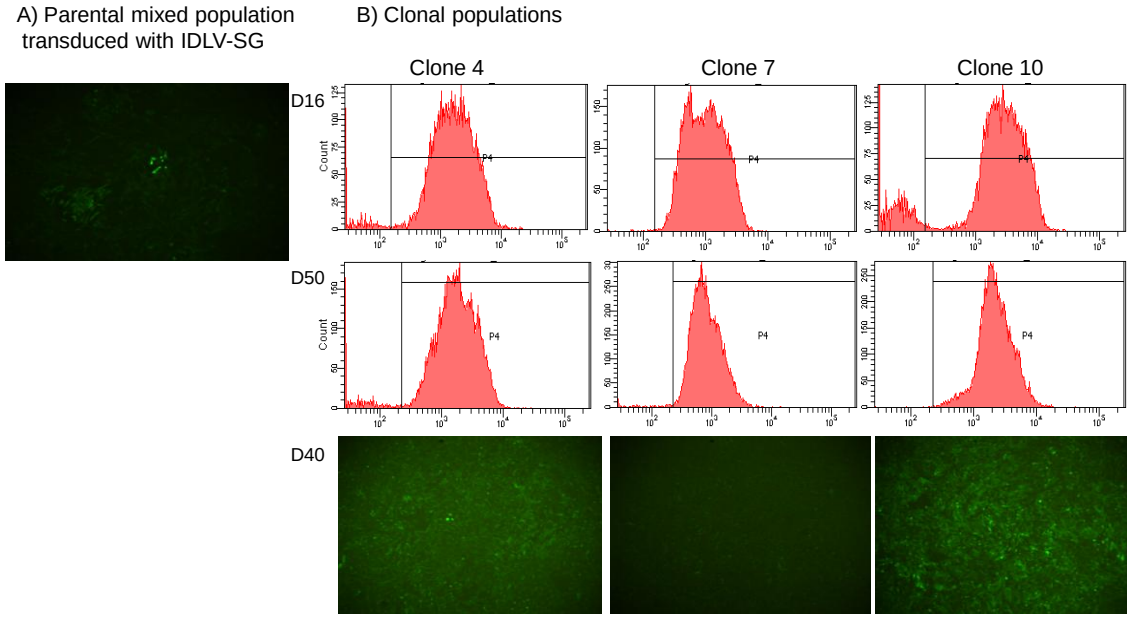


Figure 5.8 GFP expression is stable during long-term proliferation of IDLV-SV40-GFP –transduced CHO cell lines derived by dilution cloning.

CHO cell lines derived by dilution cloning of mixed parent populations transduced with IDLV-SG vector and subjected to transient cell cycle arrest (clones 4, 7 and 10) were maintained in continuous proliferating culture for up to 60 days with routine passaging approximately 3 times weekly at a 1:5 split. The clonal populations were examined every 10 days by flow cytometry and fluorescence microscopy. (A) GFP fluorescence micrograph for the parental population of transduced CHO cells prior to dilution cloning. (B) Flow cytometry plots at D16 and D50 and GFP fluorescence micrograph (D40) for clones 4, 7 and 10.

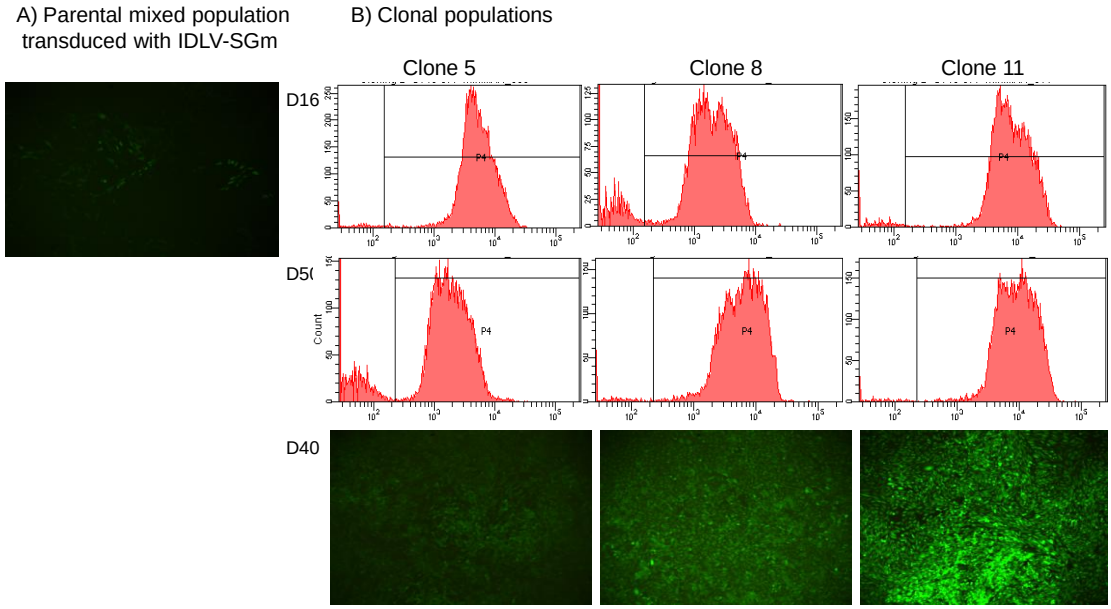


Figure 5.9 GFP expression is stable during long-term proliferation of IDLV-SV40-GFP-mMAR –transduced CHO cell lines derived by dilution cloning.

CHO cell lines derived by dilution cloning of mixed parent populations transduced with IDLV-SGm vector and subjected to transient cell cycle arrest (clones 5, 8 and 11) were maintained in continuous proliferating culture for up to 60 days with routine passaging approximately 3 times weekly at a 1:5 split. The clonal populations were examined every 10 days by flow cytometry and fluorescence microscopy. (A) GFP fluorescence micrograph for the parental population of transduced CHO cells prior to dilution cloning. (B) Flow cytometry plots at D16 and D50 and GFP fluorescence micrograph (D40) for clones 5, 8 and 11.

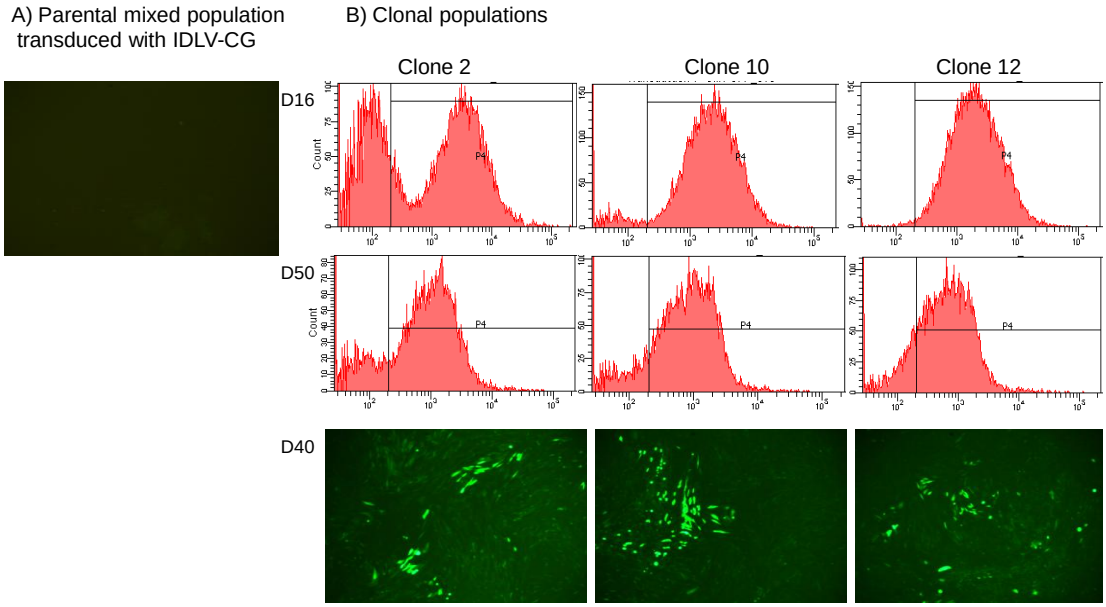


Figure 5.10 GFP expression is stable during long-term proliferation of IDLV-CMV-GFP –transduced CHO cell lines derived by dilution cloning.

CHO cell lines derived by dilution cloning of mixed parent populations transduced with IDLV-CG vector and subjected to transient cell cycle arrest (clones 2, 10 and 12) were maintained in continuous proliferating culture for up to 60 days with routine passaging approximately 3 times weekly at a 1:5 split. The clonal populations were examined every 10 days by flow cytometry and fluorescence microscopy. (A) GFP fluorescence micrograph for the parental population of transduced CHO cells prior to dilution cloning. (B) Flow cytometry plots at D16 and D50 and GFP fluorescence micrograph (D40) for clones 2, 10 and 12.

5.2.5 Evaluation of episomal status of IDLV genomes in CHO cell clones stably transduced following a period of induced cell cycle arrest

The clonal, stably GFP+ cell populations described in section 5.2.4 were derived from transduction with IDLVs, and hence the transgene cassette is unlikely to have integrated into the host genome via integrase-mediated pathways. We decided to employ a battery of techniques to ascertain whether the transgene in these stably expressing cells remains episomal or has been integrated into the host genome.

There are several methods which can be employed to assess the episomal status of vector genomes, each of them with their own caveats. To overcome limitations associated with each technique, we decided to combine several of them and use the combined results to draw final conclusions. To this end, vectors in the transduced nuclei were visualised using FISH, however the small size of the vector cassette meant operating on the close to the limit of sensitivity of the technique. Southern blotting was used to assess vector genome restriction enzyme digest patterns in the highly GFP+ clonal populations, but was not sensitive enough to be used for the polyclonal populations. To increase sensitivity, we applied PCR-based techniques. Quantitative real-time PCR (qRT-PCR) was used to estimate the relative amounts of various vector forms, and finally linear amplification-mediated PCR (LAM-PCR) was used to amplify regions adjacent to the LTRs, which were then subjected to 2^{nd} generation deep sequencing to distinguish non-integrated vector genomes from those associated with host DNA. Although each technique has its own limitations, used together they can provide evidence regarding the integration status of the vectors.

5.2.5.1 Agarose gel analysis and high throughput sequencing of LAM-PCR products amplified from IDLV-transduced CHO cell clones

LAM-PCR is a method whereby unknown DNA sequences adjacent to a known sequence can be amplified for sequencing. It is generally used for analysing viral vector integration sites, and in the case of IDLVs it can be used to estimate the proportional frequency of integrants and episomal forms.

This method was used to analyse the 9 selected GFP+ clonal populations of CHO cells derived from transduction with IDLV-SG, IDLV-SGm and IDLV-CG. Diagrammatic representation of the steps involved in the LAM-PCR protocol is shown in Figure 5.11. Briefly, each sample was subjected to an initial linear PCR reaction starting with the primer binding to the LTR, followed by second strand synthesis using random hexanucleotide primers. The dsDNA products were then digested with the restriction enzyme Tsp509I which cuts at 210 bp from the beginning of the LTR if the strand has been synthesised from an LTR towards the vector genome, or at 105 bp if it has been synthesised from an LTR joined to another LTR. The 210 bp fragments are produced by all vector genome conformations, linear, circular and integrated; the 105 bp fragments are only produced by 2-LTR circles. A variety of fragment lengths will additionally result from an integrated vector genome. The digested dsDNA fragments were ligated to a linker using the overhang created by the restriction enzyme, and the resulting double-stranded DNA molecules with known sequences at either end were subjected to 2 cycles of nested PCR with primers binding to LTR and linker sequences. After the linker ligation and 2 rounds of PCR amplification, the 210 bp fragment becomes ~280 bp long, and the 105 bp fragment becomes ~180 bp long (Figure 5.12).

The products of the second exponential PCR reactions were visualised on a gel to obtain information on the fragment sizes prior to sequencing (Figure 5.12). All clones derived from cells transduced with IDLV-SG produced a band consistent in size (280 bp) with an LTR to vector genome junction. Of the clones derived from cells transduced with IDLV-SGm, two (8 and 11) yielded bands corresponding to the LTR-vector genome junction, although clone 11 also produced a second band. The single band in the lane for IDLV-SGm clone 5 does not correspond to either the LTR-vector genome or the LTR-to-LTR band size. All clones derived from cells transduced with IDLV-CG produced bands corresponding to LTR-vector genome junctions and, interestingly, also bands corresponding to LTR-to-LTR junctions (180 bp).

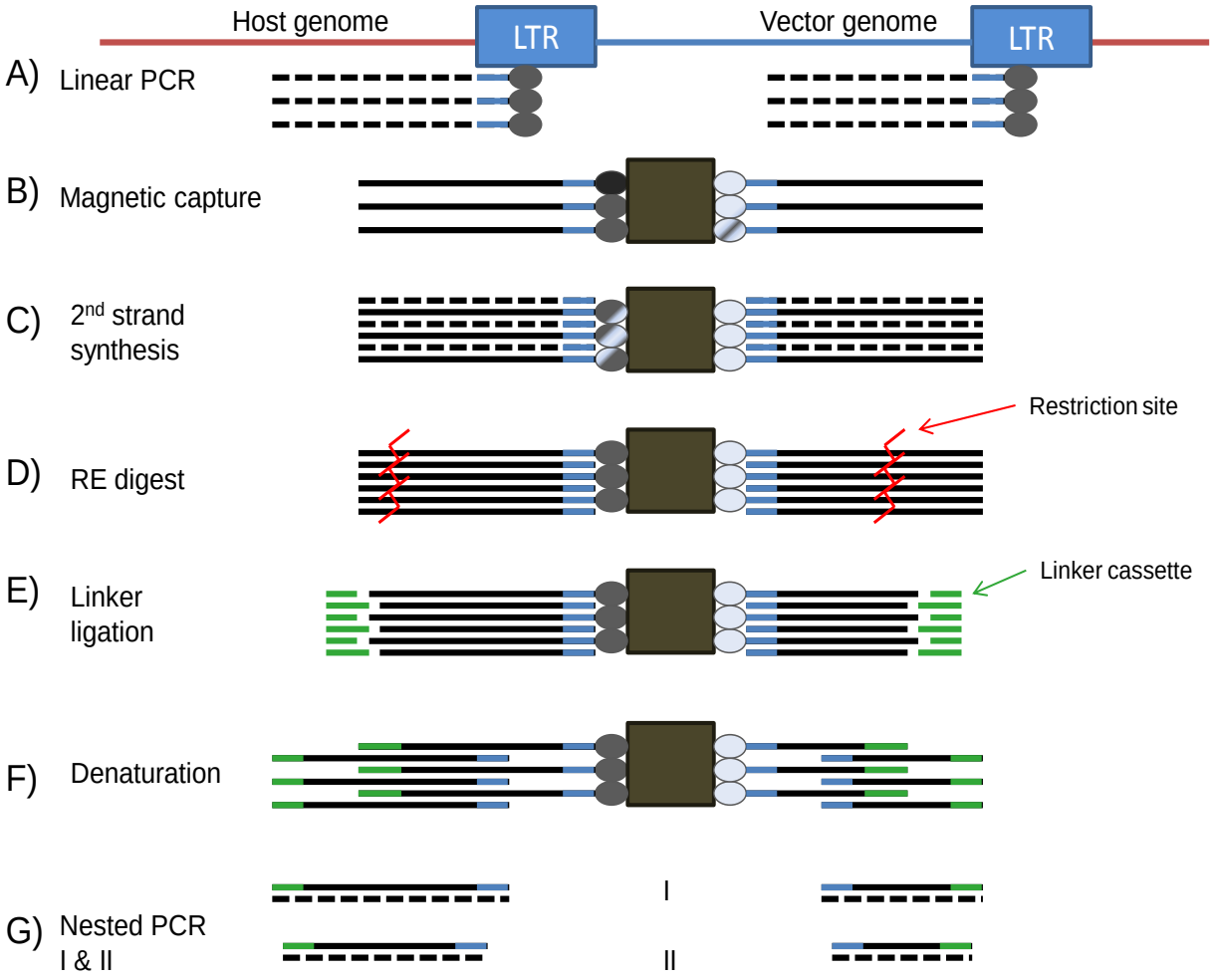


Figure 5.11 Diagrammatic representation of the steps involved in linear amplification-mediated PCR.

A) Target DNA sequences are amplified into linear DNA fragments using a biotin-labelled primer binding to the LTR (biotin denoted by grey oval). B) Amplified DNA fragments are captured on magnetic streptavidin beads (brown squares). Second strand is synthesised using random hexanucleotide primers (C), and the resulting dsDNA digested with a RE that cuts 200-300bp from the LTR into the vector sequence (D), creating defined sizes of vector fragments and/or a variety of fragment sizes from vector-genome junctions. A linker is ligated to the sticky end created by the RE (E) resulting in double-stranded DNA fragments with the LTR sequence at one end and linker cassette sequence at the other end (denoted in green and blue throughout the diagram). The strands are denatured (F) and subjected to 2 consecutive nested PCR reactions using primers complementary to the LTR and linker cassette sequences (G). Modified from (Schmidt et al., 2007).

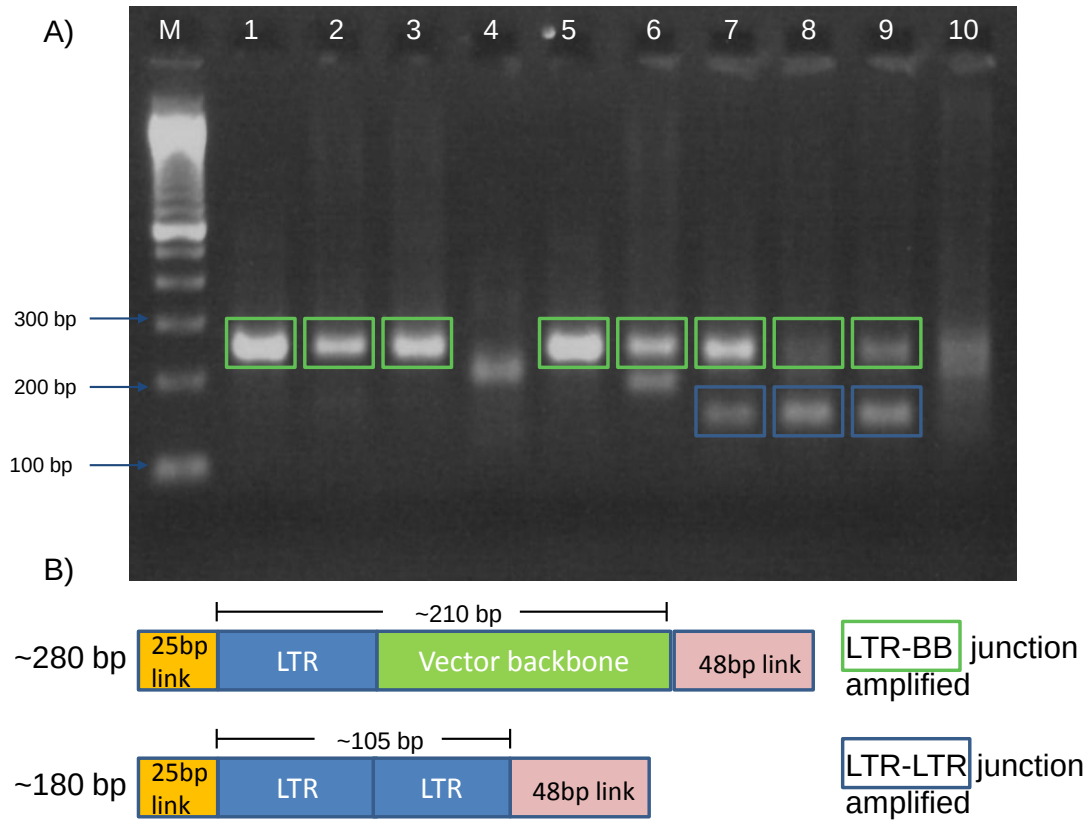


Figure 5.12 The majority of products generated by nested exponential PCR at the final stage of the LAM-PCR protocol are indicative of episomal vector sequences.

Genomic DNA was prepared from the stably GFP+ CHO clones derived from transduction and transient cell cycle arrest with IDLV-SG (clones 4, 7 and 10 on tracks 1, 2 & 3), IDLV-SGm (clones 5, 8 and 11 on tracks 4, 5 & 6) and IDLV-CG (clones 2, 10 and 12 on tracks 7, 8 & 9), and from a polyclonal population transduced with integrating LV-SG (track 10). The DNA was subject to LAM-PCR as described in Fig 5.11, and the final products were analysed by agarose-EtBr gel electrophoresis. Track 1 shows a 100 bp marker ladder.

Dominant bands at 280 bp (green boxes) correspond to LTR to vector backbone junctions present in all vector forms, 180 bp bands (blue boxes) correspond to LTR-LTR junctions only present in 2-LTR circles. The integrating clonal control (track 10) shows a smear of bands of varying sizes, as expected.

The PCR products from LAM-PCR were subjected to high throughput sequencing, resulting in an average of 1200 reads per sample. Three samples resulted in less than 50 sequence reads (IDLV-SG clone 4, IDLV-SGm clone 5, and IDLV-CG clone 10), and were discounted from further analysis. The sequencing results were divided into two categories, the first containing only vector DNA, and the second non-vector DNA (Figure 5.13).

The majority of LAM-PCR sequencing reads for the CHO clones transduced with IDLV-SG and IDLV-SGm yielded only vector sequences. In contrast, a significant proportion (12-24%) of CHO clones transduced with IDLV-CG yielded non-vector sequences. To investigate the sequence reads classed as non-vector (red sections of bars in Figure 5.13), they were aligned against the mouse and rat genomic sequences using the basic local alignment tool (BLAST). Murine sequences were used due to the absence of an available sequence for the Chinese hamster genome at the time of the experiment.

The non-vector sequences generated by IDLV-SG clone 7 and IDLV-SGm clone 11 did not yield any BLAST hits. This suggests that the non-vector sequences in these samples were the result of contamination from surrounding lanes containing HIV-transduced human cells as the samples were sequenced in a facility usually processing such samples. The non-vector sequences in IDLV-SG clone 10 and IDLV-SGm clone 8 did result in some BLAST hits, however these were excluded after further analysis established them as too partial and fragmented to be likely true matches. Of all the IDLV samples, only IDLV-CG clone 2 contains a single likely integrant, at a site corresponding to the mouse chromosome 4. The integrating control sample resulted in high read counts for 2 integration sites, one corresponding to the murine chromosome 1 and the other to chromosome 2. The one read from chromosome 1 detected in IDLV-CG clone 12 is identical to hundreds of reads for the same site on the integrating control, and hence was discounted as contamination (Table 5.1).

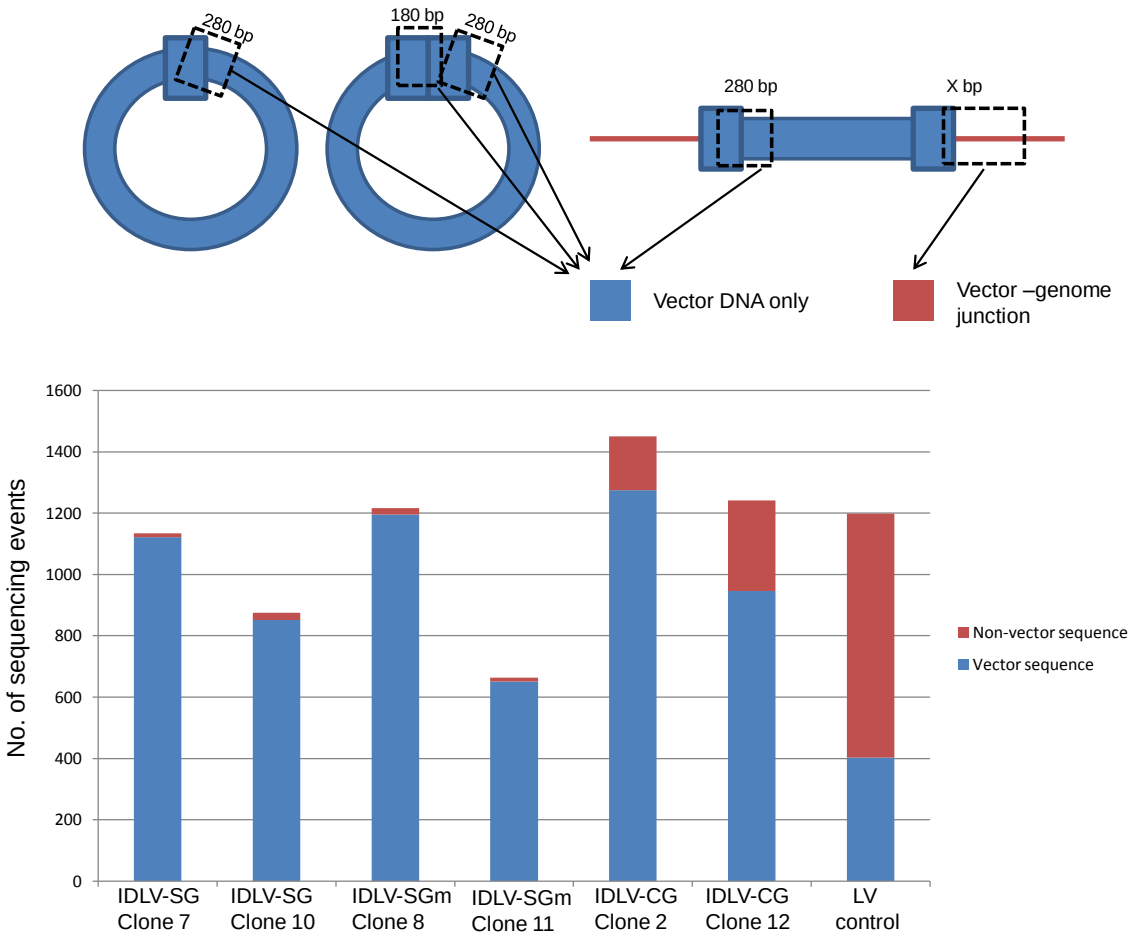


Figure 5.13 The majority of reads generated by high-throughput sequencing of LAM-PCR products from clonal IDLV-transduced CHO populations contain only vector sequences.

Genomic DNA was prepared from the stably GFP+ CHO clones derived from transduction followed by transient cell cycle arrest, and cells transduced with an integrating LV control. The DNA was subject to LAM-PCR as described in Fig 5.11, and the products were analysed by high-throughput deep sequencing and aligned against each vector sequence. The histogram shows the numbers of sequence reads generated by each sample either containing only vector sequences (blue) or also other (red). Samples generating <50 analysable sequences were disregarded. The population transduced with the integrating control vector (LV control) generated high amounts of non-vector sequence reads.

Following the publication of a draft genomic sequence for the Chinese hamster, the sequences corresponding to loci in the mouse chromosomes 1, 2 and 4 were verified by BLAST analysis to be present also in Chinese hamster. As the draft genome is not yet annotated, no information could be retrieved regarding the location of these sites in hamster (Xu *et al.*, 2011).

	Total Reads	Vector	Verified Host	Host location (mouse)
IDLV-SG clone 7	1191	1122	0	-
IDLV-SG clone 10	906	852	0	-
IDLV-SGm clone 8	1269	1196	0	-
IDLV-SGm clone 11	1010	651	0	-
IDLV-CG clone 2	1500	1275	1	Chr 4
IDLV-CG clone 12	1299	946	1	Chr 1
LV control	1279	403	248	Chr 1, 239 reads; Chr 2, 9 reads

Table 5.1 BLAST analyses of sequences generated by high-throughput sequencing of CHO clones transduced with IDLVs reveal very few integration events.

Genomic DNA from the stably GFP+ CHO clones and cells transduced with an integrating LV was subject to LAM-PCR as described in Fig 5.11. The products were analysed by high-throughput deep sequencing followed by BLAST analysis of all generated sequences against the vector genome and the mouse genomic sequence as the closest available sequenced relative of the Chinese hamster. The table shows numbers of successful sequencing reads per sample (Total Reads), and numbers of reads containing only vector sequences (Blue) and other sequences verified by BLAST analysis against the mouse genome (Red). The green column indicates the corresponding location of the BLAST result in the mouse genome.

5.2.5.2 Copy number and conformation analysis of vector genomes in IDLV-transduced CHO cells by quantitative real-time PCR

To gather additional data on episome numbers and integration status, DNA extracted from the polyclonal parental samples of IDLV-transduced CHO cells subjected to a transient cell cycle arrest and from the 3 clonal samples derived from each parent population was analysed using quantitative real-time PCR (qRT-PCR) to quantify vector sequences present in the samples. qRT-PCR allows for a simultaneous

amplification and quantification of target DNA using a fluorescently labelled reporter. Here, the method was used to quantify 3 different vector sequences: i) vector backbone adjacent to the LTR, which should be present once per vector molecule in all vector-containing samples; ii) LTR, which should be present once or twice per vector molecule in every sample; and iii) LTR to LTR junction, which should be present in samples containing 2-LTR circles (Figure 5.14 A). The quantities of these 3 sequences were normalised to the amount of cellular DNA present in each sample by quantifying the amount of CHO β-actin gene.

High molecular weight DNA from both polyclonal and clonal cell populations was subjected to all 4 qRT-PCR reactions, and the copy number of each of the vector sequences was then calculated per cell using β-actin data to estimate number of cell equivalents per sample (Figure 5.14 B). Each vector genome is expected to produce one backbone (BB) signal, one or two LTR signals, and one or none LTR-to-LTR signals, depending on the conformation and integration status. The amount of BB signal can be used to estimate vector copy number and the ratios of the 3 independent vector signals can be used to infer conformation and integration status.

Figure 5.14 B shows the amounts of the 3 independent vector signals as calculated per cell equivalent. Most samples resulted in low amounts of vector signal, corresponding to 1-15 vector genomes per cell. The 3 polyclonal samples indicated average vector copy numbers of 1.6/cell for IDLV-SG, 1.1/cell for IDLV-SGm and 0.6/cell for IDLV-CG, which reflects the differing percentages of GFP-expressing cells in these populations at the time of investigation (Figure 5.4 C).
The clonal populations transduced with IDLV-SG averaged 2.8 (Clone 4), 0.6 (Clone 7) and 1.1 (Clone 10) backbone copies per cell. For cells transduced with IDLV-SGm, the average backbone copy numbers were 0.1 (Clone 5), 4.8 (Clone 8) and 6.0 (Clone 11) per cell. The average BB copy numbers for clonal populations derived from IDLV-CG –transduced cells were 4.4 (Clone 2), 1.8 (Clone 10) and 1.4 (Clone 12) signals per cell.

Three controls were included in the experiment: DNA extracted from CHO cells transduced with IDLV-SG and integrating LV-SG 24 h after transduction, and a clone derived from cells transduced with integrating LV-SG at MOI=2 by FACS (approximately corresponding to MOI=200 by qPCR). All 3 yielded high amounts of

vector signal, with vector copy numbers estimated at 35, 37 and 45 for IDLV, ICLV and ICLV clone, respectively. The ICLV-transduced clone did not yield any 2-LTR signal.

In addition to data on vector copy numbers, the qPCR data was used to provide information on vector genome conformation. This was based on the following principles:

i) If the vector has integrated via the integrase-mediated pathway, two copies of the LTR signal should be present for each of the backbone signals, with no LTR-to-LTR signal present (first row of Table 5.2).

ii) If the vector is present as episomal 1-LTR circles, the ratio of LTR to backbone signals should be 1:1 with no LTR-to-LTR signal present (second row of Table 5.2).

iii) If the vector is present mostly as 2-LTR episomes, each vector genome should produce one LTR-LTR signal, two LTR signals, and one backbone signal (third row of Table 5.2).

The ratios of 1-LTR and 2-LTR signals per copy of backbone signal were calculated for each vector (Table 5.2). Only one sample produced a ratio suggesting a possible Integrase-mediated integration event; IDLV SV40 GFP clone 4 gave a ratio of 2:1 of LTR:BB (2.07:1 averaged over all data points). The remaining clones produced ratios indicative of either 1-LTR or 2-LTR episomes.

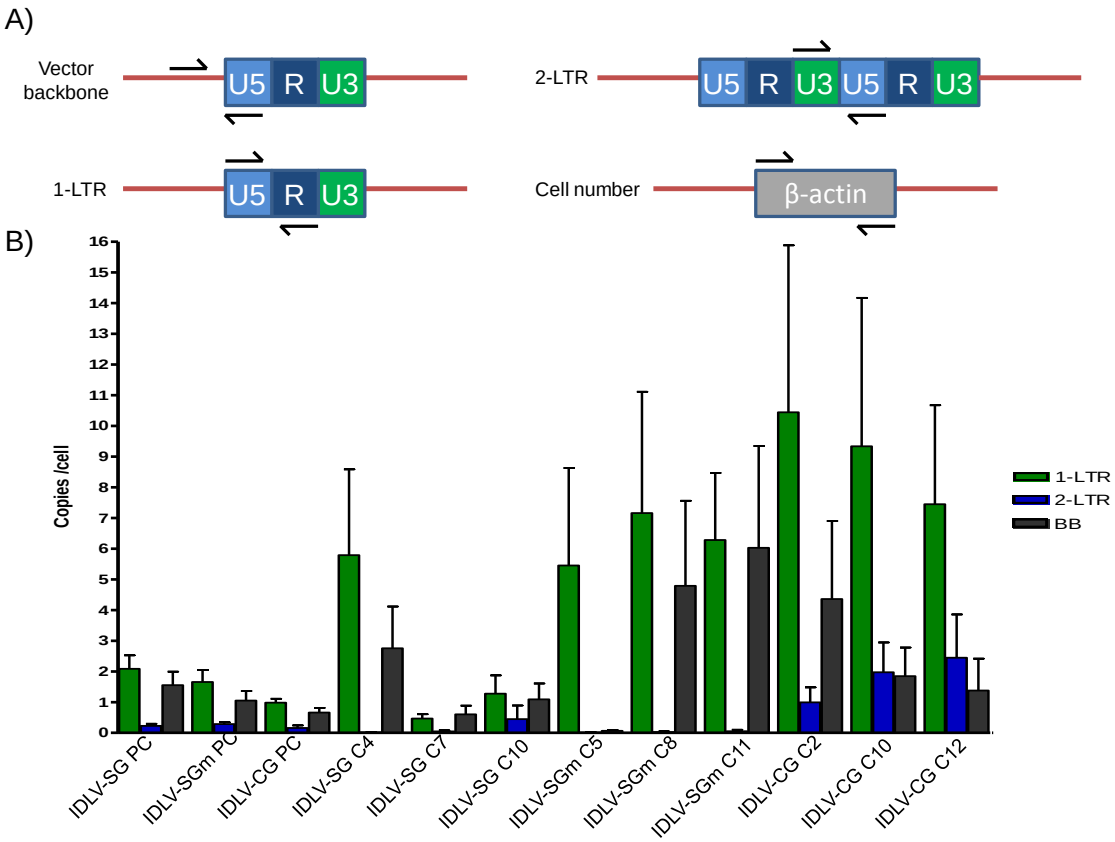


Figure 5.14 Quantitative real-time PCR results indicate IDLV vector genomes persist in copy numbers of 1-15 per cell in stably GFP+ transduced CHO cells.

DNA was extracted from the stably GFP+ IDLV-transduced CHO cells subjected to a transient cell cycle arrest, and the derived clonal populations. Four independent qRT-PCR reactions were set up to quantify the amount of vector genomes present in the cells; 3 different sections of the vector DNA were amplified in addition to Chinese hamster β-actin. Primer binding sites are shown with black arrows (A). From the absolute quantities of these signals, the amounts of 1-LTR, 2-LTR and backbone signals per cell were calculated using the amount of β-actin signal to set cell number (B). PC = polyclonal, C = clone. Data shown are the mean ± s.e.m, n = 9.

Clone	BB	1-LTR	2-LTR	Inferred Conformation
INTEGRATED	1	2	0	
EPISOMAL 1-LTR CIRCLE	1	1	0	
EPISOMAL 2-LTR CIRCLE	1	2	1	
IDLV- SV-GFP Clone 4	1	2.07	0.01	Integrated
IDLV-SV-GFP Clone 7	1	0.65	0.07	1-LTR Episomal
IDLV-SV-GFP Clone 10	1	1.12	0.42	1-LTR & 2-LTR Episomal
IDLV-SV-GFP mMAR Clone 5	N/A	N/A	N/A	N/A
IDLV-SV-GFP mMAR Clone 8	1	1.45	0.01	1-LTR Episomal
IDLV-SV-GFP mMAR Clone 11	1	0.96	0.01	1-LTR Episomal
IDLV-CMV-GFP Clone 2	1	2.35	0.23	2-LTR Episomal
IDLV-CMV-GFP Clone 10	1	5.02	1.07	2-LTR Episomal
IDLV-CMV-GFP Clone 12	1	5.14	1.80	2-LTR Episomal

Table 5.2 Ratios of vector signals calculated from the amounts of qRT-PCR product suggest the majority of vectors persist in episomal forms.

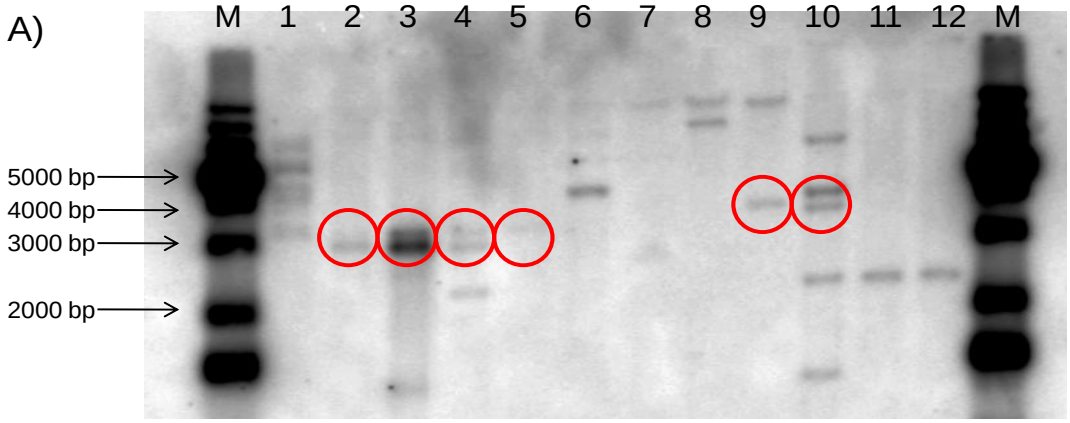
DNA was extracted from clonal IDLV-transduced CHO cells subjected to a transient cell cycle arrest, and 4 independent qRT-PCR reactions were performed to quantify the amounts of BB, 1-LTR, 2-LTR and β-actin signals per sample. Absolute amounts of qPCR signals were normalised for cell number, and used to calculate the ratios of 1-LTR and 2-LTR signals per copy of backbone signal in each sample. The expected ratios of signals for each of the 3 possible vector conformations are shown in the first 3 rows (blue). The predominant conformations for each sample as inferred from the signal ratios are shown in the right hand column. The qRT-PCR reaction for the BB signal in clone 5 of cells transduced with IDLV-SGm yielded a very low amount of product and the ratios could therefore not be reliably calculated. BB: Vector backbone (see Fig 5.14 A).

5.2.5.3 Integration status analysis of vector genomes in stably transduced CHO cell clones by Southern blotting

Southern blotting is a method that can be used to detect the presence and abundance of a known DNA fragment within a sample, as well as to inspect the integration status of the sequence by probing DNA digested with a restriction enzyme. Here, high molecular weight DNA extract from transduced cells was subjected to a restriction enzyme digest that cuts once within the vector genome and separated on an agarose gel. The DNA was then transferred onto a membrane and probed with a radioactive probe. If episomal, the resulting bands should correspond to the expected size of the episome; if integrated, each integration event should produce a band not likely to correspond to the expected episome size.

The results show multiple integration sites in the clonal control transduced with integration-competent SV40 GFP control vector (Figure 5.15, lane 1). The controls in lanes 2 and 3 contain DNA harvested 24 hours after transduction with IDLV-SG and integrating LV-SG, and both show bands corresponding to 1-LTR and 2-LTR circles.
Lanes 4-6 contain DNA from clonal populations transduced with IDLV-SG; clone 4 shows bands corresponding to both episomal forms and clone 7 shows a faint 2-LTR sized band. Of the 3 clonal populations transduced with IDLV-SGm in lanes 7-9, clone 11 shows a band corresponding to the 1-LTR episome (lane 9). The DNA in the other two clones is potentially only partially digested and hence has not completely entered the gel. Clone 2 of IDLV-CG (lane 10) shows a band corresponding to 1-LTR and 2-LTR episomes in addition to several other bands. All IDLV-CG clones have the same band at ~2200 bp, which is too small to contain the entire episome and may therefore indicate either rearrangement or integration.

A)



B)

Lane	Sample	Expected 1-LTR band	Expected 2-LTR band
1	**ICLV** clonal control	3049 bp	3283 bp
2	IDLV SV40 GFP 24h control	3049 bp	3283 bp
3	**ICLV** 24h control	3049 bp	3283 bp
4	IDLV SV40 GFP clone 4	3049 bp	3283 bp
5	IDLV SV40 GFP clone 7	3049 bp	3283 bp
6	IDLV SV40 GFP clone 10	3049 bp	3283 bp
7	IDLV SV40 GFP mMAR clone 5	3782 bp	4016 bp
8	IDLV SV40 GFP mMAR clone 8	3782 bp	4016 bp
9	IDLV SV40 GFP mMAR clone 11	3782 bp	4016 bp
10	IDLV CMV GFP clone 2	3771	4005
11	IDLV CMV GFP clone 10	3771	4005
12	IDLV CMV GFP clone 12	3771	4005

Figure 5.15 Southern blotting of DNA from IDLV –transduced CHO cells indicates several clones containing episomal vectors.

High molecular weight DNA was harvested from cells transduced with vectors, and digested with either EcoRI or XhoI (mMAR-containing vector only). 10 µg digested DNA was loaded per lane, and probed with ^{32}P –labelled GFP probe. The expected band sizes for each of the 2 episomal conformations are shown in table 5.11 B. Grey squares indicate the presence of bands corresponding to the expected size(s), circled in red.

5.2.5.4 Conformation and copy number analysis of vector genomes using fluorescent in situ hybridisation

FISH is a cytogenetic technique used commonly to scan for relatively large chromosomal abnormalities, but some successes have also been reported with the detection of smaller single-copy targets (Rupprecht *et al.*, 2009). In order to examine the cytogenetic distribution of the GFP transgenes in the CHO cell clones stably transduced with IDLVs following a period of induced cell cycle arrest, FISH analyses were performed on 13 cell populations: IDLV-SG –transduced polyclonal population and clones 4, 7 and 10, IDLV-SGm –transduced polyclonal population and clones 5, 8 and 11, two control populations transduced with IDLV-SG and integrating LV-SG and fixed 24 h after transduction, a control polyclonal population transduced with integrating LV-SG and 3 derived clones, and untransduced control population.

The samples were analysed in 2 different ways: firstly, a qualitative analysis of signal quality on FISH metaphase images was performed. Although the cells were not synchronised prior to fixing, at least 20 clear metaphases were obtained on each slide, and 10-20 were photographed for further analysis. Secondly, one clone transduced with each of the 3 vectors IDLV-SG clone 4, IDLV-SGm clone 11, and integrating LV-SG clone 6n as well as a negative control were chosen for further analysis. The number of signals and nuclei were counted in 20 fields across the whole slide to account for variation within slides and minimise observer bias. Statistical analyses were then applied to the obtained data to investigate differences between the samples.

The FISH probes used in the experiment were based on the IDLV-SG transfer plasmid, labelled either with red or with green fluorochromes. An equal amount of each probe was used simultaneously for the *in situ* hybridisation, and only signals positive for both colours were accepted as true signals. This was done to minimise the possibility for overlap between background and true signals, which was important due to the small size of the target sequence. In the negative control slide containing untransduced cells, some signals containing only one colour were observed. In contrast, the majority of signals arising in transduced cells are yellow (Figure 5.16).

As the IDLV-CG sequence was not sufficiently similar to the IDLV-SG for the same probe to be used for cells transduced with IDLV-CG, a separate pair of red and green probes were labelled for use in these cells. However, the probe consistently resulted in

very high background, blurring any true signals, and therefore neither the IDLV-CG – transduced polyclonal nor the clonal populations were included in further analyses.

We also performed a statistical analysis to assess the specificity of the FISH signal. The samples included in the analysis were the selected clonal cell lines transduced with IDLV-SG, IDLV-SGm and integrating LV-SG, as well as an untransduced negative control. For each sample slide, 20 fields from every part of the slide were included, and for each field the number of signals within nuclei, outside nuclei, the total number of nuclei and the number of empty nuclei were counted. The ratio of signals inside nuclei versus outside nuclei was calculated for each field. A Student's t-test shows that each sample slide is significantly different to the negative control slide (P <0.0001), which supports the conclusion that the signals arise from the target sequence rather than background (Figure 5.17 A). The average numbers of signals were also calculated for the three samples, with similar results for all of the clones, varying from ~1.6 per nucleus for both IDLV-SG and integrating LV-SG, to ~1.9 per nucleus for IDLV-SGm.

An integrated transgene is expected to produce a doublet signal where the signal is present at the same location on both sister chromatids. Only 1 doublet signal was found in a total of 52 metaphase images taken of cell populations transduced with IDLVs. In contrast, 8 clear doublets were observed in a total of 35 metaphase images taken of clonal samples transduced with integrating LVs (Figure 5.17 B). The metaphase images therefore provide supportive evidence for the presence of episomes in the cells transduced with non-integrating vectors.

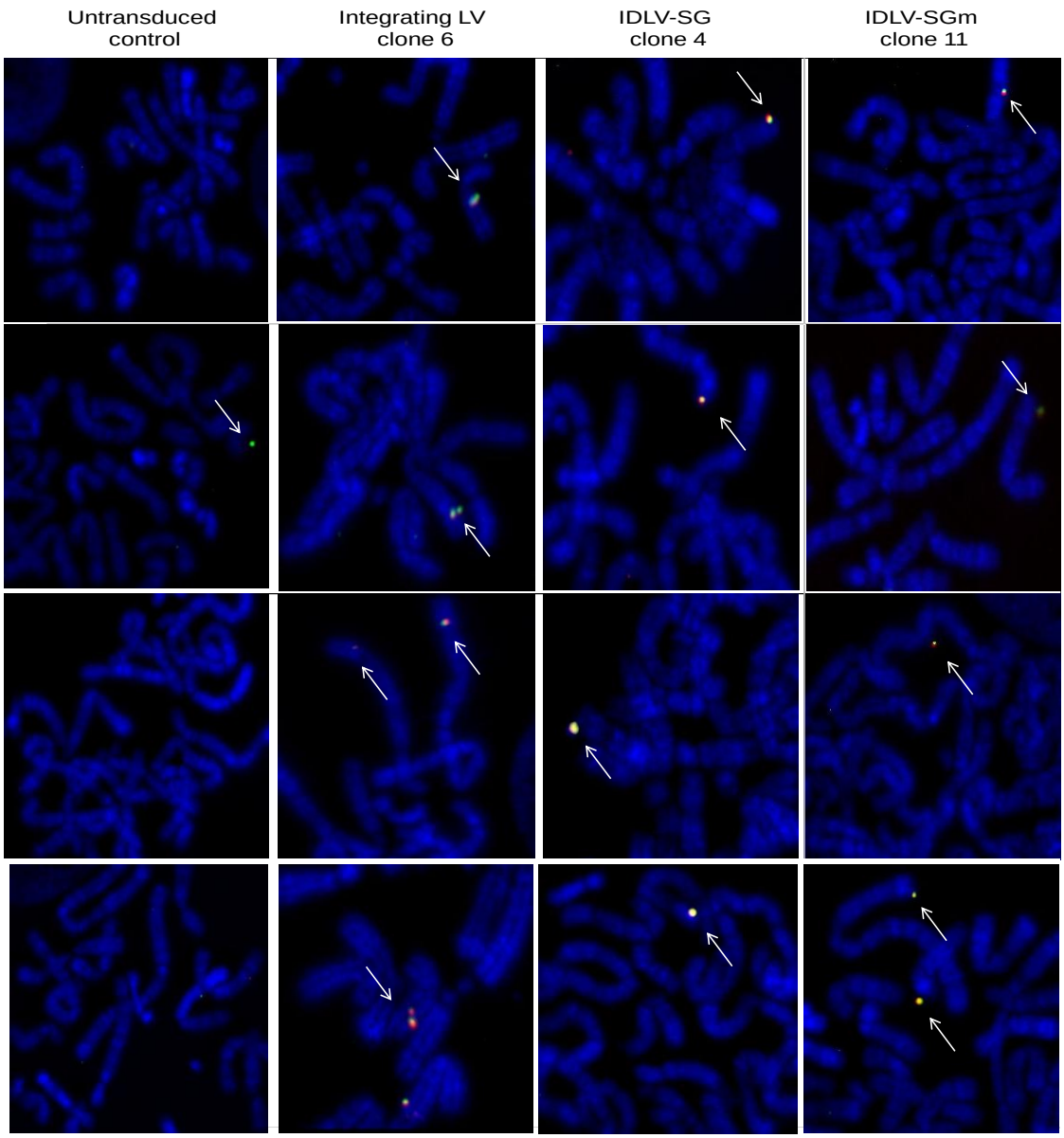


Figure 5.16 Fluorescent in situ hybridisation analysis of IDLV and integrating LV -transduced cells supports the nuclear episome hypothesis.

Clonal CHO cells transduced with an integrating LV-SG (clone 6,) IDLV-SG (clone 4), IDLV-SGm (clone 11) and a non-transduced control population were fixed on slides and the vector genomes were localised with probes labelled with both red and green fluorochromes. At least 20 metaphases were located on each slide. Several doublet signals indicative of integration events can be seen in the integrating control slide. Single strong signals, often located near telomeres, can be seen in IDLV-transduced cells. The second image from the negative control slide shows a typical background spot positive for only one colour.

Average diameter of a metaphase, 30 µm (partial metaphases shown).

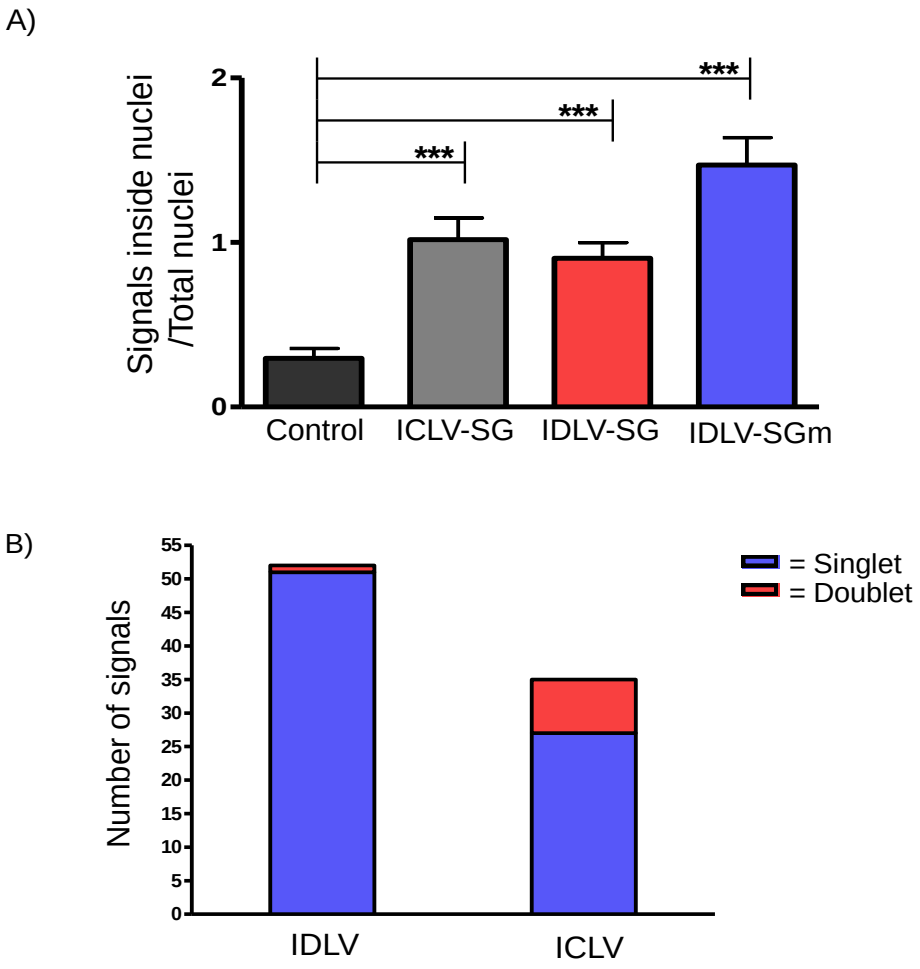


Figure 5.17 Fluorescent in situ hybridisation analysis of transduced cells supports the hypothesis that a majority of vector genomes remain episomal.

CHO cells were transduced with IDLV-SG and IDLV-SGm and subjected to an initial cell cycle arrest, a control population was transduced with integrating LV-SG, and clonal cell lines were derived from all 3. The clonal cells and a non-transduced control were fixed on slides and the vector genomes located with fluorochrome-labelled probes. (A) Signal specificity. The number of signals inside nuclei and the number of total nuclei were counted in 20 fields (n=20) for each population. A statistical analysis comparing the ratios of signals occurring inside nuclei per total nuclei in each sample shows a significant difference between each sample and negative control (t-test, P< 0.0001).

(B) Singlet and doublet signals in IDLV and integrating LV populations. A total of 52 images of metaphases in IDLV-transduced cells and a total of 35 images in integrating LV-transduced cells were acquired. The numbers of visible doublets in the metaphase images were 1 and 8, respectively.

5.3 Discussion

In this chapter, the possibility of producing mitotically stable episomes was explored via transduction with non-integrating lentivectors. Three different IDLVs were used in the experiment; two of them containing the SV40 immediate early promoter, and one containing the CMV promoter, both driving the eGFP reporter gene. One of the SV40-driven vectors also contained a truncated miniMAR element. The combination of SV40 and S/MAR elements has previously shown success, and hence it was also tested here on the IDLV platform (Piechaczek *et al.*, 1999).

Another variable introduced into the transduction experiment was the induction of a period of transient cell cycle arrest 24 h after transduction. A modest positive effect on the retention levels of S/MAR –containing plasmids (Chapter 3, Figure 3.4) and ssAAV vectors (Chapter 4, Figure 4.5) was observed after an initial cell cycle arrest; it was therefore important to ascertain whether any effect would be observed in the context of lentivectors.

5.3.1 SIDD analysis does not reveal a significant impact on destabilisation profile upon addition of S/MAR element

Prior to embarking on the *in vitro* experiments described in this chapter, an *in silico* analysis of the proposed constructs was conducted in order to ascertain in advance the suitability of the constructs as well as provide a basis for interpretation of the results.

The SIDD analysis uncovered no difference in the destabilisation profile between the 1-LTR and 2-LTR circular episomes, indicating that neither of these forms is likely to be favoured over the other for long-term episome retention. The addition of the IFN-S/MAR or the miniMAR element caused duplex condensation across the rest of the circular episome, in accordance with other circular structures investigated thus far (Chapter 3, Section 3.2.1). This could be predicted to result in decreased transgene expression due to promoter condensation, or decreased episomal replication ability if regions other than the S/MAR element acting as the primary origins of replication are also condensed.

5.3.2 Evaluation of the effect of DNA elements and cell culture conditions on generation of mitotically stable episomes

5.3.2.1 The effect of the S/MAR element on episome retention

The truncated miniMAR –element was included in the IDLV-SV40-GFP vector cassette immediately following the transgene to provide the episome with a potential replication origin. However, the presence of the miniMAR in the vector cassette did not increase the level of retention of the vector, either when combined with an initial cell cycle arrest or when cells were kept in continuously proliferating culture. The mean fluorescence intensity was also not altered by the inclusion of the miniMAR, suggesting that the augmentation of transcription levels does not occur when combining S/MAR elements with IDLVs.

Furthermore, the SIDD analysis indicated that the miniMAR element would have a detrimental effect on the function of a replication origin if the miniMAR itself was not the primary Ori. This hypothesis was experimentally supported, as a lower episome establishment frequency was indeed observed for the mMAR-containing vector when compared to the equivalent vector without a mMAR-element (Figure 5.4). Thus it can be inferred that elements other than the S/MAR are likely to act as replication origins in the stably retained IDLV episomes.

5.3.2.2 The effect of a period of induced cell cycle arrest on episome retention

A transient period of cell cycle arrest was induced following IDLV transduction in order to allow for potential episome establishment to take place in the nuclear environment prior to the diluting out of the IDLV genomes due to the rapid cycling of the CHO cells.

The transduction of CHO cells followed by a 5-day period of cell cycle arrest induced by methionine depletion resulted in a substantial percentage (15-50%) of cells expressing the transgene in a mitotically stable fashion. Previously reported levels of stably expressing cells achieved with S/MAR-element-containing episomal vectors are up to 10-fold lower at 5% (Broll *et al.*, 2010) in the absence of selection pressure.

Again, it is unlikely that the difference between the populations either subjected to a cell cycle arrest or kept in continuous culture arise purely from differences in the number of population doublings. As CHO cells divide every 16 hours, by 72 days post-transduction the difference between the arrested and the control population is 8 doublings (100 and 108). Assuming that, by chance, one of the daughter cells retains an episome at each doubling, even then the percentage of positive cells would be expected to drop from 100% to 12.5% in only 4 population doublings. It is therefore reasonable to assume that the episome is actively being retained in the cell cycle arrested populations.

There are two possible explanations for the unprecedentedly high retention of the transgenes in the transduced cells initially held in cell cycle arrest: either the viral episomes became associated with the nuclear matrix during cell cycle arrest and developed into mitotically stable entities, or the cell cycle arrest induced a higher than usual rate of integration. The evidence for and against the episome hypothesis is discussed in section 5.3.3.

For the episomal retention hypothesis to be supported, a functional origin of replication must be present in the IDLV circular episomes. As the presence of the miniMAR – element did not increase retention levels, it is unlikely to act as the Ori. Furthermore, the mitotic segregation potential of the IDLVs only becomes established upon a period of residence in a nuclear environment which is arrested at the G_1/M checkpoint. The exact contents of the IDLVs did not appear to have a significant effect on the vector genome retention, and conversely, without the cell cycle arrest none of the IDLVs was capable of mitotic stability. This suggests that the replication potential is related to a feature present in all lentiviral genomes, which then interacts with cellular elements during the cell cycle arrest. Based on the SIDD analysis, the most strongly destabilised vector backbone region that is present in all vector genomes is the central polypurine tract (cPPT) (see Figure 5.2). Whether this region alone or in combination with others is capable of inducing mitotic segregation remains to be investigated.

The significance of the method for inducing cell cycle arrest as well as the duration also remain to be explored. It is possible that an arrest for 4-5 days at the G_1/M checkpoint is crucial; likewise, it is possible that any protocol that causes the cells to enter a reversible cell cycle arrest following transduction will achieve a similar result.

Although the method used for inducing cell cycle arrest was relatively gentle and did not involve the use of any chemicals known to cause DNA damage, it is possible that the cellular DNA repair machinery for some reason became more active during cell cycle arrest, increasing the amount of vector DNA integration. One of the most consistent characteristics of CHO cells appears to be genome instability, as demonstrated by several studies into the genome structure of cell lines in laboratory use. Karyotypic rearrangements have been found both on the basis of banding (Deaven *et al.*, 1973) and, more recently, upon constructing a chromosomal map by using CHO genomic BAC library clones as FISH hybridisation probes (Omasa *et al.*, 2009). However, confirming rearrangements at a sequence level, whether in CHO-K1 cells in general or our clonal cell lines in particular is not currently straightforward since the Chinese hamster genome remains to be annotated. This propensity for genomic rearrangements may also indicate a tendency towards random integration of any episomal material present. Interestingly, gene amplification in another CHO –derived cell line has been found to result in the amplified gene nesting between two inverted genomic regions (Omasa *et al.*, 2009). One method by which large genomic rearrangements could be examined is FISH. The pre-requisite for constructing probes to look for such events is knowledge of the genomic sequence and chromosome structure of the cells under scrutiny. Most likely the recently published Chinese Hamster draft genomic sequence will soon be annotated, and the FISH approach will be possible as soon as chromosome structure information becomes available.

5.3.2.3 Clonal stability of GFP expression in IDLV-transduced CHO cells

GFP expression in the CHO cells transduced with IDLVs and subjected to a period of transient cell cycle arrest stabilised at high levels at 15-50% GFP +ve cells, which was maintained for over 100 cell generations in culture. To investigate whether the stability and level of GFP expression was a clonally stable phenomenon, dilution cloning was performed on the polyclonal populations transduced with all 3 vectors, and 12 clonal populations were obtained for each vector. The GFP expression in the clonal populations was monitored for over 50 cell generations and both the percentage of positive cells and the mean fluorescence intensity were found to remain stable within each clone (Figures 5.5 to 5.10). This suggests that the association of the episomes with

the chromosomes or the nuclear matrix is stable, resulting in constant copy numbers as well as expression levels.

The transduction experiment and acquisition of clonal positive populations was only performed once. As 20-50% of the transduced populations remained GFP% following a transient cell cycle arrest, and this included triplicates of cell populations transduced with three different vectors, this was considered supportive enough to discount the possibility of an isolated artefact. Moreover, the dilution cloning experiment showed that the GFP expression was a clonally stable phenomenon. For further supporting evidence, the transduction experiment should nevertheless be repeated to ascertain the reproducibility of the result.

5.3.3 Integration status analysis of vector genomes in clonal IDLV-transduced CHO cell populations

The most important question regarding the IDLV-transduced clonal CHO cell populations stably expressing GFP is whether the transgene has integrated or remains episomal. To answer this question, an array of experiments was conducted investigating the integration status of the vectors. This combined approach was taken to avoid any bias resulting from the weaknesses of each type of experiment when drawing the final conclusions on the integration status in each clone.

Since all the IDLVs used carried a specific incapacitating D64V mutation in the *IN* enzyme that usually facilitates lentiviral integration, it is unlikely that any integration events were *IN*-mediated (Gaur *et al.*, 1998). The mitotic stability of gene expression must therefore either stem from stable episomes or integrants facilitated by cellular mechanisms. Although approximately 1% of cells transduced with IDLVs have been shown to contain integrants (Wanisch *et al.*, 2009; Yanez-Munoz *et al.*, 2006), the rate of spontaneous integration events observed in the cells in induced cell cycle arrest would have had to increase 20 to 30-fold to account for the percentage of stably expressing cells. It was hypothesised that the induced cell cycle arrest either helped the episomes to become associated with the nuclear matrix and therefore mitotically stable, or otherwise significantly increased the activity of the cellular DNA repair mechanisms which promoted integration.

The previously reported levels of residual integration from IDLVs vary from 0.3% - 2% depending on the cell type, and in HeLa cells remained between 1.1% and 1.3% even when a much higher vector dose of MOI 25 was used (Wanisch *et al.*, 2009). Estimates of integration frequency in IDLVs with the D64V mutation such as the one used in this chapter set it approximately 3 logs lower than wild-type vector.

5.3.3.1. LAM-PCR and deep sequencing suggest the majority of IDLV vector genomes are not integrated

Linear amplification –mediated PCR and subsequent deep sequencing of the PCR products is the most reliable method for detecting even the rarest of integration events in LV-transduced cells. Here, it was applied to the clonal populations derived from IDLV –transduced CHO cells subjected to an initial period of cell cycle arrest. The protocol amplifies all sequences immediately adjacent to each LTR, and therefore each integration event will produce amplicons containing host sequences. All amplicons produced were analysed computationally and compared against both the vector and the closest available host (mouse) sequences.

An average of 1200 amplicons was generated by the LAM-PCR protocol for each IDLV-transduced clone. A subsequent BLAST analysis confirmed that the vast majority of the sequences amplified corresponded to the vector. Of the amplified sequences not corresponding to the vector, only one was confirmed by BLAST analysis to be a likely integrant, in IDLV-CG clone 2. Two other clones that produced non-vector amplicons, IDLV-SG clone 10 and IDLV-SGm clone 8, did not retrieve any BLAST results from either the mouse or the rat genomes. Therefore they are likely the result of contamination from neighbouring sequencing lanes containing HIV-1 samples in human cells. The possibility that at least some of the amplicons resulted from integrations in the CHO genome in a rare section that does not correlate well with the mouse and rat genomes cannot be completely excluded, however, it is highly unlikely considering the almost perfect alignment of the other integration sites with one or both the rat and mouse genomes.

In a similar way, the one BLAST-positive read for sample IDLV-CG clone 12 is likely to have arisen through contamination from the neighbouring lane. That lane contained the positive control which produced ~240 amplicons for exactly the same integration site. It is very unlikely for random integration to have occurred at exactly the same site in two independent samples. However, the possibility that this may have happened cannot be entirely discounted, as the site in question could be a hotspot for rearrangement and therefore exhibit a tendency towards non-integrase –mediated integration.

One factor that may prevent integration sites from being detected by LAM-PCR is amplicon length bias. If the restriction site used to form the amplicon unit is further than ~400 bp away from the LTR in the host genome, the resulting fragment will be too long for downstream processing and deep sequencing. It is therefore possible that some integration sites were missed because of this.

Another factor introducing inherent bias in the LAM-PCR experiment is the positioning of the primers in the LTR. Although this is the best method to detect LTR-mediated integrations, in our case it is possible that some of the integration events were not *IN*-mediated and therefore the vector-genome junction is present elsewhere in the genome. These would not have been picked up by the protocol used. In a comprehensive study using LAM-PCR –based methods to detect IDLV integration frequencies in the rat eyecup 2.5 months post-injection, only one integration event was detected despite approximately 30-50% of vector genome having been scanned for vector-genome junctions (Yanez-Munoz *et al.*, 2006). In this chapter, only vector-genome junctions containing LTR sequences will have been detected, and it is therefore possible that vector genome breakage followed by integration has occurred elsewhere and would not have been detected by the LAM-PCR. It would be interesting to perform a more thorough scan of the clonal populations here; however this was not possible within the scope of the current project.

5.3.2.2 Quantitative RT-PCR provides data on intracellular vector conformation

To obtain further information on the copy number of the vector genomes, DNA samples from the stably IDLV-transduced CHO clones were subjected to 4 independent qRT-PCR reactions. Three of these reactions quantified different sections of the vector genome, and the fourth one quantified the number of cells from which the DNA was extracted. The results were used to derive two different types of information: copy number data based on the numbers of vector signal per cell, and integration status data based on the relative amounts of each signal.

The copy number calculations were based on the amount of vector backbone signal detected, as out of the 3 different vector genome sections investigated, this region is the

least likely to be subject to recombination and therefore likely to be detectable in all vector genomes. The average copy number estimates ranged from <1 to 6 per cell, which indicates that probably not all vector genomes were detected in the experiment. However, the overall copy number of 1-15 per cell is in accordance with results obtained by fluorescent in situ hybridisation (see section 5.3.2.4).

Although the qRT-PCR results cannot directly be used to state integration status, the ratios of signals can be used to infer the likely intracellular conformation of the vector genome by comparing the ratios of the 3 separate signals in each sample. Ideally, in 1-LTR circles, the LTR:BB ratio should be 1:1, and in 2-LTR circles the LTR:BB:2LTR ratio should be 2:1:1. A LTR:BB ratio of 2:1 without 2-LTR signal may be indicative of integration (See Table 5.2).

In 7 out of 8 clones investigated, the ratios of the signals supported the episome hypothesis. The signal ratios obtained for IDLV-SG clones 7 and 10 and IDLV-SGm clones 8 and 11 are indicative of 1-LTR episome circles. In all IDLV-CG clones the amount of 2-LTR signal is high, suggesting the presence of 2-LTR circles. In IDLV-SG clone 4, the LTR to BB signal ratio is close to 2:1, which is potentially indicative of integration.

As the qRT-PCR was carried out in 4 separate reactions, it needs to be borne in mind that each reaction may have varied slightly in its efficiency. The 1-LTR to BB ratio may be affected if the reverse transcription during transduction has been incomplete, as the BB signal is measured from the part which is the last one to be transcribed. Similarly, the 1-LTR and 2-LTR reactions may be affected if the primer binding site has been subject to recombination during vector circularisation. Also, as with any very sensitive PCR-based methods, different primers may have slightly differing binding efficiencies, so obtained ratios may deviate somewhat from expected. Ratios may also be affected or signals lost if the method of DNA purification causes the DNA to become fragmented, although it should not be the case here since the protocol used was specifically designed to keep the DNA molecules as intact as possible. Furthermore, any of these imperfections would have affected each clone in the same way, making it unlikely that variation in product amplification efficiency would cause spurious results for any of the samples.

5.3.2.3 Southern blotting produces results consistent with several clones containing episomal vectors

Southern blotting was used to evaluate the conformation of the vector genomes using restriction enzyme -digested genomic DNA extracted from CHO cell clones transduced with IDLVs and subjected to an initial period of cell cycle arrest. At least 3 out of 9 clones showed strong evidence supporting the episome hypothesis.

As can be seen on the gel in Figure 5.15, 2 out of 3 clones transduced with IDLV-SG (clones 4 and 7) show only bands corresponding to the expected sizes for circular episomes. Clone 10 resulted in a band approximately 4.5 kb in size, which could either result from a concatemerisation and rearrangement or integration.

One of the clones transduced with IDLV-SGm, clone 11, also produced a band consistent with the expected size for an episomal vector genome. The remaining 2 clones transduced with IDLV-SGm did not produce unambiguous results; DNA from cells transduced with IDLV-SGm was digested with a different enzyme than for the other 2 vectors (XhoI instead of EcoRI) due to the presence of several EcoRI sites within the miniMAR-element. The resulting digests were likely to be partial, and the DNA fragments did not enter the gel at the correct level for IDLV-SGm clones 5 and 8. It is therefore possible that these 2 samples, digested with a different enzyme, could produce bands corresponding to the expected episome size.

All IDLV-CG –transduced samples show a band which is shorter than the expected 1-LTR episome. It is possible that some recombination events have resulted in a shorter vector, and the presence of the same band for all 3 clones suggests the existence of a recombination hotspot within the vector genome. This may also explain the failure for the FISH probe to work for this vector, since if the episome is reduced in size it may fall below the FISH detection level. A possible explanation is episomal recombination and rearrangement, especially since all 3 independent clones exhibit the same form. In the case of an integrant, it is unlikely for the integration to have occurred in the same place in all 3 clones.

Bands corresponding to the expected sizes of the 1-LTR and 2-LTR episomes can be taken as strong evidence for the presence of the vector DNA in episomal forms. Arguably, integration-defective mutants could exhibit a tendency towards integration

through non-viral pathways, and the genomes might integrate in oligomeric tandem repeats similar to those formed after plasmid transfection (Hagino-Yamagishi *et al.*, 1987). To date, no evidence has been found of a tendency of lentiviral episomes to form integrated concatemeric forms. As the copy number of genomes is shown by qRT-PCR to be well below 10 per cell, integrated head-to-tail concatenated genomes should produce 1-2 additional restriction bands flanking the integration site, and these should be visible on the Southern blot.

5.3.2.4 Fluorescent in situ hybridisation provides supporting data on the copy number and intracellular location of the episomes

Fluorescent in situ hybridisation is a technique which is normally used to visualise DNA sequences measured in tens to hundreds of kilobases. The shorter and rarer the target sequence, the more difficult it becomes to reliably differentiate between target and background. Here, the technique was optimised for one probe, targeting both of the SV40-containing vectors.

A large amount of data was generated by examining images of metaphases naturally appearing on each slide, as well as counting signals and nuclei in several fields across each slide. Statistical analysis of the signals supports the theory that the signals observed arise from real vectors present in the cells, and may be used to infer approximate copy numbers per cell (see Figure 16 A). Copy numbers were estimated for IDLV-SG clone 4 and IDLV-SGm clone 11, which were chosen due to them being the most likely episomal clones for each vector according to the Southern blot experiment. Both clones produced the same average of 1.6 vector genomes per cell. The equivalent copy numbers as calculated from the qRT-PCR results were 2.8 and 6 respectively, the highest for each vector. These data are in relative agreement, as more genomes can be expected to be detected by qRT-PCR than by FISH.

Another probe for investigating the cell lines transduced with IDLV-CG did not produce quantifiable signals. Although both probes were made using the same protocol, it is possible that differing reaction conditions would be required for each. A putative reason for the IDLV-CG probe working include problems with target binding; the CMV sequence may hybridise very weakly and is removed during stringent washes. Another

probable cause is that the probe may bind to genomic sequences through partial homology, resulting in high weak-level background, drowning out the true signals. This was observed on all IDLV-CG FISH slides. While the technique could be optimised and the quality of signals and images potentially improved, due to time constraints this was not possible within the scope of this project.

5.3.2.5 Summary of episomal status analysis

The episomal status of the IDLV –introduced transgenes in CHO cells was analysed using several different techniques, each with their own limitations. Here, a summary of all results from the different experiments is used to analyse the integration status of each clone. The conclusions are also tabulated in Table 5.3.

The linear amplification –mediated PCR reaction was designed to reveal any existing vector-genome junctions. The products of the reaction were first visualised on an agarose gel, as products arising from exclusively episomal vector are expected to correspond to defined sizes. Here, all clones except IDLV-SGm clones 5 and 11 exhibited only bands corresponding to the expected lengths for 1-LTR and 2-LTR episomes (Figure 5.12).

High throughput sequencing of the LAM-PCR products and subsequent BLAST analysis of the sequence reads revealed one likely integrant in IDLV-CG clone 2, corresponding to a locus on the murine chromosome 4 (Table 5.1). Although some of the other samples initially produced reads not corresponding to the vector genome, these were excluded after further BLAST analysis revealed no homology to either the rat or the mouse genome. Two samples, IDLV-SG clone 10 and IDLV-SGm clone 8 did not produce successful sequence reads and could not be analysed.

The quantitative real-time PCR experiment produced data on the amount of LTR signals, LTR to LTR junctions, and vector backbone present in the cells. The ratio of the amount of signal is expected to be different in cells with episomes and integrants, if the integration has been LTR-mediated. The signals were calculated as the amount of each type of product present per cell equivalent, and expressed as ratios of 1-LTR and 2-LTR (LTR to LTR junction) signals for each backbone (BB) signal.

IDLV-SG clones 7 and 10, and IDLV-SGm clone 11 produced LTR to BB signal ratios close to 1:1 (0.7 to 1, 1.1 to 1 and 0.96 to 1 respectively), which is expected to arise from 1-LTR circles. All IDLV-CG clones produced high amounts of LTR signal as well as high amounts of 2-LTR signal, indicating the presence of 2-LTR circles. The ratios of 1-LTR: 2-LTR: BB in the clones were 2.4 to 0.2 to 1 for clone 2, 5.0 to 1.1 to 1 for clone 10, and 5.1 to 1.8 to 1 for clone 12 (see Table 5.2 and Figure 5.14).

The Southern blotting experiment indicated bands corresponding to the expected 1-LTR and 2-LTR episomes for cells transduced with IDLV-SG, clones 4 and 7, and cells transduced with IDLV-SGm, clone 11. Of these, clones 7 and 11 contain a single band, indicating that the cell line is indeed clonal and contains the transgene in a single conformation (Figure 5.15).

In conclusion, at least two of the clonal transduced CHO cell lines appear to retain the transgene in an entirely episomal configuration, IDLV-SG clone 7 and IDLV-SGm clone 11. These two clones showed no evidence for integration in any of the experiments. Furthermore, several other clones exhibit mostly episomal evidence, including IDLV-SG clone 4, in which the only evidence against episomal vector genome is the LTR to backbone ratio in the qRT-PCR.

For IDLV-CG clones 10 and 12, the Southern blotting pattern does not correspond to the expected episome size, but no other experiments found any evidence for integration. This points to either a single integration event resulting from a breakage in the vector genome followed by a non-LTR mediated integration, or rearrangement of the episome.

Vector transduction	IDLV-SG			IDLV-SGm			IDLV-CG		
CHO Cell Clone Identifier #	4	7	10	5	8	11	2	10	12
LAM-PCR amplicon size (agarose gel)	✓	✓	✓	✓	✓	✓	✓	✓	✓
LAM-PCR amplicon Sequence	?	✓	✓	?	✓	✓	✓	?	✓
qRT-PCR for 1-LTR, 2-LTR & backbone sequence	✗	✓	✓	?	✓	✓	✓	✓	✓
Southern blotting analyses	✓	✓	✗	?	?	✓	✓	?	?
FISH analyses	✓	?	?	?	?	✓	?	?	?
Overall majority consensus	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 5.3 Summary of the experimental evidence to evaluate the episomal status of IDLV genomes in CHO cell lines derived by dilution cloning from transduced cell populations subjected to a transient cell cycle arrest.

The table shows the results from 5 independent experimental techniques to evaluate the episomal status of IDLV genomes in clonal CHO cell lines derived by dilution cloning from transduced cell populations subjected to early transient cell cycle arrest. Green tick indicates that the experiment provides evidence to support the existence of replicating episome structures. The red cross indicates that the experiment does not provide evidence to support the existence of replicating episome structures. The ? indicates unavailability of information where the experiment was either not performed or judged not to provide reliable data.

Chapter 6: Translating the IDLV - cell cycle arrest method of episome establishment into clinically relevant applications

6.1. Introduction

6.1.1 Modelling IDLV / cell cycle arrest –based episome systems *in vitro* for potential translation to clinical applications

One of the main aims of current gene therapy is to develop systems for use on *ex vivo* modification of human stem cells to generate safe and stable transgene expression without the oncogenic risks posed by integrating vectors. The major challenge for gene therapy in this context is to retain stable gene expression from non-integrating vector systems. Based on the findings from CHO cells in chapter 5, this chapter assesses the mitotic stability of IDLV-mediated transgene expression *in vitro* in different model situations with relevance for clinical application in humans.

Firstly the phenomenon of episomal stability was evaluated in the human HeLa cell lines. This cell line is highly adherent, which makes it easier to distinguish between and separate live and dead cells; an especially important trait when the cells are undergoing methionine restriction which is likely to affect their morphology, metabolism and viability.

Secondly we chose to attempt episomal transduction of a clinically relevant model, murine haematopoietic stem cells, since the *ex vivo* manipulation and re-transplantation of haematopoietic stem cells is currently being developed for the treatment of several human disorders (see Chapter 1, section 1.6.2).

6.1.2 Origins and biology of HeLa cells

HeLa cells were derived from a biopsy of a cervical adenocarcinoma lesion in 1951 and have since become the most commonly used laboratory cell line due to their durability and proliferation (Scherer *et al.*, 1953). They divide every 24-48 h and are immortal due to an activated form of telomerase. The carcinoma was generated by horizontal gene transfer from the human papillomavirus (HPV), which resulted in a karyotype departing from the expected human spread. The cytogenetic constitution of HeLa cells is considered stable, and contains several numerical and structural chromosomal aberrations (Macville *et al.*, 1999). The main features of the cell line are characteristic of advanced cervical carcinomas. Specifically, the cells are hypertriploid and contain 20 clonally abnormal chromosomes as well as integrated HPV sequences, which can be considered as the cytogenetic signature of this cell line.

In this chapter, the method for inducing a period of cell cycle arrest using methionine and serum depletion was applied to HeLa cells. Arguably, cancer cells are more sensitive to nutrient deprivation than normal human cells due to their elevated metabolic rate. Restriction of the amino acid arginine, for instance, results in the demise of HeLa cells faster than primary fibroblasts (Wheatley *et al.*, 2000). Arginine is a conditionally nonessential amino acid and can be biosynthesised by human cells, but is nevertheless important for growth and proliferation. Methionine, on the other hand, is essential for the initiation of protein synthesis and cannot be synthesised by mammalian cells unless homocysteine is made available. As cancer cells, HeLa cells cannot synthesise methionine from homocysteine, and methionine restriction has been shown to induce caspase-dependent apoptosis in HeLa cells (Lu *et al.*, 2003). In this chapter, the possibility for increasing IDLV retention in HeLa cells by methionine restriction - induced cell cycle arrest will be explored.

6.1.4 Biology & applications of murine haematopoietic stem cells

Haematopoietic stem cells (HSCs) are multipotent stem cells capable of differentiating into all cells of the myeloid and lymphoid lineages, including erythrocytes, platelets, macrophages, T-cells and B-cells. They are capable of self-renewal, meaning that in every cell division at least one of the daughter cells retains is multipotent capacity. HSCs occur in the adult bone marrow at a frequency of 1 in 10^4 cells, but can also be obtained from umbilical cord blood (Schroeder, 2010). They can be maintained and propagated *in vitro* when cytokines are added to the media, enabling a plethora of potential applications, and therefore are at the centre of the development of new stem cell and gene therapies.

A heterogeneous population, murine HSCs are usually practically defined as cells from myeloid tissue that do not exhibit any of the surface markers found in more mature blood cells (lineage negative, Lin^-) but do express CD117 (c-kit) and stem cell antigen 1 (Sca-1). In contrast to human HSCs, repopulating murine HSCs are found in the low to negative CD34 fraction (Spangrude *et al.*, 1988). HSCs are separated from the surrounding cell populating by fluorescent or magnetic labelling and subsequent removal of lineage –committed cells.

Even after the application of all current state-of-the art enrichment and separation strategies, the HSC populations obtained are heterogeneous. The most stringent purification methods yield approximately 10^6 HSCs per mouse, of which 30-50% can be confirmed to be true HSCs after single cell molecular analysis (Schroeder, 2010). The ratio of myeloid and lymphoid cells generated from a HSC population can vary, as can the number of differentiated cells produced from a single HSC. Thus, the heterogeneity manifests both in differentiation bias as well as self-renewal capacity.

It has been shown that individual HSCs preferentially differentiate either into myeloid or lymphoid progeny (Muller-Sieburg *et al.*, 2002). This subtype identity is lost during *in vitro* propagation, suggesting epigenetic factors rather than environmental signalling are dominant in maintaining the bias (Dykstra *et al.*, 2007). However, transplantation of single HSCs into mice showed that at least 25% of stem cells are capable of repopulating the haematopoietic system, showing that the multipotency of HSCs is present *in vivo* and gradually lost during *in vitro* propagation (Dykstra *et al.*, 2007).

In this chapter, the transduction protocol followed by methionine deprivation was applied to mHSCs. Firstly, this was done to test the protocol on primary cells instead of laboratory cell lines, which contain several chromosomal abnormalities and potential latent integrated proviruses, both of which may influence their ability to harbour episomes. Secondly, if successful, a system whereupon a non-integrating lentivector could be used to transduce quiescent cells to establish mitotically stable episomes would find its most important clinical use in the *ex vivo* manipulation of stem cells like HSCs. Therefore it was important to explore the applicability of the stable episome generation to primary, quiescent stem cells.

There is a possibility that the special needs of stem cell culture may interfere with the application of the methionine depletion. The cultivation of mHSCs was done according to an established protocol from Institute of Child Health, UCL (personal communication with Steve Howe). This consisted of co-application of 3 cytokines, mSCF, IL-6 and Flt3-ligand, which support cell viability, maintenance and proliferation.

Recombinant Mouse Stem Cell Factor /c-kit ligand (mSCF) is a transmembrane glycoprotein which can be proteolytically cleaved in the extracellular region to produce a soluble form. It plays an important role in the regulation of HSCs, which display the same level of receptor for SCF at all stages of development (c-Kit receptor or CD117). *In vivo*, it is constitutively produced by endothelial cells and fibroblasts. Null mutant mice for either the ligand or the receptor perish *in utero*, suggesting that SCF plays an essential role during development. The binding of SCF to CD117 causes the receptor to homodimerise and initiates multiple signalling cascades. SCF has been shown to increase the survival of haematopoietic stem cells *in vitro* (Broudy, 1997).

Recombinant Human IL-6 is an interleukin and a pleiotropic cytokine that regulates haematopoiesis and immune and inflammatory responses. In combination with SCF, it promotes haematopoietic progenitor cell proliferation and differentiation in vitro, and has also been shown to promote retroviral-mediated gene transfer into HSCs (Luskey *et al.*, 1992).

Human Fms-related tyrosine kinase 3 ligand (Flt3-Ligand**)** is a haematopoietic cytokine which binds to the CD135 receptor expressed on the surface of HSCs. It synergises with other SCFs and interleukins to stimulate proliferation of early HSCs.

There is no data available on the response of HSCs to methionine depletion, but it can be expected to slow down the metabolism due to the blockage to protein synthesis. Hypothetically, the true HSCs are quiescent upon extraction from the bone marrow, and are stimulated to proliferate by the cytokines added to the media; therefore the effects of the cytokines and methionine depletion are potentially antagonistic. To investigate the interactions, it was decided to experiment with several different combinations of methionine depletion and cytokine supplements in an attempt to find a combination that would result in reversible cell cycle arrest without apoptosis.

6.2 Results

6.2.1 Optimisation of induced cell cycle arrest in HeLa cells for IDLV transduction

A proliferation assay was employed to compare the response of CHO cells and HeLa cells to methionine and serum depletion. The experiment is designed to provide information on the number of metabolically active cells in a given population. As the rate of cell proliferation is also dependent on cell density, 2 different densities were assessed. The cells were seeded at densities corresponding to either $1x10^5$ or $2x10^5$ cells per well on a 6-well plate, and cultured either in normal media or in methionine-free serum-depleted media for 5 days. The metabolic activity of the cells was then evaluated using the CellTiter assay (see Materials & Methods, section 2.7.5).

After 5 days in methionine-free media, both cell lines showed markedly reduced metabolism, indicating either cell cycle arrest or cell death (Figure 6.1). Although both cell lines show a similar response to methionine depletion, the difference in the metabolic activity between CHO cells and HeLa cells after exposure to the cell cycle arrest-inducing media is statistically significant. Student's t-test indicates a more significant difference between the higher density populations (P <0.001) than the lower density populations (P <0.01). The least significant difference is found between the low density CHO population and high density HeLa population (P =0.028) (Figure 6.1).

In Chapter 5, the culture conditions that resulted in the stable retention of transduced IDLVs in CHO cells included seeding at a density of $1x10^5$ cells per well followed by 5 days in methionine-depleted media. HeLa cells exhibited a similar number of proliferating cells when seeded at $2x10^5$ cells per well, supported by the less statistically significant difference when compared with the lower cell density. To maintain the culture conditions close to the original CHO cell protocol, cell density of $2x10^5$ HeLa cells per well was therefore chosen for further experiments.

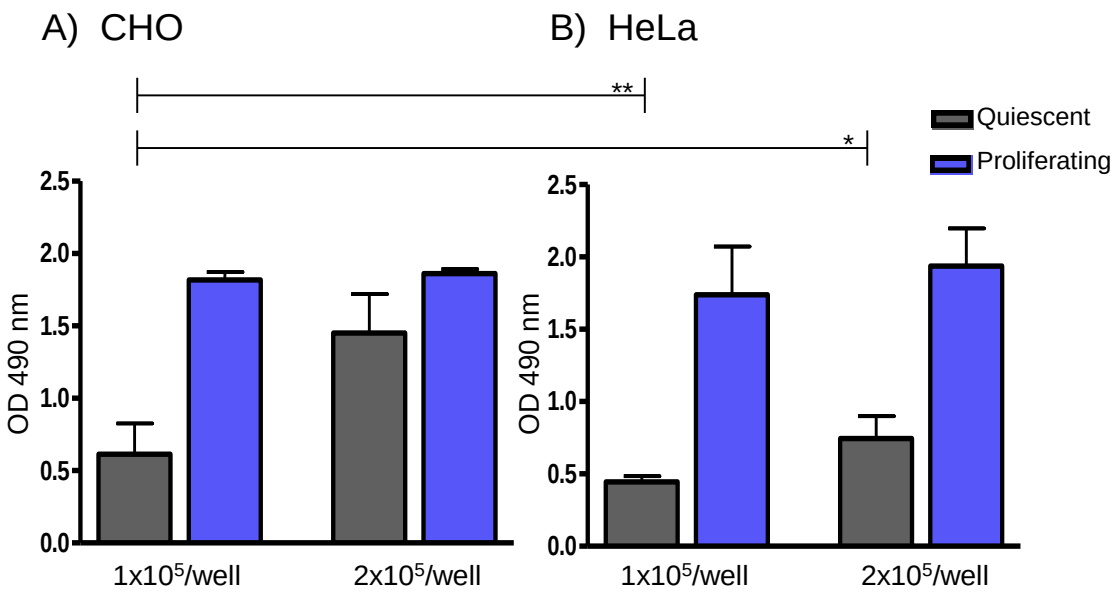


Figure 6.1 Proliferation assay indicates a significant difference between CHO and HeLa cells when subjected to methionine and serum depletion.

CHO and HeLa cells were seeded at 2 densities, corresponding to $1x10^5$ and $2x10^5$ cells/well on a 6-well plate. Half of the cells were subjected to methionine and serum depletion for 5 days to induce cell cycle arrest (grey bars). Normally proliferating controls were seeded at the same time (blue bars).

After 5 days, a proliferation assay was performed to assess the metabolic activity of the cells. Y axis shows optical density at 490 nm after adding MTS tetrazolium reagent, which is directly proportional to the number of metabolising cells. A significant difference was found between the low density CHO cells and HeLa cells for both cell densities when subjected to methionine depletion. The difference is less significant between low density CHOs and high density HeLas (= P <0.05, ** = P <0.01 by Student's t-test). Data shown are the mean ± standard deviation, n = 16.*

The effect of methionine and serum depletion on HeLa cells was also investigated by observing the morphology of the cells during induced cell cycle arrest. During 5 days in depleted media, the cells developed an increasingly elongated morphology, eventually resembling neurons (Figure 6.2 A). The distended forms begin to appear 3-4 days into the induced cell cycle arrest, and by day 5 many of the cells begin to detach from the plate. This effect is not observed in CHO cells, which retain consistent morphology throughout the induced cell cycle arrest.

The proportion of cells in each cell cycle phase was also monitored during culture in methionine-free media by quantifying the amount of DNA present in each nucleus by propidium iodide staining. For the experiment, cells were seeded at a density of $2x10^5$ cells per well and the media changed 24 h after seeding. A control population was kept proliferating in standard media. 5 days later, the proportion of cells in each cell cycle phase was determined by flow cytometric analysis of the PI stained nuclei. No significant difference was found between the populations cultured with and without methionine (Figure 6.2 B). However, there was a difference between treatments in the percentage of the total cell population classed as G_1, S or G_2 by the FlowJo software; of the proliferating cells, on average 75% could be classified, whereas only 53% of the cells in methionine-free media could be classified. This is a potential reflection of cell viability.

Considering the apparent morphological changes taking place at 4-5 days in methionine-free media (Figure 6.2 A), combined with the lack of significant change in the cell cycle phase proportions even after 5 days, it was decided that 4 days of methionine depletion should be used in further experiments to maximise the number of surviving cells.

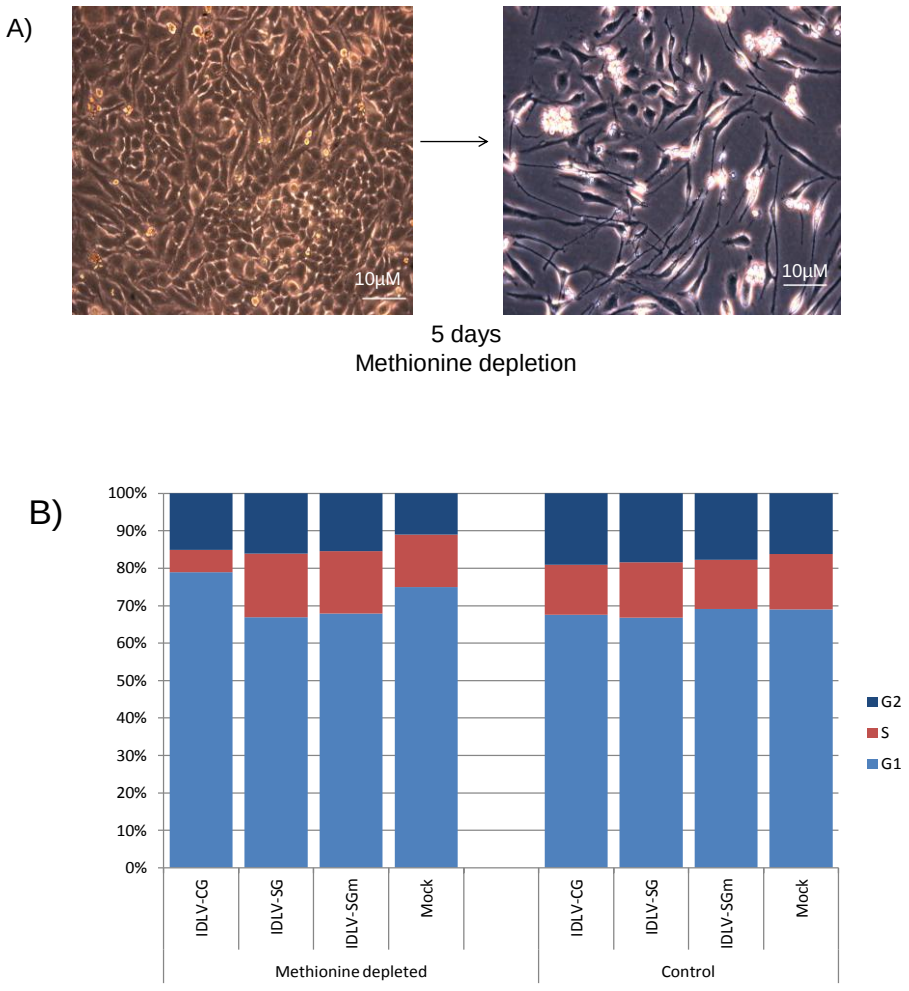


Figure 6.2 The morphology but not the cell cycle phase distribution of HeLa cells responds to methionine and serum depletion.

HeLa cells exhibit an elongated morphology after 5 days in methionine-free, serum-depleted media (A). However, propidium iodide staining and flow cytometric analysis of the nuclei of HeLa cells subjected to 5 days culture in methionine-free, serum-depleted media shows no significant differences in the proportions of cells in each cell cycle phase and the control group in normally proliferating culture (B). Statistical analyses were conducted using Student's t-test comparing each cell cycle phase separately between arrested and control at D6. The total cell population included in the analysis consists of cells successfully categorised by the FlowJo software following PI staining. This proportion varied from 51 % to 76 %, most likely due to dead cells present in the sample. N=4.

6.2.2 Optimisation of *in vitro* transduction of HeLa cells

Transduction of HeLa cells by IDLVs was optimised by seeding 40 000 and 80 000 cells / well on a 12-well plate (corresponding to $1x10^5$ and $2x10^5$cells /well on a 6-well plate by factoring cell number for surface area of the well). 24 h after seeding the cells were transduced with IDLV-SG. The MOIs used were 1,2,3,4 & 5, as calculated from the 2nd dilution of FACS titration, with polybrene added to a final concentration of 8 µg/ml. The cells were harvested 72 h post-transduction and GFP expression was analysed by flow cytometry. All MOIs resulted in a high percentage of the cells becoming transduced, although the transduction efficiencies were higher at the lower cell density (Figure 6.3 A).

To investigate the effect of the promoter on transduction efficiency, HeLa cells were further seeded at a density corresponding to $2x10^5$cells /well on a 6-well plate and transduced with IDLV-SG and IDLV-CG an MOI of 1. The cells were analysed for GFP expression 72 h after transduction, and the CMV promoter was found to express at a slightly higher level (97 %) than the SV40 promoter (91 %) in this cell type. Based on the results, MOI 1 was chosen for the transduction experiment as this resulted in 65-97 % GFP positive cells for vectors with both promoters.

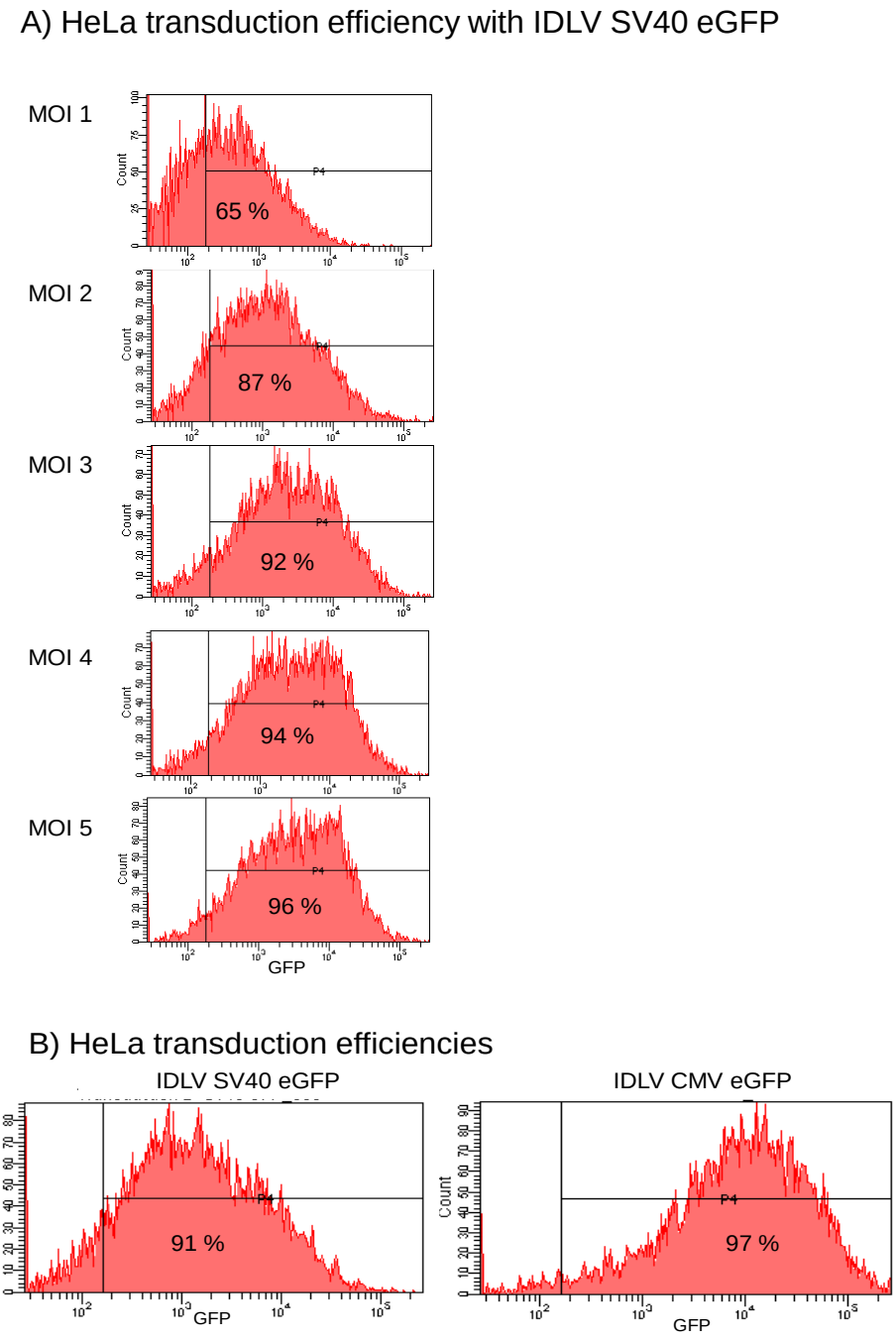


Figure 6.3 Flow cytometric analysis of GFP expression in HeLa cells transduced with IDLVs shows that HeLas can efficiently be transduced with an MOI of 1.

HeLa cells were transduced with IDLVs using polybrene, and GFP expression was measured after 3 days using flow cytometry. The histograms for HeLa cells show transduction efficiencies varying from 65% to 96% at MOIs 1 to 5 (A).

Efficient transduction using an MOI of 1 was confirmed using both IDLV-SG and IDLV-CG (B.) The vector containing the CMV promoter resulted in a higher MFI (21 980) than the SV40 promoter (MFI 4780). The mock was set at 0.5 %. N = 10 000 cells /vector.

6.2.3 Methionine depletion does not result in increased retention of lentivector genomes in HeLa cells

HeLa cells were transduced using seeding density of $2x10^5$ and MOI of 1. The initial transduction efficiencies as measured at 24 h post-transduction were 83.7 % for IDLV-SG, 76.8 % for IDLV-SGm and 84 % for IDLV-CG. Cell cycle arrest was induced 24 h after transduction for 4 days. The proportion of GFP -expressing cells was slightly higher for most populations at the end of cell cycle arrest, measured as 89.4 % for IDLV -SG, 84.9 % for IDLV-SGm and 90.0 % for IDLV-CG.

By one week after release into proliferating culture, the GFP expression had declined to 19.8 % for IDLV-SG, 22.0 % for IDLV-SGm and 53.0 % for IDLV-CG. The proportion of GFP expressing cells continued to decline further until, by 3 weeks post-transduction, transgene expression had dropped to levels 0 – 1.1% above background (Figure 6.6, A and C). The cells in continuously proliferating culture exhibited a similar decline in GFP expression within the same time frame. By 3 weeks post-transduction, the percentages of expressing cells had dropped to 0 -1.4% above background (Figure 6.4, B and D).

The differences in stable GFP expression levels between vectors and treatments were very small. Although the vector containing the CMV promoter exhibited slightly more long-term expression and the final expression level was fractionally higher in the cell set not held in cell cycle arrest, the differences are not statistically significant (Figure 6.4, C and D).

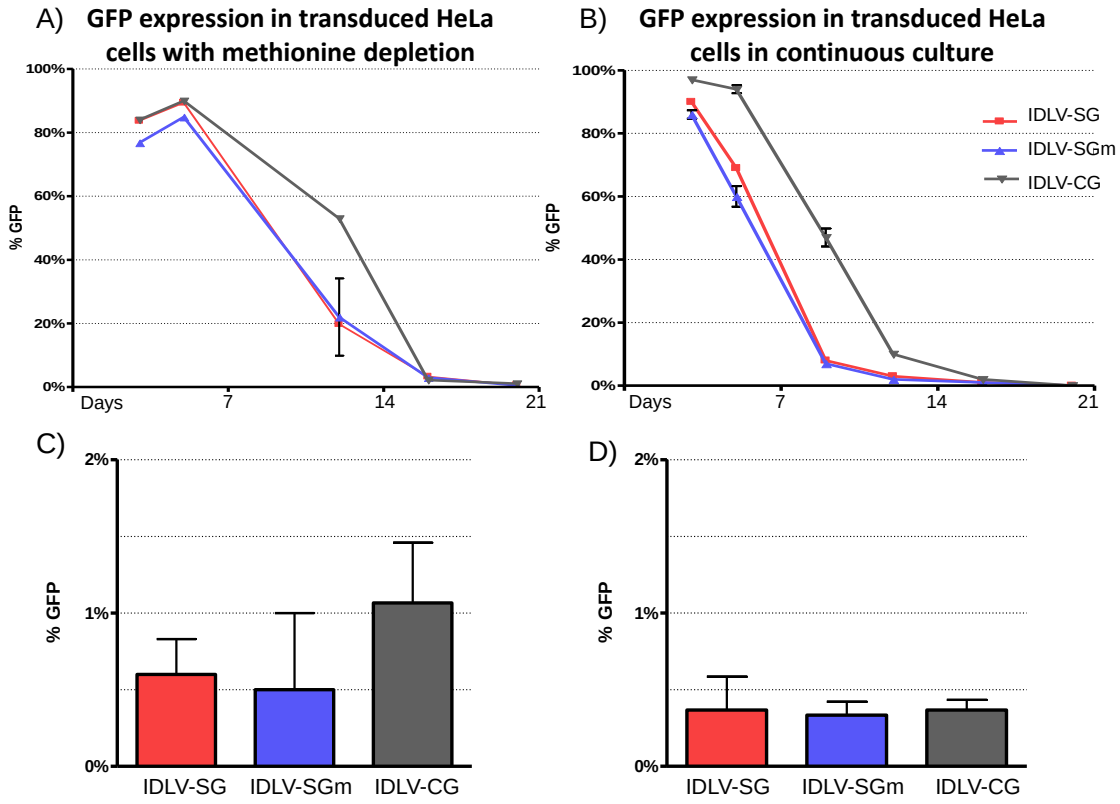


Figure 6.4 GFP expression is not retained in HeLa cells after IDLV transduction followed by methionine and serum depletion.

HeLa cells were plated into culture, and after 1 day transduced with IDLV-DG, IDLV-SGm and IDLV-CG vectors (MOI=1). Cells were placed in methionine and serum depletion of the culture media at D2 (A and C) or allowed to proliferate freely (B and D). From D7, all cells were allowed to proliferate normally with routine passaging for a period up to D21. In (C) and (D) the stabilised transduction levels at the D21 time point are shown. Data are means ± s.e.mean, n=3, SD not shown for first 2 time points in A) as only one data point was available. The difference between vectors and treatments is not statistically significant (ANOVA and t-test). IDLV-SG = IDLV SV40 GFP; IDLV-SGm = IDLV SV40 GFP mMAR; IDLV-CG = IDLV CMV GFP. The mock transduced population was set to 0.5% GFP.

6.2.4 Optimisation of induced cell cycle arrest in murine HSCs

In order to evaluate the effect of methionine depletion on viability of mHSCs, the cells were seeded at $8x10^4$ per well in 24-well culture plates in complete HSC medium to promote survival. After 2 days, the culture was continued either in the presence or the absence of methionine for 20 days, and the live and dead proportions of cells were counted at regular intervals. The cells in the methionine –containing media were split 1:4 whenever the cell count exceeded $1x10^5$ cell per well, but the cells in methionine-free media were not split as the cell count did not exceed $1x10^5$ cell per well throughout the experiment.

After 20 days, > $3.5x10^4$ cells per well in both groups were still viable, showing that both methionine-free and methionine-containing DME/F-12 media with added cytokines were able to support mHSC viability and proliferation *in vitro*. The methionine –free media resulted in a reduction in cell number, with less than 50 % of the total cell number observed in the methionine-free media versus cells in methionine-containing media after 5 days (Figure 6.5), with $2.2x10^5$ cells in methionine-containing media and $1x10^5$ cells in methionine-free media.

For future experiments, the modification of always keeping ≥ $1x10^5$ cells per well was added to the protocol, as the number of cells per well was observed to have an effect on viability and the cells in methionine –containing media exhibited more cell death after being seeded at a density less than $1x10^5$.

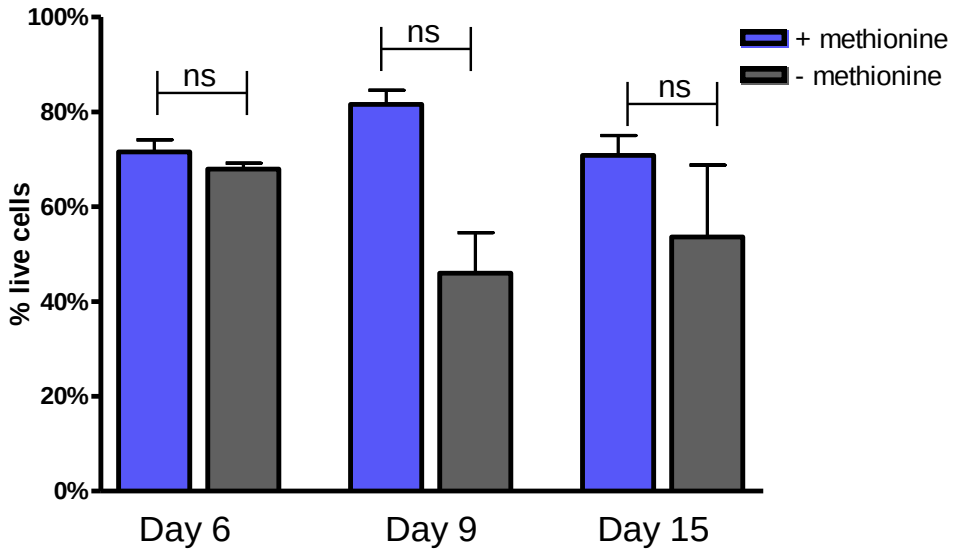


Figure 6.5 Murine HSCs can survive in vitro in the absence of methionine.

Primary mHSCs were seeded at 8x10^4 cells per well in methionine-containing media at Day 0. At Day 2, media was changed to methionine-free 2%FCS in half of the wells. Proliferation and viability was monitored by counting live and dead cells for 20 d using Trypan Blue. The cells in methionine-containing media were split 1:4 every 3-4 days, the cells in methionine-free media were not split. There is no significant difference between the populations at any of the time points (Student's t-test). The results show that mHSCs can be kept alive in both medias for at least 15 days although a decline in both proliferation and viability is observed. Methionine depletion partially decreased cell numbers. N=3.

6.2.5 IDLV Transduction and induced cell cycle arrest result in delayed GFP expression kinetics in mHSCs

The culture media requirements of haematopoietic stem cells are more complicated than that of standard cell lines, as is the prolonged culture of these cells *ex vivo*. We decided to approach the induction of cell cycle arrest simultaneously from several different angles and tested a total of 11 different combinations of media and cytokines (Table 6.1) for their ability to induce or retain the quiescent state of the cells, whilst allowing for transduction and viability of the cells.

6.2.5.1 Induction of cell cycle arrest in mHSCs requires both methionine and cytokine restriction

Murine HSCs were extracted from the bone marrow of two animals as previously, obtaining $1.1x10^6$ cells from a total of $9.83x10^7$ BM cells. The cells were seeded at $1x10^5$cells/well and transduced with IDLV-CG immediately after lineage depletion. They were seeded in 300 µl of the first media, and 1 day post-transduction 200 µl of another media was added onto the well (see Table 6.1). Four days after seeding, complete mHSC media was added to all wells to re-stimulate growth and survival (Figure 6.6).

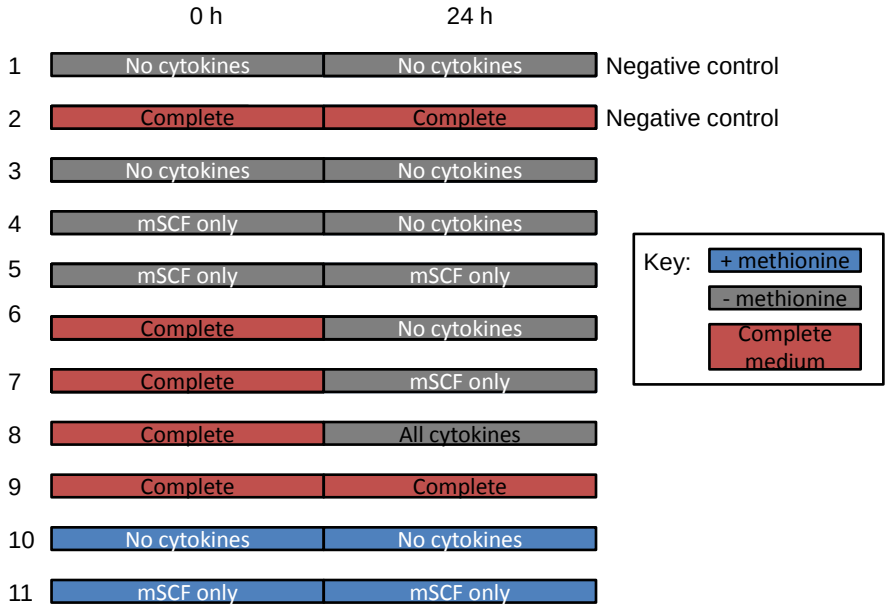


Table 6.1 Diagrammatic representation of the different media combinations used in the experiment. *Numbers on left correspond to each condition. Grey = -Met 2% FCS, blue = +Met 20% FCS, red = Complete mHSC medium. mSCF, murine stem cell factor.*

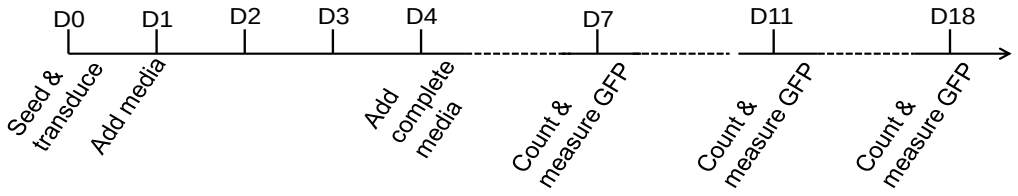


Figure 6.6 Diagrammatic representation of the timeline for the HSC transduction experiment.

HSCs were obtained from murine bone marrow by lineage depletion. They were transduced with IDLV-CG (MOI 40) immediately upon seeding at 1x10^{5}cells/well on a 24-well plate in 300 µl initial medium. 24 h later, 200 µl of second medium was added to each well. The cells were allowed to remain undisturbed until day 4 after seeding, at which point they were counted and GFP expression was measured. The cells were re-seeded in complete media at 1x10^{5}cells/well. Cells were counted again, GFP expression measured by flow cytometry, and re-seeded in complete media at 1x10^5 cells/well at 7, 11 and 18 days post-transduction.

Conditions 4, 5 and 8 resulted in the most convincing and reversible non-proliferative phases, and recovered fully after introduction into complete media. Well 5 contained no methionine and only mSCF for the first 90 h, and well 4 was seeded with the same methionine-free media containing only mSCF but the media added 24 h later contained no methionine or cytokines. Cells in well 8 were seeded in full media followed by methionine-free media with all 3 cytokines 24 h later. Thus it appears that methionine depletion was successful in slowing the cycling of the cells without affecting the viability.

Conditions 1, 3, and 10 were seeded in media containing no cytokines, and a vast majority of the cells became apoptotic or necrotic by 4 days after seeding. None of the populations could be recovered even after the addition of full media with cytokines 90 h after seeding, suggesting that at least mSCF is essential to cell viability. Supporting this hypothesis, cells seeded in media containing only mSCF but not the other cytokines survived (samples 5 and 11).

However, condition 5, which initially contained no methionine, recovered much better than condition 11 seeded with methionine and mSCF. Whereas cells in well 5 had

approximately doubled in number by 4 days after seeding, the number of viable cells in well 11 had decreased in number by the same time point and only returned back to the original cell number seeded after 11 days.

As expected, the 4 media combinations (2, 6, 7 and 9) in which the cells were seeded in methionine –containing media with all 3 cytokines proliferated most in the first 7 days. Conditions 2 and 9 were essentially replicates in terms of media added, the only difference being that condition 9 was additionally transduced with IDLV and condition 2 was not. As observed for the other transduced/ non-transduced pair of samples (1 and 3), there is little variation in the rate of proliferation.

Interestingly, the highest proliferation in the first week after seeding was achieved in well 7 where the cells were seeded in full media followed by methionine –free media with mSCF only at 24 h. In well 6, full media was followed by methionine –free media but with no cytokines, resulting in slower proliferation for up to one week after plating.

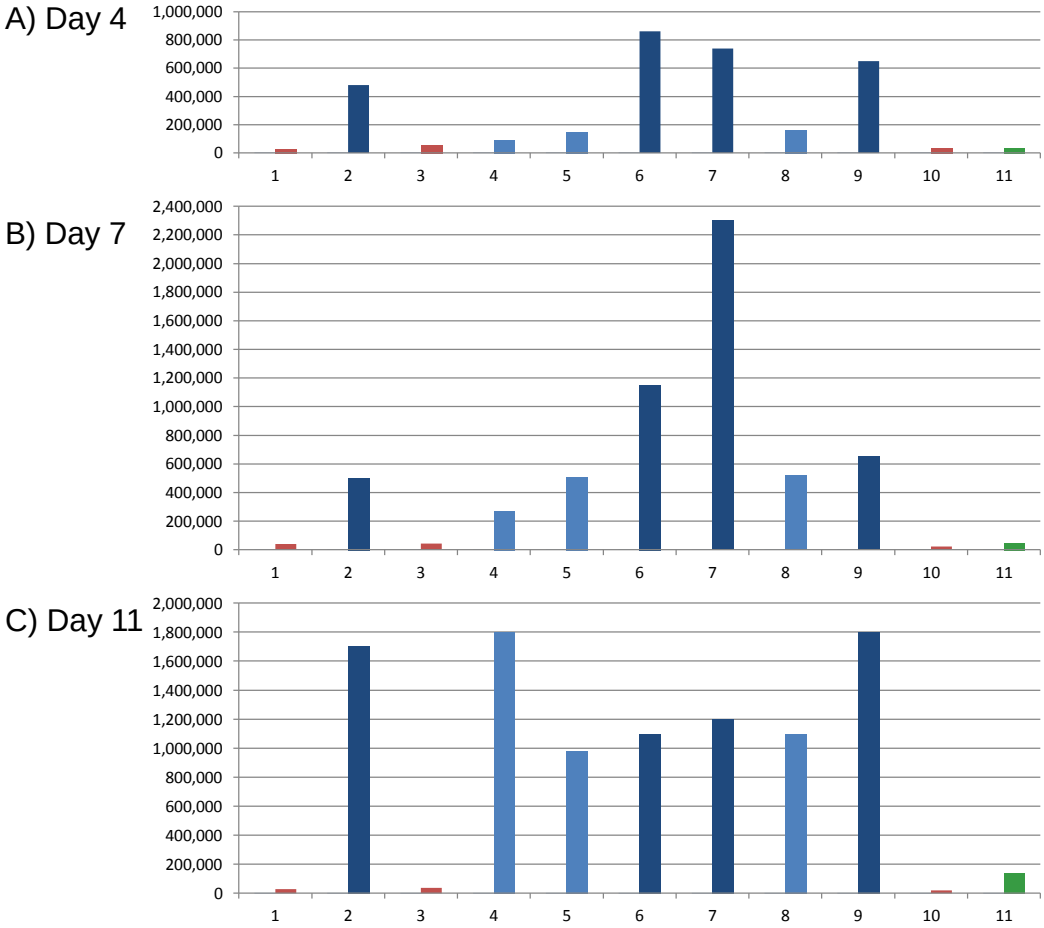


Figure 6.6 mHSCs respond to different combinations of methionine and cytokines in culture media.

Murine HSCs were obtained by lineage depletion and seeded at a density of 1x10^{5}cells/well using 11 different media conditions to test their capacity for retaining HSCs in a state of reversible cell cycle arrest. Cells were counted at 4, 7 and 11 days after seeding. The bar height indicates the number of live cells present at each time point. Three of the conditions resulted in a reversible pause in proliferation, namely conditions 4, 5 and 8 (pale blue). One condition experienced high cell death but recovered within 11 days (11, green). Three of the conditions (1, 3 & 10) resulted in excessive cells death (red), and four did not reduce cell proliferation (2, 6, 7 & 9, dark blue). N=1.

6.2.5.2 GFP expression is not retained in transduced mHSCs following induced cell cycle arrest

The HSCs were transduced with IDLV CMV GFP immediately after seeding, with an MOI of 40. GFP expression was measured at 3-4 day intervals, but only if there were more than $1x10^5$ cells per well as mHSCs suffer at densities of less than $\sim10^5$ cells /well. If the viable cell number was less than the threshold, the cells were re-plated in fresh media without removing the 10^4 cells needed for flow cytometric analysis in order to conserve the population.

The transduction efficiency was first measured at 4 days post-transduction. Only 7 out of 11 samples could be analysed at this stage, and the percentages of GFP expressing cells at this stage varied from 53 % to 91 %.

The percentage of GFP expressing cells was higher in the wells which had been successfully held back from proliferation. Samples 2 and 7 produced the highest GFP levels at 86 % and 91 %, respectively. Although sample 3 could not be measured at this stage, the initial transduction efficiency can be extrapolated to have been high, as the proportion of GFP –positive cells in this sample was the highest at the next measurement at day 7, at 3.9 %. In contrast, samples 2 and 7 only expressed at 2.1 % and 1.7 % by day 7 (Figure 6.10 A, light blue).

The samples which proliferated at higher rates lost their GFP expression more quickly. The fast-growing samples 4, 5 and 6 were only 52 - 61 % GFP positive by day 4, with the most prolific sample 4 the least positive at 52 %, and 5 and 6 having similar growth rates and GFP expression levels at 60 % for sample 5 and 61 % for sample 6 (Figure 6.10 A, dark blue). By day 7, the GFP expression had dropped to below 2% in all 3 highly proliferating cell lines.

A trend towards higher expression in the samples which underwent a successful quiescent period can be seen in the GFP percentages at 11 days post-transduction (Figure 6.10 B). Especially samples 3 and 11, which were the slowest proliferating cell populations to survive, show a tendency for higher level of GFP expression. However, during the following week GFP expression declined in all populations, and by 18 days post-transduction had dropped to background levels in all populations.

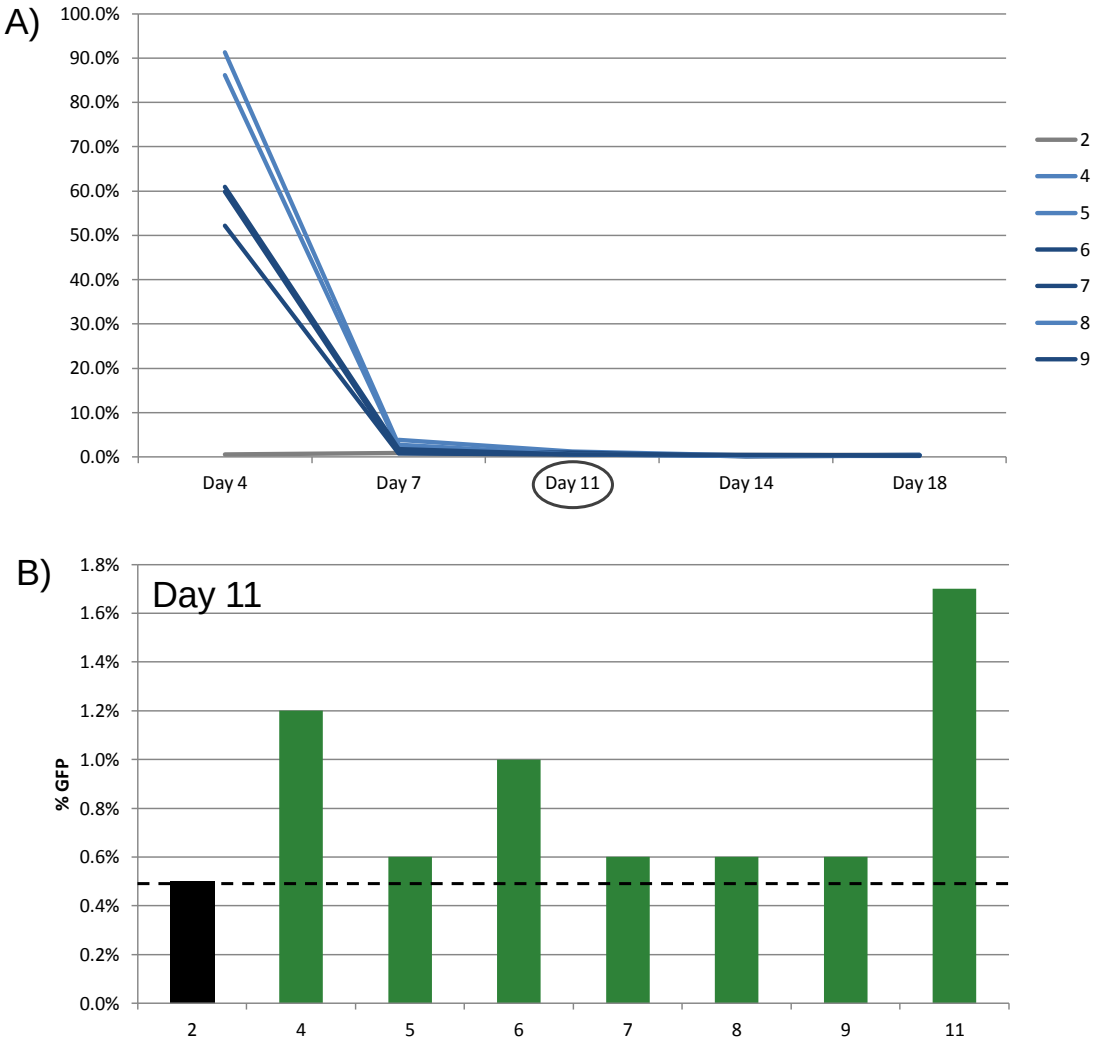


Figure 6.7 GFP expression is not retained in IDLV –transduced murine haematopoietic stem cells.

Murine HSCs were transduced with IDLV-CMV-GFP immediately upon seeding. After 4 days, non-dividing samples 4, 5 and 8 (light blue) expressed GFP at a higher level than proliferating samples 6, 7 and 9 (dark blue). Sample 2 (grey) is a non-transduced control. GFP expression declined as cells continued to proliferate, and reached background levels by 18 days post-transduction. Samples for which data could not be collected at all data points are omitted, n = 10 000 cells /measurement /sample.

B) The bars represent the percentage of GFP+ cells in IDLV-transduced HSCs at 11 days post-transduction. The samples which have experienced the least proliferation (4 and 11) exhibit the highest levels of GFP expression. The non-transduced control population (sample 2) was set to 0.5% GFP, n = 10 000 cells /sample. Cell populations containing less than 10^5 cells at this time point were not included in the analysis.

6.3 Discussion

The experiments in this chapter were designed to begin to investigate the potential of IDLV transduction to lead to the formation and establishment of stable replicating episomes in different cell types. In addition, the effect of S/MAR elements, different promoters, and cell culture manipulations to induce an initial period of cell cycle arrest were examined. The advantages of using IDLVs in such protocols are two-fold: firstly, lentivectors are able to efficiently transduce quiescent cells such as stem cells, and secondly, non-integrating vectors are considered safer compared to integrating vectors, and are therefore more likely to find wide application.

6.3.1 Methionine and serum depletion does not induce cell cycle arrest or retention of IDLV-based GFP expression in HeLa cells

Initially we wished to determine if the phenomenon of establishment of replicating episomes seen in CHO cells following IDLV transduction and induced cell cycle arrest could be transposed to a human cell culture. The human cell line chosen for the experiment was the rapidly proliferating immortal HeLa cell line. HeLa cells have a highly unusual chromosome number of 82 instead of the usual 46. The highly transformed nature and extreme cytogenetic traits of HeLa cells may have affected the ability of these cells to undergo and recover from cell cycle arrest induced by methionine depletion.

A proliferation assay was done to ascertain the appropriate seeding cell density in order to achieve a similar density of metabolising cells as exhibited by CHO cells after the same amount of time in methionine-free media. The assay uses a colorimetric method to measure the metabolic activity in the cells (Cory *et al.*, 1991). As the quantity of formazan produced is directly proportional to the number of living, metabolising cells in the culture, this information was used to compare the numbers of actively metabolising cells in each population. The HeLa cell population producing a similar amount of formazan after 5 days in cell cycle arrest, as compared with the CHO cell population used in chapter 5, was chosen for the subsequent experiments in this chapter.

The morphology of HeLa cells was also observed during 5 days culture in methionine-free, serum-depleted media. After 3 days, the cells were seen to increasingly adopt an elongated form. This may in part be due to the lack of attachment factors usually provided by the foetal calf serum. Whereas there was no visible difference in the cells after 4 or 5 days, the number of cells attached to the culture plate continued to decline with time and it was therefore concluded that 4 days in methionine-free media was the maximum time possible before too much cell death occurred.

Cell cycle phase analysis by PI staining of the nuclei following 4 days in cell cycle arrest showed the majority of the live HeLa cells to be in G_0/G_1. Despite the major difference in morphology, there was no significant difference in the cell cycle phase distribution between the populations in methionine-free serum depleted media and in normal media. However, a difference was observed in the proportion of cells unclassified by the cell cycle analysis software as one of the 3 phases (G_1, S or G_2); the percentage of unclassified cells increased from 53% to 75% on average, reflecting cell death.

Despite the fact that we failed to demonstrate induction of cell cycle arrest in HeLa cells following methionine depletion, experiments were conducted to examine IDLV transduction efficiency and stability. However, in both methionine-depleted and non-depleted cultures the percentage of GFP+ cells, which initially was high (>60%), decreased during proliferation to background levels with no evidence of significant stable transgene retention.

6.3.2 IDLV Transduction and cell cycle arrest studies in mHSCs

The establishment of stable episomes following a period of cell cycle arrest would find its most important clinical use in the *ex vivo* transduction of stem cells, which by definition already exist in a quiescent state. As lentiviral vectors are able to transduce quiescent cells, we decided to explore the possibility for transducing primary murine haematopoietic stem cells with IDLVs, followed by maintaining the cells in a viable yet quiescent state for a number of days, and the subsequent *in vitro* culture of the transduced cells to assess episome stability.

HSCs require a cocktail of cytokines as well as comprehensive media for *in vitro* culture and propagation. For this reason, a specialised media was designed containing all amino acids required by the cells except for L-methionine, which could be added or omitted as needed. Since the synergistic effects between the cells, cytokines and other components of the media have not yet been elucidated, different combinations of cytokines, methionine and FCS concentration were tested for their ability to facilitate both transduction and allow survival but avoid apoptosis.

The cytokines IL-6 and Flt3 have been indicated as promoting stem cell proliferation in culture, and so these cytokines were omitted from some of the media combinations. Murine stem cell factor (mSCF) was still added as it is considered to promote survival rather than cell division (Broudy, 1997). Interestingly, the samples seeded in methionine-free media containing mSCF (4 and 5) survived better than the sample seeded in methionine-containing media with mSCF (11), potentially indicating synergistic effect between methionine depletion and maintenance of viable stem cell cycle arrest.

For the murine HSCs, cell proliferation studies were only conducted by cell counting. Cell cycle phase analysis by propidium iodide was not possible due to the large number of cells ($>10^6$) required for this type of analysis. However, cell cycle phase analysis can also be done using DAPI staining of the nuclei followed by flow cytometric analysis. For future investigations, access to the appropriate facilities could be arranged to investigate the DNA content in cells in the more promising culture conditions to ascertain the cell cycle phase distribution.

Transduction efficiency was not dependent on the proliferation status of the cells. Previous reports indicate that lentivectors are able to transduce quiescent human HSCs, and this finding was replicated here in the samples showing minimal proliferation and high transduction (Figure 6.10, samples 2,3 and 7) (Case *et al.*, 1999).
Three conditions resulted in a large proportion of the cells being transduced, followed by successful reduction in cell number indicating potentially cell cycle arrest, and successful cell expansion once in full growth media. The GFP expression was nevertheless lost in all samples by 18 days after start of the experiment.

In this case, the loss of GFP expression may be due to loss of the episome during mitosis; however, it may also be a result of CMV promoter shut-down, which has been shown to occur in several cell types including human CD34+ T-lymphocytes (Sirven *et al.*, 2001). Circumventing promoter silencing by using the SV40 promoter –driven vector remains a possibility to be explored. Silencing of the pEPI episomal vector has been reported in the murine erythroleukemic cell line (Papapetrou *et al.*, 2006). It is therefore conceivable that silencing of the lentiviral episomes would also occur in this closely related cell line.

An interesting follow-up experiment would be to repeat the transduction and the 3 conditions that resulted in a successful cell cycle arrest, using IDLV-SV40-GFP. Unlike the CMV promoter, the SV40 promoter is not likely to be shut down in the course of the experiment, and may provide an insight into the mechanism of the loss of transgene expression.

Chapter 7: General discussion, conclusions & future directions

The objective of the investigations in this thesis was to explore methods for establishing mitotically stable episomal vectors. The paramount aim was to contribute to the development of clinically applicable, safe and effective non-integrating gene delivery vectors.

In this final chapter, the overall value of SIDD analysis in the construction of gene transfer vectors will be discussed, together with the influence of S/MAR elements on transgene retention. Also evaluated is the effectiveness of methionine depletion as a method for inducing a reversible period of cell cycle arrest in different cell types. The main results and conclusions of each chapter will be considered in turn along with the issues that remain unresolved within each area of research. The ways in which each investigation could be pursued further will also be evaluated. Finally, the importance of the work in this thesis in the context of the general field of gene therapy will be considered.

7.1 SIDD analysis in vector design and the effect of S/MAR elements on transgene retention

Computational analysis of stress-induced duplex destabilisation along the length of each transfer vector was conducted in each chapter prior to embarking on functional *in vitro* testing of the vectors. The need for SIDD analysis originally stemmed from the inclusion of S/MAR elements in the vectors, as S/MARs are known to be highly destabilised regions with the potential for affecting other regions of the molecule that require strand separation in order to function efficiently, such as promoters.

The rationale for including the analysis here was bi-fold. Firstly, it was done to preclude avoidable mistakes in vector design prior to manufacture and functional testing. Two factors need to be considered when constructing a vector containing a S/MAR element; the need for transcription to run through the S/MAR, and potential condensation of the

double helix in neighbouring regions. At least in the case of larger transfer vectors, SIDD analysis allows for the optimisation of these two occasionally conflicting requirements.

Secondly, SIDD analysis can be useful when evaluating functional data. In a S/MAR - related publication, SIDD analysis was used to indicate the best possible configuration for the transfer vector (Giannakopoulos *et al.*, 2009). The SIDD analysis was used in conjunction with functional data to explain the behaviour of the construct and to make an improved version. Similarly, in this thesis SIDD analysis was most useful when used in conjunction with functional data for analysis of vector performance, especially in Chapter 4 for the AAV constructs.

The predicted strand separation potential exhibited major changes in the self-complementary AAV vectors upon the introduction of the miniMAR element, and based on the SIDD algorithm this was predicted to decrease gene expression significantly due to promoter condensation (Figure 4.2). This was indeed found to be the case, and therefore it can be suggested that the S/MAR elements investigated here are not suitable for use in scAAV vectors.

Neither the full-length S/MAR nor the miniMAR element increased episome retention to remarkable levels when incorporated into any of the vectors assessed in this thesis. However, the full-length β-Ifn S/MAR did result in approximately 3 % of the cells becoming permanently transduced when incorporated into both the pEPI plasmid and the single-stranded AAV construct. This suggests that the original S/MAR element does have an effect, albeit small, on transgene retention even when no initial selection pressure is applied. If this investigation were to be pursued further, an integration analysis of the stably expressing cells would be of paramount importance to ascertain the episomal status of the transgene in these cells. The investigations were not continued in this thesis in order to concentrate on the more exciting findings described in Chapter 5.

Although the miniMAR element was found to confer episome retention equivalent to the full-length S/MAR in a minicircle system (Broll *et al.*, 2010), the investigations in this thesis found no supporting evidence for any episome stabilising effect of this element. This may either be due to the different backbone associated with the element, or the fact that initial selection pressure was applied in the previous study. Either way,

the element was not found to enhance the mitotic stability of any of the vectors investigated here.

The cell line used in these *in vitro* experiments was the Chinese Hamster Ovary (CHO) K1 cell line. The rationale for using these cells stems from research by Bode et al into the ability of S/MAR elements to assist in the establishment of mitotically stable episomes, as this work was done with CHO-K1 cells (Broll *et al.*, 2010). Initial experiments therefore focused on the use of these cells, to ensure similar results to Bode et al were obtained, before investigating the potential of other, more clinically relevant cell lines.

In a study looking to increase transgene expression in CHO cells following transfection with an integrating plasmid, it was found that the S/MAR element from human β-globin gene offered the highest level of expression and therefore the most efficient insulator effect (Kim *et al.*, 2004). An interesting future direction might be to try combining this element in some of the episomal systems introduced in this thesis to find out if the beneficial effect is also observed in a non-integrating format.

7.2 Induction of cell cycle arrest using methionine depletion

Methionine and serum depletion is not a commonly used method for the induction of a reversible period of cell cycle arrest in the context of gene addition therapy; drugs such as butyrate are more commonly used in comparable experiments (Broll *et al.*, 2010). Quite possibly for exactly this reason some unexpected results were obtained when using this method to assist in episome establishment.

CHO cells responded to methionine depletion beautifully, entering cell cycle arrest reliably without suffering much cell death upon release into normal cycling. In particular, the proportion of cells in the G_1 phase was always observed to increase at the end of the quiescent period, indicating that the cell cycle arrest occurs at the G_1/S phase checkpoint. Although some variation was observed between experiments in the significance of the differences in proportion of cells in each cell cycle phase, these can mostly be attributed to the somewhat variable growth rate of the control population. CHO cells consistently show a tendency for a decrease in the proportion in S-phase as the population approaches confluency.

Discussion on the effect of methionine depletion on other cell types examined in this thesis is continued in section 7.6.

7.3 The full-length β-IFN S/MAR but not the miniMAR confers a small advantage in retention of non-viral vectors when combined with a transient cell cycle arrest

The non-viral vectors systems investigated in this thesis consisted of pEPI plasmid vectors and viral transfer plasmids containing S/MAR elements, and a minicircle without a S/MAR element.

The purpose of including the pEPI vectors, which have been investigated by other groups, was to find out if the results reported previously can be replicated when no initial selection pressure was applied. The answer was mildly positive; the pEPI vector containing the full-length β-Ifn S/MAR did result in 2-3 % of the CHO cells becoming permanently transfected. However, it is reasonable to assume that at least a proportion of the positive cells contain integrants, as the S/MAR has been noted to promote integration when present in cultured cells for an extended period of time. Indeed, in a selected population of cells transduced with a S/MAR -containing minicircle, approximately 40 % of nuclei showed signs of integration events (Nehlsen *et al.*, 2006). It is therefore reasonable to assume that a similar percentage of the pEPI -transfected cells here contain integrated vectors, which in turn brings the likely proportion of cells containing true episomes below clinical applicability.

The retention of a minicircle vector following a period of induced cell cycle arrest was investigated in order to find out whether the cell cycle arrest in itself is enough to establish an exogenous piece of DNA in the nucleus. As minicircles are devoid of any sequences of bacterial origin, the DNA will not be recognised by the cells for containing prokaryotic methylation patterns, and therefore any established episomes should remain stable. However, no transgene retention was observed. It would be interesting to experiment with the minicircle containing the miniMAR –element described previously, to find out whether retention levels with no selection but with induced cell cycle arrest are similar to those achieved with antibiotic selection. However, as the in-house production of minicircles was not set up, this approach would require a major financial and temporal commitment beyond the possibilities of this thesis.

7.4 The full-length β-IFN S/MAR but not the miniMAR confers a small advantage in retention of AAV vectors when combined with a transient cell cycle arrest

The AAV vectors investigated for potential episomal retention in this thesis included vectors both in the single-stranded and self-complementary format. The single-stranded AAV was engineered to include both the full-length β-IFN S/MAR element and the truncated miniMAR –element. The self-complementary AAV could only carry the miniMAR –element due to its limited packaging capacity.

In both cases, retention of transgene expression was not achieved above 1-3 %. The miniMAR –element was not found to increase retention levels of either vector tested; in fact, it appeared to have a detrimental effect and reduce the retention levels achieved with a comparable vector with no S/MAR element. The only vector displaying notable retention was the ssAAV vector containing the full-length S/MAR when combined with an initial period of cell cycle arrest; however, the 3 % positive population is unlikely to contain only episomes, and therefore the clinical applicability of this finding is likely to be limited.

It is possible that the concatemeric structures often formed by AAV genomes benefited from the inclusion of the S/MAR element and did indeed form stable structures. The integration status of such structures could be further investigated by obtaining clonal populations of from the 3 % GFP+ fraction and subjecting them to an integration status analysis such as detailed in chapter 5. However, due to time limitations and lack of impressively positive findings, combined with simultaneous and impressively positive findings achieved with IDLV vectors, the 3 % GFP+ population was not investigated further.

7.5 IDLV transduction of CHO cells combined with a transient cell cycle arrest results in a high proportion of cells becoming stably transduced with mitotically stable episomes

The transduction of IDLVs into CHO cells followed by a reversible cell cycle arrest produced the most novel, unexpected and exciting results in this thesis. Between 10% and 45% of cells became stably transduced, with little evidence of integration events.

Very recently, a draft genomic sequence of the CHO-K1 cell line was published (Xu *et al.*, 2011). This novel resource was utilised in this thesis to verify the reliability of the few putative integration sites found when comparing amplified sequence fragments against the mouse and rat genomic sequences. Had the hamster sequence been available earlier, it would have enabled a blast analysis of the amplified sequences directly against the hamster genomic sequence. However, the results are very unlikely to be any different, as all sequences found were also found to correlate with the hamster sequence. Moreover, the hamster genome is not assembled or annotated yet, and therefore it is not currently possible to obtain more accurate data on the chromosomal location of the integrants.

To maximise the probability of finding all integrants, a more thorough LAM-PCR using primers binding to the vector backbone as well as the LTRs could be done. Such an analysis was done in another study investigating IDLV integration, although the only integrant was still amplified by the LTR primers (Yanez-Munoz *et al.*, 2006). This method may be more efficient in uncovering integration events that might have happened through non-LTR mediated pathways.

One of the main questions that remain unanswered is how the IDLV -generated episomes are able to replicate; the minimal requirement for mitotic retention is a initiation of replication signal, and considering that the IDLV episomes without S/MAR elements performed as well as ones containing S/MARs, the S/MAR cannot be the replication signal responsible for their retention. The long terminal repeats are one of the features of the IDLVs that distinguish them from other types of episomes investigated in this thesis, and it is possible they could be acting to initiate replication. Another possibility for a replication initiation region in the IDLV episomes is the cPPT, which the SIDD analysis indicates is highly destabilised (Figure 5.1).

To test whether these DNA sequences can confer mitotic stability, minicircles devoid of bacterial backbone but containing one or two copies of the lentiviral LTRs or the cPPT could be constructed and used to transform cells in combination with the methionine depletion protocol; if these sequences are indeed what causes episomal retention, such minicircles should behave in much the same way as the IDLV episomes.

CHO cells were originally derived from a tissue sample from an inbred laboratory animal. They are noted for their chromosome instability (Tjio *et al.*, 1958). It is possible that the stress caused by methionine depletion induces further DNA instability, chromosome breaks and integration of any exogenous DNA present in the nucleus. However, very little evidence for integration events was found during extensive investigations, favouring instead the hypothesis that the quiescent period encouraged the episomes to become permanently associated either with chromosomes or the nuclear matrix.

7.6 IDLV transduction and cell cycle arrest studies in HeLa cells and murine haematopoietic stem cells

Methionine deprivation appears in the literature as a method for slowing down the differentiation of muscle stem cells, as well as a method for inducing apoptosis in cancer cells (Guo *et al.*, 1993; Kitzmann *et al.*, 1998). The results of this thesis suggest that CHO cells as a spontaneously immortalised cell line are able to enter cell cycle arrest induced by methionine depletion, whereas the human tumour cell line HeLa becomes apoptotic instead.

As the successful induction of cell cycle arrest was found to be crucial for the establishment of mitotically stable episomes, the induction of cell cycle arrest in various cell types needs to be optimised further in order to translate the results for clinical applications. Non-tumour human cells would provide an ideal target for such experiments. Further investigations using haematopoietic stem cells would also be beneficial.

HeLa is a noted tumour cell line, which likely affected its ability to undergo cell cycle arrest upon methionine depletion. In this thesis, cell cycle arrest was investigated by monitoring cells numbers, and by employing a propidium iodide stain to assess the

amount of DNA present in each nucleus. Another method for investigating cell proliferation is a BrdU assay, in which bromodeoxyuridine is supplemented in the growth medium and replaces thymidine in any nascent DNA strands during replication. The amount of BrdU can be detected by using antibodies and is indicative of the proportion of cells that were actively replicating during the assay. For any further investigations on the ability of cultured cell to enter cell cycle arrest, it would be useful to employ this assay alongside PI staining.

7.7 Concluding remarks and future directions

The experiments in this thesis were set out to investigate the possibility for creating stable, episomal gene transfer for therapeutic purposes. By combining an IDLV transduction with an induced cell cycle arrest, this aim was achieved. Currently, the method has been optimised for Chinese hamster ovary cells, and work is underway to translate the method for use in more clinically applicable settings. Overall, the ability to generate mitotically stable gene expression from non-integrating episomes would significantly advance the safety aspect of gene addition therapy.

8. Bibliography

Aboody, KS, Najbauer, J, Danks, MK (2008) Stem and progenitor cell-mediated tumor selective gene therapy. *Gene Ther* **15**(10)**:** 739-752.

Ahmed, MR, Berthet, A, Bychkov, E, Porras, G, Li, Q, Bioulac, BH, Carl, YT, Bloch, B, Kook, S, Aubert, I, Dovero, S, Doudnikoff, E, Gurevich, VV, Gurevich, EV, Bezard, E (2010) Lentiviral overexpression of GRK6 alleviates L-dopa-induced dyskinesia in experimental Parkinson's disease. *Sci Transl Med* **2**(28)**:** 28ra28.

Aiuti, A, Cattaneo, F, Galimberti, S, Benninghoff, U, Cassani, B, Callegaro, L, Scaramuzza, S, Andolfi, G, Mirolo, M, Brigida, I, Tabucchi, A, Carlucci, F, Eibl, M, Aker, M, Slavin, S, Al-Mousa, H, Al Ghonaium, A, Ferster, A, Duppenthaler, A, Notarangelo, L, Wintergerst, U, Buckley, RH, Bregni, M, Marktel, S, Valsecchi, MG, Rossi, P, Ciceri, F, Miniero, R, Bordignon, C, Roncarolo, MG (2009) Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med* **360**(5)**:** 447-458.

Alcaro, S, Gaspar, A, Ortuso, F, Milhazes, N, Orallo, F, Uriarte, E, Yanez, M, Borges, F (2010) Chromone-2- and -3-carboxylic acids inhibit differently monoamine oxidases A and B. *Bioorg Med Chem Lett* **20**(9)**:** 2709-2712.

Allocca, M, Doria, M, Petrillo, M, Colella, P, Garcia-Hoyos, M, Gibbs, D, Kim, SR, Maguire, A, Rex, TS, Di Vicino, U, Cutillo, L, Sparrow, JR, Williams, DS, Bennett, J, Auricchio, A (2008) Serotype-dependent packaging of large genes in adeno-associated viral vectors results in effective gene delivery in mice. *J Clin Invest* **118**(5)**:** 1955-1964.

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

Argyros, O, Wong, SP, Niceta, M, Waddington, SN, Howe, SJ, Coutelle, C, Miller, AD, Harbottle, RP (2008) Persistent episomal transgene expression in liver following delivery of a scaffold/matrix attachment region containing non-viral vector. *Gene Ther* **15**(24)**:** 1593-1605.

Azzouz, M, Le, T, Ralph, GS, Walmsley, L, Monani, UR, Lee, DC, Wilkes, F, Mitrophanous, KA, Kingsman, SM, Burghes, AH, Mazarakis, ND (2004) Lentivector-mediated SMN replacement in a mouse model of spinal muscular atrophy. *J Clin Invest* **114**(12)**:** 1726-1731.

Baiker, A, Maercker, C, Piechaczek, C, Schmidt, SB, Bode, J, Benham, C, Lipps, HJ (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**(3)**:** 182-184.

Barre-Sinoussi, F, Chermann, JC, Rey, F, Nugeyre, MT, Chamaret, S, Gruest, J, Dauguet, C, Axler-Blin, C, Vezinet-Brun, F, Rouzioux, C, Rozenbaum, W, Montagnier, L (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**(4599)**:** 868-871.

Barteau, B, Chevre, R, Letrou-Bonneval, E, Labas, R, Lambert, O, Pitard, B (2008) Physicochemical parameters of non-viral vectors that govern transfection efficiency. *Curr Gene Ther* **8**(5)**:** 313-323.

Bauer, S, Kirschning, CJ, Hacker, H, Redecke, V, Hausmann, S, Akira, S, Wagner, H, Lipford, GB (2001) Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proc Natl Acad Sci U S A* **98**(16)**:** 9237-9242.

Bayer, M, Kantor, B, Cockrell, A, Ma, H, Zeithaml, B, Li, X, McCown, T, Kafri, T (2008) A large U3 deletion causes increased in vivo expression from a nonintegrating lentiviral vector. *Mol Ther* **16**(12)**:** 1968-1976.

Benham, C, Kohwi-Shigematsu, T, Bode, J (1997) Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. *J Mol Biol* **274**(2)**:** 181-196.

Benham, CJ (1996) Duplex destabilization in superhelical DNA is predicted to occur at specific transcriptional regulatory regions. *J Mol Biol* **255**(3)**:** 425-434.

Benham, CJ (2001) Stress-induced DNA duplex destabilization in transcriptional initiation. *Pac Symp Biocomput***:** 103-114.

Benham, CJ, Bi, C (2004) The analysis of stress-induced duplex destabilization in long genomic DNA sequences. *J Comput Biol* **11**(4)**:** 519-543.

Bernabei, PA, Santini, V, Silvestro, L, Dal Pozzo, O, Bezzini, R, Viano, I, Gattei, V, Saccardi, R, Rossi Ferrini, P (1989) In vitro chemosensitivity testing of leukemic cells: development of a semiautomated colorimetric assay. *Hematol Oncol* **7**(3)**:** 243-253.

Bi, C, Benham, CJ (2004) WebSIDD: server for predicting stress-induced duplex destabilized (SIDD) sites in superhelical DNA. *Bioinformatics* **20**(9)**:** 1477-1479.

Bigger, BW, Tolmachov, O, Collombet, JM, Fragkos, M, Palaszewski, I, Coutelle, C (2001) An araC-controlled bacterial cre expression system to produce DNA minicircle vectors for nuclear and mitochondrial gene therapy. *J Biol Chem* **276**(25)**:** 23018-23027.

Blaese, RM, Culver, KW, Miller, AD, Carter, CS, Fleisher, T, Clerici, M, Shearer, G, Chang, L, Chiang, Y, Tolstoshev, P, Greenblatt, JJ, Rosenberg, SA, Klein, H, Berger, M, Mullen, CA, Ramsey, WJ, Muul, L, Morgan, RA, Anderson, WF (1995) T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science* **270**(5235)**:** 475-480.

Bode, J, Kohwi, Y, Dickinson, L, Joh, T, Klehr, D, Mielke, C, Kohwi-Shigematsu, T (1992) Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science* **255**(5041)**:** 195-197.

Bode, J, Maass, K (1988) Chromatin domain surrounding the human interferon-beta gene as defined by scaffold-attached regions. *Biochemistry* **27**(13)**:** 4706-4711.

Bode, J, Schlake, T, Rios-Ramirez, M, Mielke, C, Stengert, M, Kay, V, 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, KP, Benham, C (2006) Correlations between scaffold/matrix attachment region (S/MAR) binding activity and DNA duplex destabilization energy. *J Mol Biol* **358**(2)**:** 597-613.

Boztug, K, Schmidt, M, Schwarzer, A, Banerjee, PP, Diez, IA, Dewey, RA, Bohm, M, Nowrouzi, A, Ball, CR, Glimm, H, Naundorf, S, Kuhlcke, K, Blasczyk, R, Kondratenko, I, Marodi, L, Orange, JS, von Kalle, C, Klein, C (2010) Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N Engl J Med* **363**(20)**:** 1918-1927.

Brantly, ML, Spencer, LT, Humphries, M, Conlon, TJ, Spencer, CT, Poirier, A, Garlington, W, Baker, D, Song, S, Berns, KI, Muzyczka, N, Snyder, RO, Byrne, BJ, Flotte, TR (2006) Phase I trial of intramuscular injection of a recombinant adeno-associated virus serotype 2 alphal-antitrypsin (AAT) vector in AAT-deficient adults. *Hum Gene Ther* **17**(12)**:** 1177-1186.

Braun, S (2008) Muscular gene transfer using nonviral vectors. *Curr Gene Ther* **8**(5)**:** 391-405.

Broll, S, Oumard, A, Hahn, K, Schambach, A, Bode, J (2010) Minicircle performance depending on S/MAR-nuclear matrix interactions. *J Mol Biol* **395**(5)**:** 950-965.

Broudy, VC (1997) Stem cell factor and hematopoiesis. *Blood* **90**(4)**:** 1345-1364.

Brown, PO, Bowerman, B, Varmus, HE, Bishop, JM (1989) Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc Natl Acad Sci U S A* **86**(8)**:** 2525-2529.

Buller, RM, Janik, JE, Sebring, ED, Rose, JA (1981) Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. *J Virol* **40**(1)**:** 241-247.

Burns, JC, Friedmann, T, Driever, W, Burrascano, M, Yee, JK (1993) Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc Natl Acad Sci U S A* **90**(17)**:** 8033-8037.

Calcedo, R, Morizono, H, Wang, L, McCarter, R, He, J, Jones, D, Batshaw, ML, Wilson, JM (2011) Adeno-associated virus antibody profiles in newborns, children and adolescents. *Clin Vaccine Immunol*.

Cartier, N, Hacein-Bey-Abina, S, Bartholomae, CC, Veres, G, Schmidt, M, Kutschera, I, Vidaud, M, Abel, U, Dal-Cortivo, L, Caccavelli, L, Mahlaoui, N, Kiermer, V, Mittelstaedt, D, Bellesme, C, Lahlou, N, Lefrere, F, Blanche, S, Audit, M, Payen, E, Leboulch, P, l'Homme, B, Bougneres, P, Von Kalle, C, Fischer, A, Cavazzana-Calvo, M, Aubourg, P (2009) Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* **326**(5954)**:** 818-823.

Case, SS, Price, MA, Jordan, CT, Yu, XJ, Wang, L, Bauer, G, Haas, DL, Xu, D, Stripecke, R, Naldini, L, Kohn, DB, Crooks, GM (1999) Stable transduction of quiescent CD34(+)CD38(-) human hematopoietic cells by HIV-1-based lentiviral vectors. *Proc Natl Acad Sci U S A* **96**(6)**:** 2988-2993.

Cavazzana-Calvo, M, Payen, E, Negre, O, Wang, G, Hehir, K, Fusil, F, Down, J, Denaro, M, Brady, T, Westerman, K, Cavallesco, R, Gillet-Legrand, B, Caccavelli, L, Sgarra, R, Maouche-Chretien, L, Bernaudin, F, Girot, R, Dorazio, R, Mulder, GJ, Polack, A, Bank, A, Soulier, J, Larghero, J, Kabbara, N, Dalle, B, Gourmel, B, Socie, G, Chretien, S, Cartier, N, Aubourg, P, Fischer, A, Cornetta, K, Galacteros, F, Beuzard, Y, Gluckman, E, Bushman, F, Hacein-Bey-Abina, S, Leboulch, P (2010) Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* **467**(7313)**:** 318-322.

Chen, ZY, He, CY, Ehrhardt, A, Kay, MA (2003) Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo. *Mol Ther* **8**(3)**:** 495-500.

Chen, ZY, He, CY, Kay, MA (2005) Improved production and purification of minicircle DNA vector free of plasmid bacterial sequences and capable of persistent transgene expression in vivo. *Hum Gene Ther* **16**(1)**:** 126-131.

Chen, ZY, Riu, E, He, CY, Xu, H, Kay, MA (2008) Silencing of episomal transgene expression in liver by plasmid bacterial backbone DNA is independent of CpG methylation. *Mol Ther* **16**(3)**:** 548-556.

Choi, VW, Samulski, RJ, McCarty, DM (2005) Effects of adeno-associated virus DNA hairpin structure on recombination. *J Virol* **79**(11)**:** 6801-6807.

Conese, M, Auriche, C, Ascenzioni, F (2004) Gene therapy progress and prospects: episomally maintained self-replicating systems. *Gene Ther* **11**(24)**:** 1735-1741.

Conner, J, Rixon, FJ, Brown, SM (2005) Herpes simplex virus type 1 strain HSV1716 grown in baby hamster kidney cells has altered tropism for nonpermissive Chinese hamster ovary cells compared to HSV1716 grown in vero cells. *J Virol* **79**(15)**:** 9970-9981.

Cook, PR (1999) The organization of replication and transcription. *Science* **284**(5421)**:** 1790-1795.

Cory, AH, Owen, TC, Barltrop, JA, Cory, JG (1991) Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. *Cancer Commun* **3**(7)**:** 207-212.

Dang, Q, Auten, J, 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**(6)**:** 2671-2678.

Darquet, AM, Cameron, B, Wils, P, Scherman, D, Crouzet, J (1997) A new DNA vehicle for nonviral gene delivery: supercoiled minicircle. *Gene Ther* **4**(12)**:** 1341-1349.

Darquet, AM, Rangara, R, Kreiss, P, Schwartz, B, Naimi, S, Delaere, P, Crouzet, J, Scherman, D (1999) Minicircle: an improved DNA molecule for in vitro and in vivo gene transfer. *Gene Ther* **6**(2)**:** 209-218.

Davis, HE, Rosinski, M, Morgan, JR, Yarmush, ML (2004) Charged polymers modulate retrovirus transduction via membrane charge neutralization and virus aggregation. *Biophys J* **86**(2)**:** 1234-1242.

Davis, PK, Ho, A, Dowdy, SF (2001) Biological methods for cell-cycle synchronization of mammalian cells. *Biotechniques* **30**(6)**:** 1322-1326, 1328, 1330-1321.

Deaven, LL, Petersen, DF (1973) The chromosomes of CHO, an aneuploid Chinese hamster cell line: G-band, C-band, and autoradiographic analyses. *Chromosoma* **41**(2)**:** 129-144.

Di Mayorca, G, Callender, J, Marin, G, Giordano, R (1969) Temperature-sensitive mutants of polyoma virus. *Virology* **38**(1)**:** 126-133.

Dong, B, Nakai, H, Xiao, W (2010) Characterization of genome integrity for oversized recombinant AAV vector. *Mol Ther* **18**(1)**:** 87-92.

Duan, D, Sharma, P, Dudus, L, Zhang, Y, Sanlioglu, S, Yan, Z, Yue, Y, Ye, Y, Lester, R, Yang, J, Fisher, KJ, Engelhardt, JF (1999a) Formation of adeno-associated virus circular genomes is differentially regulated by adenovirus E4 ORF6 and E2a gene expression. *J Virol* **73**(1)**:** 161-169.

Duan, D, Sharma, P, Yang, J, Yue, Y, Dudus, L, Zhang, Y, Fisher, KJ, Engelhardt, JF (1998) Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. *J Virol* **72**(11)**:** 8568-8577.

Duan, D, Yan, Z, Yue, Y, Engelhardt, JF (1999b) Structural analysis of adeno-associated virus transduction circular intermediates. *Virology* **261**(1)**:** 8-14.

Duan, D, Yue, Y, Engelhardt, JF (2001) Expanding AAV packaging capacity with trans-splicing or overlapping vectors: a quantitative comparison. *Mol Ther* **4**(4)**:** 383-391.

Dull, T, Zufferey, R, Kelly, M, Mandel, RJ, Nguyen, M, Trono, D, Naldini, L (1998) A third-generation lentivirus vector with a conditional packaging system. *J Virol* **72**(11)**:** 8463-8471.

Durocher, Y, Pham, PL, St-Laurent, G, Jacob, D, Cass, B, Chahal, P, Lau, CJ, Nalbantoglu, J, Kamen, A (2007) Scalable serum-free production of recombinant adeno-associated virus type 2 by transfection of 293 suspension cells. *J Virol Methods* **144**(1-2)**:** 32-40.

Dykstra, B, Kent, D, Bowie, M, McCaffrey, L, Hamilton, M, Lyons, K, Lee, SJ, Brinkman, R, Eaves, C (2007) Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* **1**(2)**:** 218-229.

Eliopoulos, N, Al-Khaldi, A, Crosato, M, Lachapelle, K, Galipeau, J (2003) A neovascularized organoid derived from retrovirally engineered bone marrow stroma leads to prolonged in vivo systemic delivery of erythropoietin in nonmyeloablated, immunocompetent mice. *Gene Ther* **10**(6)**:** 478-489.

Emerson, M, Renwick, L, Tate, S, Rhind, S, Milne, E, Painter, HA, Boyd, AC, McLachlan, G, Griesenbach, U, Cheng, SH, Gill, DR, Hyde, SC, Baker, A, Alton, EW, Porteous, DJ, Collie, DD (2003) Transfection efficiency and toxicity following delivery of naked plasmid DNA and cationic lipid-DNA complexes to ovine lung segments. *Mol Ther* **8**(4)**:** 646-653.

Engelman, A, Englund, G, Orenstein, JM, Martin, MA, Craigie, R (1995) Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol* **69**(5)**:** 2729-2736.

Evans, K, Ott, S, Hansen, A, Koentges, G, Wernisch, L (2007) A comparative study of S/MAR prediction tools. *BMC Bioinformatics* **8:** 71.

Fabre, JW, Grehan, A, Whitehorne, M, Sawyer, GJ, Dong, X, Salehi, S, Eckley, L, Zhang, X, Seddon, M, Shah, AM, Davenport, M, Rela, M (2008) Hydrodynamic gene delivery to the pig liver via an isolated segment of the inferior vena cava. *Gene Ther* **15**(6)**:** 452-462.

Feigin, A, Kaplitt, MG, Tang, C, Lin, T, Mattis, P, Dhawan, V, During, MJ, Eidelberg, D (2007) Modulation of metabolic brain networks after subthalamic gene therapy for Parkinson's disease. *Proc Natl Acad Sci U S A* **104**(49)**:** 19559-19564.

Feng, Y, Broder, CC, Kennedy, PE, Berger, EA (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**(5263)**:** 872-877.

Ferrari, FK, Samulski, T, Shenk, T, Samulski, RJ (1996) Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J Virol* **70**(5)**:** 3227-3234.

Fujiwara, T, Mizuuchi, K (1988) Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**(4)**:** 497-504.

Gabriel, R, Eckenberg, R, Paruzynski, A, Bartholomae, CC, Nowrouzi, A, Arens, A, Howe, SJ, Recchia, A, Cattoglio, C, Wang, W, Faber, K, Schwarzwaelder, K, Kirsten, R, Deichmann, A, Ball, CR, Balaggan, KS, Yanez-Munoz, RJ, Ali, RR, Gaspar, HB, Biasco, L, Aiuti, A, Cesana, D, Montini, E, Naldini, L, Cohen-Haguenauer, O, Mavilio, F, Thrasher, AJ, Glimm, H, von Kalle, C, Saurin, W, Schmidt, M (2009) Comprehensive genomic access to vector integration in clinical gene therapy. *Nat Med* **15**(12)**:** 1431-1436.

Gallo, RC, Wong-Staal, F (1985) A human T-lymphotropic retrovirus (HTLV-III) as the cause of the acquired immunodeficiency syndrome. *Ann Intern Med* **103**(5)**:** 679-689.

Gao, G, Vandenberghe, LH, Alvira, MR, Lu, Y, Calcedo, R, Zhou, X, Wilson, JM (2004) Clades of Adeno-associated viruses are widely disseminated in human tissues. *J Virol* **78**(12)**:** 6381-6388.

Gascoigne, KE, Taylor, SS (2009) How do anti-mitotic drugs kill cancer cells? *J Cell Sci* **122**(Pt 15)**:** 2579-2585.

Gaur, M, Leavitt, AD (1998) Mutations in the human immunodeficiency virus type 1 integrase D,D(35)E motif do not eliminate provirus formation. *J Virol* **72**(6)**:** 4678-4685.

Gelderblom, HR, Hausmann, EH, Ozel, M, Pauli, G, Koch, MA (1987) Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* **156**(1)**:** 171-176.

Giannakopoulos, A, Stavrou, EF, Zarkadis, I, Zoumbos, N, Thrasher, AJ, 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**(5)**:** 1239-1249.

Girod, A, Wobus, CE, Zadori, Z, Ried, M, Leike, K, Tijssen, P, Kleinschmidt, JA, Hallek, M (2002) The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *J Gen Virol* **83**(Pt 5)**:** 973-978.

Goetze, S, Baer, A, Winkelmann, S, Nehlsen, K, Seibler, J, Maass, K, Bode, J (2005) Performance of genomic bordering elements at predefined genomic loci. *Mol Cell Biol* **25**(6)**:** 2260-2272.

Goncalves, MA (2005) Adeno-associated virus: from defective virus to effective vector. *Virol J* **2:** 43.

Gopal, V, Xavier, J, Dar, GH, Jafurulla, M, Chattopadhyay, A, Rao, NM (2011) Targeted liposomes to deliver DNA to cells expressing 5-HT receptors. *Int J Pharm*.

Goren, A, Dahan, N, Goren, E, Baruch, L, Machluf, M (2010) Encapsulated human mesenchymal stem cells: a unique hypoimmunogenic platform for long-term cellular therapy. *FASEB J* **24**(1)**:** 22-31.

Groth, AC, Olivares, EC, Thyagarajan, B, Calos, MP (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc Natl Acad Sci U S A* **97**(11)**:** 5995-6000.

Guo, H, Lishko, VK, Herrera, H, Groce, A, Kubota, T, Hoffman, RM (1993) Therapeutic tumor-specific cell cycle block induced by methionine starvation in vivo. *Cancer Res* **53**(23)**:** 5676-5679.

Haase, R, Argyros, O, Wong, SP, Harbottle, RP, Lipps, HJ, Ogris, M, Magnusson, T, Vizoso Pinto, MG, Haas, J, Baiker, A (2010) pEPito: a significantly improved non-viral episomal expression vector for mammalian cells. *BMC Biotechnol* **10:** 20.

Habeshaw, JA, Dalgleish, AG, Bountiff, L, Newell, AL, Wilks, D, Walker, LC, Manca, F (1990) AIDS pathogenesis: HIV envelope and its interaction with cell proteins. *Immunol Today* **11**(11)**:** 418-425.

Hacein-Bey-Abina, S, Garrigue, A, Wang, GP, Soulier, J, Lim, A, Morillon, E, Clappier, E, Caccavelli, L, Delabesse, E, Beldjord, K, Asnafi, V, MacIntyre, E, Dal Cortivo, L, Radford, I, Brousse, N, Sigaux, F, Moshous, D, Hauer, J, Borkhardt, A, Belohradsky, BH, Wintergerst, U, Velez, MC, Leiva, L, Sorensen, R, Wulffraat, N, Blanche, S, Bushman, FD, Fischer, A, Cavazzana-Calvo, M (2008) Insertional

oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**(9)**:** 3132-3142.

Hacein-Bey-Abina, S, Le Deist, F, Carlier, F, Bouneaud, C, Hue, C, De Villartay, JP, Thrasher, AJ, Wulffraat, N, Sorensen, R, Dupuis-Girod, S, Fischer, A, Davies, EG, Kuis, W, Leiva, L, Cavazzana-Calvo, M (2002) Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med* **346**(16)**:** 1185-1193.

Hacein-Bey-Abina, S, Von Kalle, C, Schmidt, M, McCormack, MP, Wulffraat, N, Leboulch, P, Lim, A, Osborne, CS, Pawliuk, R, Morillon, E, Sorensen, R, Forster, A, Fraser, P, Cohen, JI, de Saint Basile, G, Alexander, I, Wintergerst, U, Frebourg, T, Aurias, A, Stoppa-Lyonnet, D, Romana, S, Radford-Weiss, I, Gross, F, Valensi, F, Delabesse, E, Macintyre, E, Sigaux, F, Soulier, J, Leiva, LE, Wissler, M, Prinz, C, Rabbitts, TH, Le Deist, F, Fischer, A, Cavazzana-Calvo, M (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**(5644)**:** 415-419.

Hagedorn, C, Wong, SP, Harbottle, R, Lipps, HJ (2011) Scaffold/Matrix attached region-based nonviral episomal vectors. *Hum Gene Ther* **22**(8)**:** 915-923.

Hagino-Yamagishi, K, Donehower, LA, Varmus, HE (1987) Retroviral DNA integrated during infection by an integration-deficient mutant of murine leukemia virus is oligomeric. *J Virol* **61**(6)**:** 1964-1971.

Hansen, S, Lehr, CM (2011) Nanoparticles for transcutaneous vaccination. *Microb Biotechnol*.

Harraghy, N, Gaussin, A, Mermod, N (2008) Sustained transgene expression using MAR elements. *Curr Gene Ther* **8**(5)**:** 353-366.

Herweijer, H, Wolff, JA (2003) Progress and prospects: naked DNA gene transfer and therapy. *Gene Ther* **10**(6)**:** 453-458.

Herzog, RW, Yang, EY, Couto, LB, Hagstrom, JN, Elwell, D, Fields, PA, Burton, M, Bellinger, DA, Read, MS, Brinkhous, KM, Podsakoff, GM, Nichols, TC, Kurtzman, GJ, High, KA (1999) Long-term correction of canine hemophilia B by gene transfer of blood coagulation factor IX mediated by adeno-associated viral vector. *Nat Med* **5**(1)**:** 56-63.

Higashimoto, T, Urbinati, F, Perumbeti, A, Jiang, G, Zarzuela, A, Chang, LJ, Kohn, DB, Malik, P (2007) The woodchuck hepatitis virus post-transcriptional regulatory element reduces readthrough transcription from retroviral vectors. *Gene Ther* **14**(17)**:** 1298-1304.

Higuchi, R, Dollinger, G, Walsh, PS, Griffith, R (1992) Simultaneous amplification and detection of specific DNA sequences. *Biotechnology (N Y)* **10**(4)**:** 413-417.

Hoess, R, Wierzbicki, A, Abremski, K (1985) Formation of small circular DNA molecules via an in vitro site-specific recombination system. *Gene* **40**(2-3)**:** 325-329.

Hollon, T (2000) Researchers and regulators reflect on first gene therapy death. *Nat Med* **6**(1)**:** 6.

Howden, SE, Voullaire, L, Wardan, H, Williamson, R, Vadolas, J (2008) Site-specific, Rep-mediated integration of the intact beta-globin locus in the human erythroleukaemic cell line K562. *Gene Ther* **15**(20)**:** 1372-1383.

Huser, D, Weger, S, Heilbronn, R (2002) Kinetics and frequency of adeno-associated virus site-specific integration into human chromosome 19 monitored by quantitative real-time PCR. *J Virol* **76**(15)**:** 7554-7559.

Hyde, SC, Pringle, IA, Abdullah, S, Lawton, AE, Davies, LA, Varathalingam, A, Nunez-Alonso, G, Green, AM, Bazzani, RP, Sumner-Jones, SG, Chan, M, Li, H, Yew, NS, Cheng, SH, Boyd, AC, Davies, JC, Griesenbach, U, Porteous, DJ, Sheppard, DN, Munkonge, FM, Alton, EW, Gill, DR (2008) CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression. *Nat Biotechnol* **26**(5)**:** 549-551.

Intra, J, Salem, AK (2011) Rational design, fabrication, characterization and in vitro testing of biodegradable microparticles that generate targeted and sustained transgene expression in HepG2 liver cells. *J Drug Target* **19**(6)**:** 393-408.

Ivics, Z, Izsvak, Z (2011) Non-viral Gene Delivery with the Sleeping Beauty Transposon System. *Hum Gene Ther*.

Jacobs, F, Snoeys, J, Feng, Y, Van Craeyveld, E, Lievens, J, Armentano, D, Cheng, SH, De Geest, B (2008) Direct comparison of hepatocyte-specific expression cassettes following adenoviral and nonviral hydrodynamic gene transfer. *Gene Ther* **15**(8)**:** 594-603.

Jakobsen, J, Mikkelsen, J, Nielsen, A (2010) Elimination of the plasmid bacterial backbone in site-directed transgenesis. *Biotechniques* **48**(4)**:** 313-316.

Jenke, AC, Stehle, IM, Herrmann, F, Eisenberger, T, Baiker, A, Bode, J, Fackelmayer, FO, Lipps, HJ (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**(31)**:** 11322-11327.

Jenke, BH, Fetzer, CP, Stehle, IM, Jonsson, F, Fackelmayer, FO, Conradt, H, Bode, J, Lipps, HJ (2002) An episomally replicating vector binds to the nuclear matrix protein SAF-A in vivo. *EMBO Rep* **3**(4)**:** 349-354.

Jia, F, Wilson, KD, Sun, N, Gupta, DM, Huang, M, Li, Z, Panetta, NJ, Chen, ZY, Robbins, RC, Kay, MA, Longaker, MT, Wu, JC (2010) A nonviral minicircle vector for deriving human iPS cells. *Nat Methods* **7**(3)**:** 197-199.

Jiang, WQ, Zhong, ZH, Henson, JD, Reddel, RR (2007) Identification of candidate alternative lengthening of telomeres genes by methionine restriction and RNA interference. *Oncogene* **26**(32)**:** 4635-4647.

Kao, FT, Puck, TT (1967) Genetics of somatic mammalian cells. IV. Properties of Chinese hamster cell mutants with respect to the requirement for proline. *Genetics* **55**(3)**:** 513-524.

Kao, FT, Puck, TT (1970) Genetics of somatic mammalian cells: linkage studies with human-Chinese hamster cell hybrids. *Nature* **228**(5269)**:** 329-332.

Kaplitt, MG, Feigin, A, Tang, C, Fitzsimons, HL, Mattis, P, Lawlor, PA, Bland, RJ, Young, D, Strybing, K, Eidelberg, D, During, MJ (2007) Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial. *Lancet* **369**(9579)**:** 2097-2105.

Katsumoto, T, Hattori, N, Yamada, O, Kurimura, T (1988) Intermediate virions in maturation of human immunodeficiency virus, strain LAV. *J Electron Microsc (Tokyo)* **37**(4)**:** 205-207.

Kay, MA, He, CY, Chen, ZY (2010) A robust system for production of minicircle DNA vectors. *Nat Biotechnol* **28**(12)**:** 1287-1289.

Khaled, EG, Saleh, M, Hindocha, S, Griffin, M, Khan, WS (2011) Tissue engineering for bone production- stem cells, gene therapy and scaffolds. *Open Orthop J* **5 Suppl 2:** 289-295.

Khorsandi, SE, Bachellier, P, Weber, JC, Greget, M, Jaeck, D, Zacharoulis, D, Rountas, C, Helmy, S, Helmy, A, Al-Waracky, M, Salama, H, Jiao, L, Nicholls, J, Davies, AJ, Levicar, N, Jensen, S, Habib, N (2008) Minimally invasive and selective hydrodynamic gene therapy of liver segments in the pig and human. *Cancer Gene Ther* **15**(4)**:** 225-230.

Khoury, AT, Hanafusa, H (1976) Synethesis and integration of viral DNA in chicken cells at different time after infection with various multiplicities of avian oncornavirus. *J Virol* **18**(2)**:** 383-400.

Kim, JM, Kim, JS, Park, DH, Kang, HS, Yoon, J, Baek, K, Yoon, Y (2004) Improved recombinant gene expression in CHO cells using matrix attachment regions. *J Biotechnol* **107**(2)**:** 95-105.

King, JA, Dubielzig, R, Grimm, D, Kleinschmidt, JA (2001) DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J* **20**(12)**:** 3282-3291.

Kingsman, SM, Mitrophanous, K, Olsen, JC (2005) Potential oncogene activity of the woodchuck hepatitis post-transcriptional regulatory element (WPRE). *Gene Ther* **12**(1)**:** 3-4.

Kipp, M, Gohring, F, Ostendorp, T, van Drunen, CM, van Driel, R, Przybylski, M, Fackelmayer, FO (2000) SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol Cell Biol* **20**(20)**:** 7480-7489.

Kitzmann, M, Carnac, G, Vandromme, M, Primig, M, Lamb, NJ, Fernandez, A (1998) The muscle regulatory factors MyoD and myf-5 undergo distinct cell cycle-specific expression in muscle cells. *J Cell Biol* **142**(6)**:** 1447-1459.

Klar, M, Stellamanns, E, Ak, P, Gluch, A, Bode, J (2005) Dominant genomic structures: detection and potential signal functions in the interferon-beta domain. *Gene* **364:** 79-89.

Kotin, RM, Linden, RM, Berns, KI (1992) Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J* **11**(13)**:** 5071-5078.

Kotin, RM, Siniscalco, M, Samulski, RJ, Zhu, XD, Hunter, L, Laughlin, CA, McLaughlin, S, Muzyczka, N, Rocchi, M, Berns, KI (1990) Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci U S A* **87**(6)**:** 2211-2215.

Kues, WA, Anger, M, Carnwath, JW, Paul, D, Motlik, J, Niemann, H (2000) Cell cycle synchronization of porcine fetal fibroblasts: effects of serum deprivation and reversible cell cycle inhibitors. *Biol Reprod* **62**(2)**:** 412-419.

Lai, Y, Li, D, Yue, Y, Duan, D (2008) Design of trans-splicing adeno-associated viral vectors for Duchenne muscular dystrophy gene therapy. *Methods Mol Biol* **433:** 259-275.

Lai, Y, Yue, Y, Duan, D (2009) Evidence for the failure of adeno-associated virus serotype 5 to package a viral genome > or = 8.2 kb. *Mol Ther* **18**(1)**:** 75-79.

Leavitt, AD, Robles, G, Alesandro, N, Varmus, HE (1996) Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. *J Virol* **70**(2)**:** 721-728.

Ledwith, BJ, Manam, S, Troilo, PJ, Barnum, AB, Pauley, CJ, Griffiths, TG, 2nd, Harper, LB, Beare, CM, Bagdon, WJ, Nichols, WW (2000) Plasmid DNA vaccines: investigation of integration into host cellular DNA following intramuscular injection in mice. *Intervirology* **43**(4-6)**:** 258-272.

Leis, J, Baltimore, D, Bishop, JM, Coffin, J, Fleissner, E, Goff, SP, Oroszlan, S, Robinson, H, Skalka, AM, Temin, HM, et al. (1988) Standardized and simplified nomenclature for proteins common to all retroviruses. *J Virol* **62**(5)**:** 1808-1809.

Lewis, P, Hensel, M, Emerman, M (1992) Human immunodeficiency virus infection of cells arrested in the cell cycle. *EMBO J* **11**(8)**:** 3053-3058.

Ley, KD, Tobey, RA (1970) REGULATION OF INITIATION OF DNA SYNTHESIS IN CHINESE HAMSTER CELLS : II. Induction of DNA Synthesis and Cell Division by Isoleucine and Glutamine in G(1)-Arrested Cells in Suspension Culture. *J Cell Biol* **47**(2)**:** 453-459.

Li, W, Asokan, A, Wu, Z, Van Dyke, T, DiPrimio, N, Johnson, JS, Govindaswamy, L, Agbandje-McKenna, M, Leichtle, S, Redmond, DE, Jr., McCown, TJ, Petermann, KB, Sharpless, NE, Samulski, RJ (2008) Engineering and selection of shuffled AAV genomes: a new strategy for producing targeted biological nanoparticles. *Mol Ther* **16**(7)**:** 1252-1260.

Liebich, I, Bode, J, Frisch, M, Wingender, E (2002) S/MARt DB: a database on scaffold/matrix attached regions. *Nucleic Acids Res* **30**(1)**:** 372-374.

Logan, AC, Haas, DL, Kafri, T, Kohn, DB (2004) Integrated self-inactivating lentiviral vectors produce full-length genomic transcripts competent for encapsidation and integration. *J Virol* **78**(16)**:** 8421-8436.

Lostal, W, Bartoli, M, Bourg, N, Roudaut, C, Bentaib, A, Miyake, K, Guerchet, N, Fougerousse, F, McNeil, P, Richard, I (2010) Efficient recovery of dysferlin deficiency by dual adeno-associated vector-mediated gene transfer. *Hum Mol Genet* **19**(10)**:** 1897-1907.

Lu, S, Hoestje, SM, Choo, E, Epner, DE (2003) Induction of caspase-dependent and -independent apoptosis in response to methionine restriction. *Int J Oncol* **22**(2)**:** 415-420.

Lufino, MM, Edser, PA, Wade-Martins, R (2008) Advances in high-capacity extrachromosomal vector technology: episomal maintenance, vector delivery, and transgene expression. *Mol Ther* **16**(9)**:** 1525-1538.

Luskey, BD, Rosenblatt, M, Zsebo, K, Williams, DA (1992) Stem cell factor, interleukin-3, and interleukin-6 promote retroviral-mediated gene transfer into murine hematopoietic stem cells. *Blood* **80**(2)**:** 396-402.

Macville, M, Schrock, E, Padilla-Nash, H, Keck, C, Ghadimi, BM, Zimonjic, D, Popescu, N, Ried, T (1999) Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. *Cancer Res* **59**(1)**:** 141-150.

Maguire, AM, High, KA, Auricchio, A, Wright, JF, Pierce, EA, Testa, F, Mingozzi, F, Bennicelli, JL, Ying, GS, Rossi, S, Fulton, A, Marshall, KA, Banfi, S, Chung, DC, Morgan, JI, Hauck, B, Zelenaia, O, Zhu, X, Raffini, L, Coppieters, F, De Baere, E, Shindler, KS, Volpe, NJ, Surace, EM, Acerra, C, Lyubarsky, A, Redmond, TM, Stone, E, Sun, J, McDonnell, JW, Leroy, BP, Simonelli, F, Bennett, J (2009) Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. *Lancet* **374**(9701)**:** 1597-1605.

Manfredsson, FP, Rising, AC, Mandel, RJ (2009) AAV9: a potential blood-brain barrier buster. *Mol Ther* **17**(3)**:** 403-405.

Manzini, S, Vargiolu, A, Stehle, IM, Bacci, ML, Cerrito, MG, Giovannoni, R, Zannoni, A, Bianco, MR, Forni, M, Donini, P, Papa, M, Lipps, HJ, Lavitrano, M (2006) Genetically modified pigs produced with a nonviral episomal vector. *Proc Natl Acad Sci U S A* **103**(47)**:** 17672-17677.

Marie, C, Vandermeulen, G, Quiviger, M, Richard, M, Preat, V, Scherman, D (2010) pFARs, plasmids free of antibiotic resistance markers, display high-level transgene expression in muscle, skin and tumour cells. *J Gene Med* **12**(4)**:** 323-332.

Martinez-Chantar, ML, Latasa, MU, Varela-Rey, M, Lu, SC, Garcia-Trevijano, ER, Mato, JM, Avila, MA (2003) L-methionine availability regulates expression of the methionine adenosyltransferase 2A gene in human hepatocarcinoma cells: role of S-adenosylmethionine. *J Biol Chem* **278**(22)**:** 19885-19890.

Matrai, J, Cantore, A, Bartholomae, CC, Annoni, A, Wang, W, Acosta-Sanchez, A, Samara-Kuko, E, De Waele, L, Ma, L, Genovese, P, Damo, M, Arens, A, Goudy, K, Nichols, TC, von Kalle, C, MK, LC, Roncarolo, MG, Schmidt, M, Vandendriessche, T,

Naldini, L (2011) Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces antigen-specific tolerance in mice with low genotoxic risk. *Hepatology* **53**(5)**:** 1696-1707.

McCarty, DM (2008) Self-complementary AAV vectors; advances and applications. *Mol Ther* **16**(10)**:** 1648-1656.

McCarty, DM, Fu, H, Monahan, PE, Toulson, CE, Naik, P, Samulski, RJ (2003) Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo. *Gene Ther* **10**(26)**:** 2112-2118.

McCarty, DM, Monahan, PE, Samulski, RJ (2001) Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene Ther* **8**(16)**:** 1248-1254.

McCormack, MP, Rabbitts, TH (2004) Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **350**(9)**:** 913-922.

McDougal, JS, Nicholson, JK, Cross, GD, Cort, SP, Kennedy, MS, Mawle, AC (1986) Binding of the human retrovirus HTLV-III/LAV/ARV/HIV to the CD4 (T4) molecule: conformation dependence, epitope mapping, antibody inhibition, and potential for idiotypic mimicry. *J Immunol* **137**(9)**:** 2937-2944.

McLachlan, G, Davidson, H, Holder, E, Davies, LA, Pringle, IA, Sumner-Jones, SG, Baker, A, Tennant, P, Gordon, C, Vrettou, C, Blundell, R, Hyndman, L, Stevenson, B, Wilson, A, Doherty, A, Shaw, DJ, Coles, RL, Painter, H, Cheng, SH, Scheule, RK, Davies, JC, Innes, JA, Hyde, SC, Griesenbach, U, Alton, EW, Boyd, AC, Porteous, DJ, Gill, DR, Collie, DD (2011) Pre-clinical evaluation of three non-viral gene transfer agents for cystic fibrosis after aerosol delivery to the ovine lung. *Gene Ther*.

Mendell, JR, Campbell, K, Rodino-Klapac, L, Sahenk, Z, Shilling, C, Lewis, S, Bowles, D, Gray, S, Li, C, Galloway, G, Malik, V, Coley, B, Clark, KR, Li, J, Xiao, X, Samulski, J, McPhee, SW, Samulski, RJ, Walker, CM (2010) Dystrophin immunity in Duchenne's muscular dystrophy. *N Engl J Med* **363**(15)**:** 1429-1437.

Mendell, JR, Rodino-Klapac, LR, Rosales-Quintero, X, Kota, J, Coley, BD, Galloway, G, Craenen, JM, Lewis, S, Malik, V, Shilling, C, Byrne, BJ, Conlon, T, Campbell, KJ, Bremer, WG, Viollet, L, Walker, CM, Sahenk, Z, Clark, KR (2009) Limb-girdle muscular dystrophy type 2D gene therapy restores alpha-sarcoglycan and associated proteins. *Ann Neurol* **66**(3)**:** 290-297.

Miao, CH, Ohashi, K, Patijn, GA, Meuse, L, Ye, X, Thompson, AR, Kay, MA (2000) Inclusion of the hepatic locus control region, an intron, and untranslated region increases and stabilizes hepatic factor IX gene expression in vivo but not in vitro. *Mol Ther* **1**(6)**:** 522-532.

Mielke, C, Kohwi, Y, Kohwi-Shigematsu, T, Bode, J (1990) Hierarchical binding of DNA fragments derived from scaffold-attached regions: correlation of properties in vitro and function in vivo. *Biochemistry* **29**(32)**:** 7475-7485.

Miller, AM, Dean, DA (2009) Tissue-specific and transcription factor-mediated nuclear entry of DNA. *Adv Drug Deliv Rev* **61**(7-8)**:** 603-613.

Miller, MD, Farnet, CM, Bushman, FD (1997) Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* **71**(7)**:** 5382-5390.

Mitrophanous, K, Yoon, S, Rohll, J, Patil, D, Wilkes, F, Kim, V, Kingsman, S, Kingsman, A, Mazarakis, N (1999) Stable gene transfer to the nervous system using a non-primate lentiviral vector. *Gene Ther* **6**(11)**:** 1808-1818.

Miyoshi, H, Blomer, U, Takahashi, M, Gage, FH, Verma, IM (1998) Development of a self-inactivating lentivirus vector. *J Virol* **72**(10)**:** 8150-8157.

Muller-Sieburg, CE, Cho, RH, Thoman, M, Adkins, B, Sieburg, HB (2002) Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood* **100**(4)**:** 1302-1309.

Nakai, H, Storm, TA, Kay, MA (2000a) Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors. *Nat Biotechnol* **18**(5)**:** 527-532.

Nakai, H, Storm, TA, Kay, MA (2000b) Recruitment of single-stranded recombinant adeno-associated virus vector genomes and intermolecular recombination are responsible for stable transduction of liver in vivo. *J Virol* **74**(20)**:** 9451-9463.

Naldini, L, Blomer, U, Gallay, P, Ory, D, Mulligan, R, Gage, FH, Verma, IM, Trono, D (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**(5259)**:** 263-267.

Nathwani, AC, Tuddenham, EG, Rangarajan, S, Rosales, C, McIntosh, J, Linch, DC, Chowdary, P, Riddell, A, Pie, AJ, Harrington, C, O'Beirne, J, Smith, K, Pasi, J, Glader, B, Rustagi, P, Ng, CY, Kay, MA, Zhou, J, Spence, Y, Morton, CL, Allay, J, Coleman, J, Sleep, S, Cunningham, JM, Srivastava, D, Basner-Tschakarjan, E, Mingozzi, F, High, KA, Gray, JT, Reiss, UM, Nienhuis, AW, Davidoff, AM (2011) Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med* **365**(25)**:** 2357-2365.

Niidome, T, Huang, L (2002) Gene therapy progress and prospects: nonviral vectors. *Gene Ther* **9**(24)**:** 1647-1652.

O'Neill, JP, Hsie, AW (1978) Growth arrest of Chinese hamster ovary cells in serum-free medium: changes in cell morphology and the intracellular and prostaglandin-induced levels of cyclic AMP. *J Cyclic Nucleotide Res* **4**(3)**:** 169-174.

Omasa, T, Cao, Y, Park, JY, Takagi, Y, Kimura, S, Yano, H, Honda, K, Asakawa, S, Shimizu, N, Ohtake, H (2009) Bacterial artificial chromosome library for genome-wide analysis of Chinese hamster ovary cells. *Biotechnol Bioeng* **104**(5)**:** 986-994.

Opie, SR, Warrington, KH, Jr., Agbandje-McKenna, M, Zolotukhin, S, Muzyczka, N (2003) Identification of amino acid residues in the capsid proteins of adeno-associated

virus type 2 that contribute to heparan sulfate proteoglycan binding. *J Virol* **77**(12)**:** 6995-7006.

Osborn, MJ, McElmurry, RT, Lees, CJ, DeFeo, AP, Chen, ZY, Kay, MA, Naldini, L, Freeman, G, Tolar, J, Blazar, BR (2011) Minicircle DNA-based gene therapy coupled with immune modulation permits long-term expression of alpha-L-iduronidase in mice with mucopolysaccharidosis type I. *Mol Ther* **19**(3)**:** 450-460.

Ottaviani, D, Lever, E, Takousis, P, Sheer, D (2008) Anchoring the genome. *Genome Biol* **9**(1)**:** 201.

Papapetrou, EP, Ziros, PG, Micheva, ID, Zoumbos, NC, Athanassiadou, A (2006) Gene transfer into human hematopoietic progenitor cells with an episomal vector carrying an S/MAR element. *Gene Ther* **13**(1)**:** 40-51.

Park, F, Kay, MA (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**(3)**:** 164-173.

Pauwels, K, Gijsbers, R, Toelen, J, Schambach, A, Willard-Gallo, K, Verheust, C, Debyser, Z, Herman, P (2009) State-of-the-art lentiviral vectors for research use: risk assessment and biosafety recommendations. *Current gene therapy* **9**(6)**:** 459-474.

Perez-Luz, S, Diaz-Nido, J (2010) Prospects for the use of artificial chromosomes and minichromosome-like episomes in gene therapy. *J Biomed Biotechnol* **2010**.

Piechaczek, C, Fetzer, C, Baiker, A, Bode, J, Lipps, HJ (1999) A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. *Nucleic Acids Res* **27**(2)**:** 426-428.

Price, GB, Allarakhia, M, Cossons, N, Nielsen, T, Diaz-Perez, M, Friedlander, P, Tao, L, Zannis-Hadjopoulos, M (2003) Identification of a cis-element that determines autonomous DNA replication in eukaryotic cells. *J Biol Chem* **278**(22)**:** 19649-19659.

Puck, TT, Cieciura, SJ, Robinson, A (1958) Genetics of somatic mammalian cells. III. Long-term cultivation of euploid cells from human and animal subjects. *J Exp Med* **108**(6)**:** 945-956.

Qing, K, Mah, C, Hansen, J, Zhou, S, Dwarki, V, Srivastava, A (1999) Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nat Med* **5**(1)**:** 71-77.

Ramezani, A, Hawley, TS, Hawley, RG (2003) Performance- and safety-enhanced lentiviral vectors containing the human interferon-beta scaffold attachment region and the chicken beta-globin insulator. *Blood* **101**(12)**:** 4717-4724.

Rangasamy, D (2010) An S/MAR-based L1 retrotransposition cassette mediates sustained levels of insertional mutagenesis without suffering from epigenetic silencing of DNA methylation. *Epigenetics* **5**(7)**:** 601-611.

Ratner, L, Fisher, A, Jagodzinski, LL, Mitsuya, H, Liou, RS, Gallo, RC, Wong-Staal, F (1987) Complete nucleotide sequences of functional clones of the AIDS virus. *AIDS Res Hum Retroviruses* **3**(1)**:** 57-69.

Recchia, A, Mavilio, F (2011) Site-Specific Integration by the Adeno-associated Virus Rep Protein. *Curr Gene Ther*.

Recchia, A, Perani, L, Sartori, D, Olgiati, C, Mavilio, F (2004) Site-specific integration of functional transgenes into the human genome by adeno/AAV hybrid vectors. *Mol Ther* **10**(4)**:** 660-670.

Renz, A, Fackelmayer, FO (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**(5)**:** 843-849.

Riu, E, Chen, ZY, Xu, H, He, CY, Kay, MA (2007) Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo. *Mol Ther* **15**(7)**:** 1348-1355.

Romig, H, Fackelmayer, FO, Renz, A, Ramsperger, U, Richter, A (1992) Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *EMBO J* **11**(9)**:** 3431-3440.

Rose, JA, Koczot, F (1972) Adenovirus-associated virus multiplication. VII. Helper requirement for viral deoxyribonucleic acid and ribonucleic acid synthesis. *J Virol* **10**(1)**:** 1-8.

Rupprecht, S, Hagedorn, C, Seruggia, D, Magnusson, T, Wagner, E, Ogris, M, Lipps, HJ (2010) Controlled removal of a nonviral episomal vector from transfected cells. *Gene* **466**(1-2)**:** 36-42.

Rupprecht, S, Lipps, HJ (2009) Cell cycle dependent histone dynamics of an episomal non-viral vector. *Gene* **439**(1-2)**:** 95-101.

Sanz, L, Compte, M, Guijarro-Munoz, I, Alvarez-Vallina, L (2011) Non-hematopoietic stem cells as factories for in vivo therapeutic protein production. *Gene Ther*.

Schaarschmidt, D, Baltin, J, Stehle, IM, Lipps, HJ, Knippers, R (2004) An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO J* **23**(1)**:** 191-201.

Schambach, A, Baum, C (2008) Clinical application of lentiviral vectors - concepts and practice. *Curr Gene Ther* **8**(6)**:** 474-482.

Scherer, WF, Syverton, JT, Gey, GO (1953) Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med* **97**(5)**:** 695-710.

Schleef, M, Blaesen, M, Schmeer, M, Baier, R, Marie, C, Dickson, G, Scherman, D (2010) Production of non viral DNA vectors. *Curr Gene Ther* **10**(6)**:** 487-507.

Schmidt, M, Schwarzwaelder, K, Bartholomae, C, Zaoui, K, Ball, C, Pilz, I, Braun, S, Glimm, H, von Kalle, C (2007) High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat Methods* **4**(12)**:** 1051-1057.

Schnepp, BC, Clark, KR, Klemanski, DL, Pacak, CA, Johnson, PR (2003) Genetic fate of recombinant adeno-associated virus vector genomes in muscle. *J Virol* **77**(6)**:** 3495-3504.

Schroeder, T (2010) Hematopoietic stem cell heterogeneity: subtypes, not unpredictable behavior. *Cell Stem Cell* **6**(3)**:** 203-207.

Schuettrumpf, J, Liu, JH, Couto, LB, Addya, K, Leonard, DG, Zhen, Z, Sommer, J, Arruda, VR (2006) Inadvertent germline transmission of AAV2 vector: findings in a rabbit model correlate with those in a human clinical trial. *Molecular therapy : the journal of the American Society of Gene Therapy* **13**(6)**:** 1064-1073.

Schulz, TF, Cordes, S (2009) Is the Epstein-Barr virus EBNA-1 protein an oncogen? *Proc Natl Acad Sci U S A* **106**(7)**:** 2091-2092.

Schuttrumpf, J, Milanov, P, Abriss, D, Roth, S, Tonn, T, Seifried, E (2011) Transgene loss and changes in the promoter methylation status as determinants for expression duration in nonviral gene transfer for factor IX. *Hum Gene Ther* **22**(1)**:** 101-106.

Sgourou, A, Routledge, S, Spathas, D, Athanassiadou, A, Antoniou, MN (2009) Physiological levels of HBB transgene expression from S/MAR element-based replicating episomal vectors. *J Biotechnol* **143**(2)**:** 85-94.

Simonelli, F, Maguire, AM, Testa, F, Pierce, EA, Mingozzi, F, Bennicelli, JL, Rossi, S, Marshall, K, Banfi, S, Surace, EM, Sun, J, Redmond, TM, Zhu, X, Shindler, KS, Ying, GS, Ziviello, C, Acerra, C, Wright, JF, McDonnell, JW, High, KA, Bennett, J, Auricchio, A (2010) Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. *Mol Ther* **18**(3)**:** 643-650.

Sirven, A, Ravet, E, Charneau, P, Zennou, V, Coulombel, L, Guetard, D, Pflumio, F, Dubart-Kupperschmitt, A (2001) Enhanced transgene expression in cord blood CD34(+)-derived hematopoietic cells, including developing T cells and NOD/SCID mouse repopulating cells, following transduction with modified trip lentiviral vectors. *Mol Ther* **3**(4)**:** 438-448.

Smith, RH, Levy, JR, Kotin, RM (2009) A simplified baculovirus-AAV expression vector system coupled with one-step affinity purification yields high-titer rAAV stocks from insect cells. *Mol Ther* **17**(11)**:** 1888-1896.

Southern, EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**(3)**:** 503-517.

Spangrude, GJ, Heimfeld, S, Weissman, IL (1988) Purification and characterization of mouse hematopoietic stem cells. *Science* **241**(4861)**:** 58-62.

Stehle, IM, Scinteie, MF, Baiker, A, Jenke, AC, Lipps, HJ (2003) Exploiting a minimal system to study the epigenetic control of DNA replication: the interplay between transcription and replication. *Chromosome Res* **11**(5)**:** 413-421.

Stevenson, M, Bukrinsky, M, Haggerty, S (1992) HIV-1 replication and potential targets for intervention. *AIDS Res Hum Retroviruses* **8**(2)**:** 107-117.

Straus, SE, Sebring, ED, Rose, JA (1976) Concatemers of alternating plus and minus strands are intermediates in adenovirus-associated virus DNA synthesis. *Proc Natl Acad Sci U S A* **73**(3)**:** 742-746.

Strayer, D, Branco, F, Zern, MA, Yam, P, Calarota, SA, Nichols, CN, Zaia, JA, Rossi, J, Li, H, Parashar, B, Ghosh, S, Chowdhury, JR (2002) Durability of transgene expression and vector integration: recombinant SV40-derived gene therapy vectors. *Mol Ther* **6**(2)**:** 227-237.

Studeny, M, Marini, FC, Dembinski, JL, Zompetta, C, Cabreira-Hansen, M, Bekele, BN, Champlin, RE, Andreeff, M (2004) Mesenchymal stem cells: potential precursors for tumor stroma and targeted-delivery vehicles for anticancer agents. *J Natl Cancer Inst* **96**(21)**:** 1593-1603.

Suda, T, Suda, K, Liu, D (2008) Computer-assisted hydrodynamic gene delivery. *Mol Ther* **16**(6)**:** 1098-1104.

Summerford, C, Bartlett, JS, Samulski, RJ (1999) AlphaVbeta5 integrin: a co-receptor for adeno-associated virus type 2 infection. *Nat Med* **5**(1)**:** 78-82.

Temin, HM, Mizutani, S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**(5252)**:** 1211-1213.

Tjio, JH, Puck, TT (1958) Genetics of somatic mammalian cells. II. Chromosomal constitution of cells in tissue culture. *J Exp Med* **108**(2)**:** 259-268.

Tobey, RA, Ley, KD (1970) Regulation of initiation of DNA synthesis in Chinese hamster cells. I. Production of stable, reversible G1-arrested populations in suspension culture. *J Cell Biol* **46**(1)**:** 151-157.

Valle, D, Downing, SJ, Harris, SC, Phang, JM (1973) Proline biosynthesis: multiple defects in Chinese hamster ovary cells. *Biochem Biophys Res Commun* **53**(4)**:** 1130-1136.

Van Craenenbroeck, K, Vanhoenacker, P, Haegeman, G (2000) Episomal vectors for gene expression in mammalian cells. *Eur J Biochem* **267**(18)**:** 5665-5678.

Vargas, J, Jr., Gusella, GL, Najfeld, V, Klotman, ME, Cara, A (2004) Novel integrase-defective lentiviral episomal vectors for gene transfer. *Hum Gene Ther* **15**(4)**:** 361-372.

Vogt, VM, Eisenman, R, Diggelmann, H (1975) Generation of avian myeloblastosis virus structural proteins by proteolytic cleavage of a precursor polypeptide. *J Mol Biol* **96**(3)**:** 471-493.

Wang, Z, Troilo, PJ, Wang, X, Griffiths, TG, Pacchione, SJ, Barnum, AB, Harper, LB, Pauley, CJ, Niu, Z, Denisova, L, Follmer, TT, Rizzuto, G, Ciliberto, G, Fattori, E, Monica, NL, Manam, S, Ledwith, BJ (2004) Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Ther* **11**(8)**:** 711-721.

Wanisch, K, Yanez-Munoz, RJ (2009) Integration-deficient lentiviral vectors: a slow coming of age. *Mol Ther* **17**(8)**:** 1316-1332.

Weindler, FW, Heilbronn, R (1991) A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. *J Virol* **65**(5)**:** 2476-2483.

Werner, M, Kraunus, J, Baum, C, Brocker, T (2004) B-cell-specific transgene expression using a self-inactivating retroviral vector with human CD19 promoter and viral post-transcriptional regulatory element. *Gene Ther* **11**(12)**:** 992-1000.

Wheatley, DN, Scott, L, Lamb, J, Smith, S (2000) Single amino acid (arginine) restriction: growth and death of cultured HeLa and human diploid fibroblasts. *Cell Physiol Biochem* **10**(1-2)**:** 37-55.

Wong, SP, Argyros, O, Coutelle, C, Harbottle, RP (2011) Non-viral S/MAR vectors replicate episomally in vivo when provided with a selective advantage. *Gene Ther* **18**(1)**:** 82-87.

Woods, NB, Bottero, V, Schmidt, M, von Kalle, C, Verma, IM (2006) Gene therapy: therapeutic gene causing lymphoma. *Nature* **440**(7088)**:** 1123.

Wu, P, Xiao, W, Conlon, T, Hughes, J, Agbandje-McKenna, M, Ferkol, T, Flotte, T, Muzyczka, N (2000) Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. *J Virol* **74**(18)**:** 8635-8647.

Wu, Z, Asokan, A, Samulski, RJ (2006) Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol Ther* **14**(3)**:** 316-327.

Wu, Z, Yang, H, Colosi, P Effect of genome size on AAV vector packaging. *Mol Ther* **18**(1)**:** 80-86.

Wurm, FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* **22**(11)**:** 1393-1398.

Wurm, FM, Hacker, D (2011) First CHO genome. *Nat Biotechnol* **29**(8)**:** 718-720.

Xu, X, Nagarajan, H, Lewis, NE, Pan, S, Cai, Z, Liu, X, Chen, W, Xie, M, Wang, W, Hammond, S, Andersen, MR, Neff, N, Passarelli, B, Koh, W, Fan, HC, Wang, J, Gui, Y, Lee, KH, Betenbaugh, MJ, Quake, SR, Famili, I, Palsson, BO (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol* **29**(8)**:** 735-741.

Yanez-Munoz, RJ, Balaggan, KS, MacNeil, A, Howe, SJ, Schmidt, M, Smith, AJ, Buch, P, MacLaren, RE, Anderson, PN, Barker, SE, Duran, Y, Bartholomae, C, von

Kalle, C, Heckenlively, JR, Kinnon, C, Ali, RR, Thrasher, AJ (2006) Effective gene therapy with nonintegrating lentiviral vectors. *Nat Med* **12**(3)**:** 348-353.

Yang, GS, Schmidt, M, Yan, Z, Lindbloom, JD, Harding, TC, Donahue, BA, Engelhardt, JF, Kotin, R, Davidson, BL (2002) Virus-mediated transduction of murine retina with adeno-associated virus: effects of viral capsid and genome size. *J Virol* **76**(15)**:** 7651-7660.

Zabner, J, Fasbender, AJ, Moninger, T, Poellinger, KA, Welsh, MJ (1995) Cellular and molecular barriers to gene transfer by a cationic lipid. *J Biol Chem* **270**(32)**:** 18997-19007.

Zufferey, R, Donello, JE, Trono, D, Hope, TJ (1999) Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J Virol* **73**(4)**:** 2886-2892.

Zufferey, R, Dull, T, Mandel, RJ, Bukovsky, A, Quiroz, D, Naldini, L, Trono, D (1998) Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J Virol* **72**(12)**:** 9873-9880.

Zufferey, R, Nagy, D, Mandel, RJ, Naldini, L, Trono, D (1997) Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol* **15**(9)**:** 871-875.