Investigating the mechanism of establishment of a prototype episomal gene therapy vector.

A thesis submitted to The University of Manchester for the degree of PhD

In the Faculty of Life Sciences

Year of Submission 2007

Rianne Elizabeth Stacey

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

Table of Contents

Chapter Two: Materials and Methods

2.1 Molecular Biology Techniques

2.2 Cell Culture Techniques

List of Figures

List of Tables

Abstract

There is much interest at present in the rational design of gene therapy vectors. This is hardly surprising considering the potential benefits of alleviation and even eradication of many genetically inherited diseases. There has been limited success to date with many diverse gene therapy systems, ranging from retroviral based ones to artificial chromosomes. Problems that have been encountered along the way include the efficient delivery of the gene to the cell and sustained long-term expression of the transgene once there. An ideal gene therapy vector would have a large cloning capacity to allow the transfer of not only the gene but its associated regulatory sequences, be readily transferred into dividing and non-dividing cells and mediate the long-term stable expression of the transgene without the use of bacterially or virally encoded factors.

A novel gene therapy vector has been developed that is maintained episomally for many generations in various mammalian cell types (Piechazcek *et al.* 1999, Jenke *et al.* 2004a, Papapetrou *et al.* 2006). This achieved without the use of virally encoded factors and occurs via an interaction between a scaffold/matrix attachment region on the vector and the host cell nuclear matrix. This association mediates efficient mitotic segregation and confers long-term stable gene expression in established cell lines. Whilst much work has focussed on the mechanisms involved in maintaining the established state, less interest has been directed towards the establishment phase. The switch from a transient state to a stably maintained state is incredibly inefficient and the processes involved are poorly understood. In order to achieve the goal of the ideal gene therapy vector, a fuller understanding of the factors affecting the establishment as well as the maintenance phase will be needed.

This study began by investigating the factors that might influence the efficiency of the establishment phase such as the timing of gene transfer with respect to the cell cycle phase and the presence of selective pressure. It was found that the efficiency of establishment was not affected by any of the changes applied in this study. It was observed that drug selection was not an absolute requirement for the establishment phase, though it may be necessary for high level transgene expression. A number of pEPI-1 derivatives were constructed; it was hoped that integration of a *lac* operator array into the vector would allow the identification of individual episomes in live cells via binding of a Lac repressor-EGFP fusion protein. However, the sensitivity was not sufficient for detection of the array in stable clones. Immunostaining against the GFP component of the Lac repressor fusion allowed the direct visualisation of episomes in fixed cells and provided valuable information concerning the distribution of episomes at mitosis. Finally, live cells were tracked to investigate the dynamic properties of the transcription cassette and it was shown that some clones have large variations in transgene expression levels. All these observations will aid in the development of subsequent generations of pEPI-1 or other such vectors.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree of qualification of this or any other university of other institute of learning.

Copyright Statement

i) Copyright in text of this thesis rests with the author. Copies (by any process) either in full, or of extracts, may be made **only** in accordance with instructions given by the author and lodged in the John Rylands University Library of Manchester. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the author.

ii) The ownership of any intellectual property right which may be described in this thesis is vested in The University of Manchester, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement.

iii) Further information on the conditions under which disclosures and exploitation may take place is available from the Vice-President and Dean of the Faculty of Life Sciences.

Acknowledgements

I would like to thank my supervisor Dr. Dean Jackson and my advisor Dr. Anne-Marie Buckle for support and advice throughout the highs and lows of the PhD program. Thanks must also go to all the members of my group for all their help and support, particularly to Dr. Chi Tang, Dr. Apolina Maya-Mendoza, Dr. Kang Zheng and Mr. Richard Firth. A special thank you must go to Miss Rachel Mottram for all her expertise with the FACS machine.

Finally I would like to recognise and thank my family for all their support, love and belief, especially my husband, Stephen Rice.

List of Abbreviations

ATP	Adenosine triphosphate
AU	Arbitrary unit (of fluorescence)
BSA	Bovine serum albumin
BUR	Base unpairing region
ChIP	Chromatin immunoprecipitation
CHO	Chinese hamster ovary
CLSM	Confocal laser scanning microscope
CT	Chromosome territory
DAPI	4', 6-diamidino-2-phenylindole
DHFR	Dihydrofolate reductase
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulphoxide
dNTP	Deoxynucleotide triphosphate
DsRed	Discosoma sp. red fluorescent protein
EBNA-1	EBV binding nuclear antigen-1
EBV	Epstein-Barr virus
EGFP	Enhanced green fluorescent protein
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FISH	Fluorescence in situ hybridisation
HAC	Human artificial chromosome
HAT	Histone acetyltransferase
HBSS	Hank's balanced salt solution
HC	Human chromosome
HDAC	Histone deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
HMT	Histone methyltransferase
HP1	Heterochromatin protein 1
HS	(Dnase 1) Hypersensitive site
HSC	Haematopoietic stem cell
IC	Interchromatin domain

Kb	Kilobase pairs
*lac*O	*lac* operator
LacI	Lac repressor
LB	Luria-Bertani
LCR	Locus control region
Lys	Lysine
MCS	Multiple cloning site
MEL	Murine Erythroleukemia
pCMV	Cytomegalovirus promoter
NLS	Nuclear localisation signal
NPC	Nuclear pore complex
OD	Optical density
PBS	Phosphate buffered saline
PD	Population doubling
PCR	Polymerase chain reaction
PI	Propidium iodide
RCF	Relative centrifugal force
rpm	Revolutions per minute
SAF-A	Scaffold attachment factor A
SATB1	Special AT-rich sequence binding 1
S/MAR	Scaffold/matrix attachment region
SV40	Simian virus 40
TBE	Tris-Borate-EDTA
TI	Transcriptional interference
YAC	Yeast artificial chromosome

Chapter One: Introduction

1.1 Gene Therapy- Overview

The main purpose of gene therapy is to treat, cure or prevent a disease by the introduction of a therapeutic gene and its regulatory sequence, or the alteration in expression profile of endogenous genes. The ectopic gene expression can be used to complement the disease-causing loss-of-function mutations characteristic of inherited genetic defects. There is much excitement and interest in developing strategies capable of executing this central aim, but to date, successful clinical applications remain depressingly few.

Jesse Gelsinger was the first patient to die as a direct result of a gene therapy treatment. Aged just 18 years old, Jesse was taking part in a Phase I gene therapy trial that aimed to correct a partial deficiency in a metabolic enzyme (ornithine transcarbamylase). Patients on the trial were given intra-hepatic injections of a second-generation recombinant adenoviral vector containing a functional copy of the defective gene. Within hours of administration, he began to experience severe complications and died two days later. Other patients on the study received similar doses to no ill effect and there is no clear answer to why such a reaction occurred. Celebrations were relatively short lived regarding another gene therapy trial concerned with treating children with X-linked severe combined immunodeficiency disease by *ex vivo* retrovirally mediated gene transfer. The disease is characterised by an early block in T and natural killer cell differentiation due to mutations in the gene encoding the γc cytokine receptor subunit common to several interleukin receptors. Haematopoietic stem cells (HSC) were transduced with a recombinant mouse leukaemia viral vector containing the γc receptor gene. Initial reports were extremely promising (Cavazzana-Calvo *et al.* 2000), showing sustained correction of the disease in some patients up to 3.6 years after gene therapy (Hacein-Bey-Abina S. *et al.* 2002). However, there may have been problems with insertional mutagenesis in one patient that is believed to have led to acute lymphocytic leukaemia (Hacein-Bey-Abina *et al.* 2003). Despite these tragedies, the potential benefits of gene

therapy mean that there is still a huge appetite for the field and many new and exciting avenues are opening up all the time.

The process of gene therapy relies on exploitation of the host transcriptional machinery for the expression of the exogenous DNA. To best design systems that provide the correct dose of the required protein at the right time and in the target cells, there requires a thorough understanding of the mechanisms underlying the transcriptional control of a gene.

1.2 Gene Expression and Nuclear Organisation

It is more than a century since chromosomes were first identified as the basic units of heredity and fifty years since Watson and Crick elucidated the structure of DNA. The ambitious human genome project and other such studies revealed that humans have approximately 35,000 genes, roughly the same amount as other mammals and only about twice as many as the relatively primitive fruit fly (Berezney, 2002). It is therefore evident that not only the genes themselves, but also their co-ordinated expression, must be responsible for the vast differences between species. Our understanding of how the genome is constructed to allow precisely orchestrated developmental and tissue-specific gene expression and to respond to a constantly changing environment has revealed many more levels of control and regulation than were ever initially imagined.

1.2.1 Chromatin Organisation

Nucleosomes represent the most basic level of packaging and are formed by wrapping approximately 146 bp of the DNA double helix (1.65 turns) around a core histone octamer to form a 10 nm fibre (Luger *et al.* 1997). Reversible modifications of the histone N-terminal tails form a pattern or code that influences the accessibility of regulatory sequences to their binding proteins and also the affinity with which these proteins bind to their recognition motifs (Strahl & Allis, 2000). The nucleosome array is thought to coil up to form a 30nm fibre with the binding and depletion of histone H1 promoting interconversions between the 10 and 30 nm structures (Park & De Boni, 1999). The 30 nm fibres

form loops attached to the nuclear matrix at their bases by scaffold/matrix attachment regions (S/MARs) found in the non-coding part of the genome. There are estimated to be 100,000 nuclear matrix attachment sites in the human genome (Liebich *et al.* 2002) that organise the chromosomes into a series of loops of between 20 and 200 kb (Park & De Boni, 1999). It is not entirely clear what higher levels of packaging actually exist in the interphase nucleus beyond loops; it has been suggested that the loops may coil up to form larger, more compact loops. Chromatin is generally found in one of two overall states heterochromatin and euchromatin; heterochromatin is extremely compact, tightly folded chromatin and is associated with repressed, inactive genes; euchromatin is typically more open and accessible to the transcription machinery and consists of active genes.

1.2.2 Control of Gene Expression

Chromatin has a dual role in the nucleus – to compact the chromosomal DNA and to direct dynamic and highly regulated gene expression. The compactness of the chromatin affects the expression potential of the genes contained within it and each additional level of chromatin packaging described above contributes to another prospective layer of control over the expression potential of the genome it organises. Transgenes integrated at most chromosomal loci are subject to gradual silencing due to encroachment of heterochromatin from the surrounding area (Geyer, 1997) and suggests that gene expression is regulated against a default background of repressive chromatin. For genes to become, and remain, active they must recruit factors to promoter and enhancer elements to establish a chromatin structure permissive for transcription. This dynamic relationship between activation by enhancers and chromatin-mediated repression ensures a considerable difference between the induced and repressed level of gene expression allowing tighter regulation of the genome (Udvardy, 1999).

1.2.3 Chromatin Structure

There are two different classes of enzyme that interact with chromatin to alter its structure i) chromatin remodelling enzymes that act by altering nucleosome positioning and ii) enzymes that modify the histone tails by adding or removing epigenetic marks (Jenuwein & Allis, 2001). Chromatin remodelling enzymes, such as RSC, ISWI and NuRF, act by changing the position of nucleosomes with respect the DNA sequence thus modulating the ability of regulatory sequences to be bound by transcriptional activator and repressor proteins. Various enzymes catalyse the post-translational modification of histone N-terminal tails. For example, histone acetyl-transferases (HATs) and histone deacetylases (HDACs) work antagonistically by adding and removing acetyl groups respectively at specific residues on the histone N-terminal tails. Increased acetylation of specific lysine (Lys) residues on histones H3/H4 is associated with active chromatin whereas histone hypoacetylation is a hallmark of inactive heterochromatin (Pikaart, Recillas-Targa & Felsenfeld, 1998). The addition or removal of these moieties is thought to influence the packaging of nucleosomes into higher order chromatin structures both by altering the charge on the histone tails and serving as part of the recognition motif for proteins that are able to modulate chromatin structure (Cheung, Allis & Sassone-Corsi, 2000). For example, methylation of H3-Lys 9 induces binding of heterochromatin protein 1 (HP1) proteins that once bound are thought to self-dimerise generating cross-linkers to stabilise a rigid higher-order chromatin structure and generate heterochromatin (Bannister *et al.* 2001). Recent work *in vitro* has provided real evidence to support the concept of histone tail charge influencing the degree of chromatin packing (Shogren-Knaak *et al.* 2006). Alteration of the physicochemical properties by acetylation of histone H4 lysine 16 can inhibit formation of compact higher-order fibres. This more open chromatin would then be more accessible to factors that recognise regulatory sequences in the DNA inducing more modifications, perhaps resulting in the extreme case of nucleosome-free regions, identified as nuclease hypersensitive sites. It has been generally held for quite some time that modification of different combinations of amino acids within the histone N-terminal tails forms a code that can be 'read' by other proteins; where acetylation of H4-Lys5, Lys12 and Lys16 are involved in the initiation of chromatin opening and acetylation of H4-

Lys8 appears to be important for the maintenance of open chromatin at the β-globin locus (Cheung *et al.* 2000). The situation now appears to be much more complex and the classical histone code may be better thought of as a 'language' where it is the combination of modifications (words) that determines whether the domain is active or inactive rather than individual modifications having specific activating or silencing properties. This revision of thinking came about by numerous seemingly conflicting observations where the same modification appeared to be activating in some studies and inactivating in others, this is excellently reviewed by Berger (2007). It seems that the effect of a chromatin modification is dependent on those around it as they all contribute to the recruitment of regulatory factors and in addition to interacting with the chromatin they also interact with each other. Using this reasoning, the chromatin modifications within a domain would be the result of gradual evolution where successive rounds of transcription and replication may influence the nature of the modifications and factors bound to them, ultimately influencing the activity of the gene region depending on whether open or compact chromatin has been formed.

1.2.4 Chromatin Domains

Transgenes integrated randomly at various chromosomal loci exhibit differences in expression levels. This is thought to be due the transgene adopting a chromatin structure and expression potential similar to its surrounding environment. The gradual spread of heterochromatin into the majority of integrated genes reflects the fact that the bulk of the genome in differentiated cells is not active. The targeted generation of regions of specific histone modifications appears to organise the chromosomes into functionally independent domains of gene activity where all the genes within a domain share the same transcriptional potential.

This is illustrated by the integration of a β-galactosidase reporter gene fused to the HPRT promoter (HPRT is expressed in all tissues) into fertilised mouse eggs (Bonnerot *et al.* 1990). Levels of β-galactosidase expression can be monitored in the adult organism to indicate the transcriptional potential of the

site of integration. Each organism as a whole showed varying levels of β-galactosidase expression depending on the tissue examined. As the insert site was the same in every tissue within a single mouse, this suggests that the transgene was subject to the transcriptional control of the surrounding chromatin domain, the variation in expression levels resulting from the fact that the domain was only active in certain tissues. In some organisms studied, β-galactosidase expression appeared transiently, for example, in one organism it appeared early in the motor neurons of the spinal cord but disappeared rapidly after birth showing that within the normal developmental program the transcriptional potential of a domain may alter. The chromatin structure within a domain is kept independent of the surrounding genome by the presence of insulators or boundary elements at the ends of a domain. Boundary elements mark the limits of a domain and are thought to prevent the spread of heterochromatin into an active domain as well as promote an open chromatin structure within the domain (Geyer, 1997). The interaction of insulator bound proteins with the nuclear matrix or with each other creates higher order domains of chromatin that are functionally separate from the surrounding region to specify a precise pattern of gene expression (Gerasimova & Corces, 1998). The variable expression of a transgene stably integrated into a genome highlights the existence of these domains. The presence of endogenous enhancers and silencers cross-talking inappropriately with the transgene or the inability of the construct to prevent silencing by the encroachment of heterochromatin is responsible for this variability in expression and is due, in part, to a lack of insulators or boundary elements and so an inability of the gene to establish the appropriate chromatin structure (Kellum & Elgin, 1998). Flanking transgenes with insulators or boundary elements confers position independence, allows 'normal' regulation of the transgene and prevents gradual silencing by the spread of heterochromatin (Kalos & Fournier, 1995). This is probably due the formation of loops insulating the promoters in one loop from the regulatory influences of enhancers and silencers in a neighbouring loop and acting as a 'chain-terminator' for the co-operative loading of HP1 (Udvardy, 1999).

1.2.5 'Positions of Potential' – the β-globin locus

Any gene inserted into an active domain will be transcribed if the appropriate activating proteins are present, conversely, if this gene were inserted into an inactive domain then the environment would not be permissive for transcription even in the presence of transcriptional activators. The presence or absence of activator proteins combined with the generation of open or closed chromatin in any tissue gives rise to 3 distinct transcriptional states; closed, nuclease insensitive chromatin that is refractory to transcription even in the presence of activating factors; open, nuclease sensitive chromatin that has the potential for gene expression but remains inactive as no activating proteins are present; and an open active domain bound at regulatory sequences by enhancer proteins, etc that produce nuclease hypersensitive regions and facilitate a high level of regulated transcription. It follows that even though a domain may be classified as active, it does not necessarily mean that all the genes within the domain are being actively transcribed, rather they could be said to be 'poised' (Gasser, 2001).

The mammalian β-globin locus exemplifies the theory described above. The β-globin genes are expressed at high levels specifically in erythroid cells and expression is dependent on a functional LCR (locus control region) located 5' of the β-globin genes and composed of a set of nuclease hypersensitive sites (HSs) (Forrester *et al.* 1990).

Examination of non-erythroid cells reveals basal histone acetylation within the globin cluster suggesting that the genes are 'poised' in all cell types whether they are transcribed or not and that β-globin expression is probably due to erythroid-specific factors binding within the LCR. Studies have shown two distinct properties of the LCR i) formation of an active domain characterised by histone hyperacetylation and ii) activation of transcription (Bulger *et al.* 2002). Distinct peaks of histone hyperacetylation coincide with the transcribed genes and the regulatory sequences (Bulger *et al.* 2003). Constructs with deletions of the HS elements were produced to help elucidate the part each HS element plays in creating an active domain and initiating high level gene expression. The 5'HS4 element is the only DNase1 hypersensitive site in all tissue types, the hypersensitive sites downstream are erythroid-specific suggesting that 5'HS4

insulates the cluster against encroachment of the neighbouring heterochromatin, maintaining the chromatin in the region in a more open state (Chung, Bell & Felsenfeld, 1997). The formation of the erythroid-specific hypersensitive sites (5'- HS1, HS2 and HS3) is thought to occur as a result of erythroid-specific factors binding to the 5'HS4. These tissue-specific hypersensitive sites can then be bound by proteins which effect transcription of the β-globin cluster due to the strong enhancer (5'HS2) and chromatin opening activities (5'HS3) establishing an independent chromatin domain characterised by histone hyperacetylation across the region (McMorrow *et al.* 2000). Removal of the entire LCR results in a loss of high-level gene expression at the β-globin locus and is associated with an alteration in chromatin structure over the region (Martin, Fiering & Groudine, 1996). The region remains nuclease sensitive and global histone acetylation is unaffected, however, the hyperacetylation of histones H3/H4 at the β-globin promoter is lost (Schubeler, Groudine & Bender, 2001). ChIPs (chromatin immunoprecipitation) were performed at the locus of erythroid cells to probe the developmentally regulated patterns of histone acetylation and they revealed 3 distinct classes of acetylation: i) typically very low acetylation at all developmental stages – found in the condensed chromatin region ii) typically very high levels of acetylation at all developmental stages – found in the regulatory regions, particularly at the 5'HS4 element and iii) development stage dependent acetylation – low in the early stages of erythroid cell development when the β-globin genes are inactive and high in later stages when the locus is active (Litt *et al.* 2001).

1.2.6 Chromosome Territories

The application of fluorescence in situ hybridisation (FISH) using whole chromosome paints has shown that each chromosome occupies a discrete chromosome territory (CT) and is surrounded by the interchromatin compartment (IC) which contains the transcriptional machinery and splicing factors etc. (Cremer *et al.* 2000). Chromosomes are constrained within their territories by attachments to the nuclear matrix and other nuclear substructures such as the nucleoli and nuclear envelope. Disruption of the nucleoli resulted in an increase in mobility of nucleolar-associated foci (Chubb *et al.* 2002).

Chromatin architecture alters when cells progress through developmental stages from progenitor cells to cycling cells to terminally differentiated cells, or when cells exit the cell cycle and become senescent or quiescent (Bridger *et al.* 2000). This could be partly due to a reduction in the number of contacts made between the genome and the nuclear matrix via binding to transcription factories. The distinct chromatin architecture generated in different cell types or at different developmental stages is likely to have evolved with the genome to produce a pattern of chromatin loops that facilitates specific and regulated expression of only those genes required by a particular cell. These chromatin loops might create not only a domain of active or inactive chromatin, but by tethering particular nuclear substructures might actually constrain genes to specific areas of the nucleus, such as an area high in HAT activity that would maintain the chromatin in that loop in an open conformation.

Our understanding of chromosome territories and how this is related to gene expression has been advanced by the use of FISH to demonstrate differences in nuclear position and chromosome morphology between human chromosomes (HC) 18 and HC19 (Croft *et al.* 1999). Chromosomes 18 and 19 are roughly similar in size; however, chromosome 18 is relatively gene poor with late replicating DNA compared with chromosome 19, which is gene rich and replicates early in S-phase. Chromosome 19 adopts a relatively internal nuclear position, is extensively associated with the nuclear matrix and has a moderately high level of histone H4 acetylation. This is in contrast to chromosome 18, which is located close to the lamina with less nuclear matrix attachments and little hyperacetylated H4. It has also been shown that HC19 adopts a much more internal nuclear position compared with HC18 and that those differences in the subnuclear localisations of HC18 and HC19 are established very early in the cell cycle (Bridger *et al.* 2000). This level of nuclear architecture is altered in quiescent or senescent cells; HC18 moves from the nuclear periphery to a more internal location in cells entering senescence. When cells were chemically stimulated to re-enter the cell cycle, HC18 moved away from the nuclear interior back to the nuclear periphery.

An examination of specific genes on HC2 showed that their localisation with respect to the domain periphery is independent of gene activity. Both the active and the inactive genes were preferentially positioned at the periphery of the

chromosome territory and the non-coding regions were found predominantly in the interior of the territory (Kurz *et al.* 1996). This positioning is logical as genes located close to the domain peripheries could easily contact the transcription machinery located in the IC. However, photobleaching studies indicate that the whole of the nucleus is accessible to small molecules (Houtsmuller & Vermeulen, 2001) and high-resolution microscopy has shown that chromosome territories are highly porous and ramified with numerous channels (Verschure *et al.* 1999). This indicates that there will be a much greater surface area of a CT in contact with the IC and that genes that appear to be embedded deep in a CT may actually still have access to the transcription machinery. It was found that HC19 occupied a significantly larger portion of the nuclear area than HC18 and that this was also dependent on transcriptional status (Bridger *et al.* 2000). When an irreversible inhibitor of RNA polymerases I and II was added to lymphoblasts, the chromosome territory of HC19, but not HC18, was seen to occupy a significantly smaller fraction of nuclear area than in untreated cells. It should be noted, however, that there was no change in the relative positions of the chromosomes, only in the area occupied by them. Comparison of the domains of the active and inactive human X chromosomes using FISH showed they possessed a similar volume but the active X chromosome territory had a more convoluted and larger surface area than its inactive counterpart (Belmont *et al.* 1999). These examples demonstrate that the spatial organisation and localisation of chromosomes is plastic and can be influenced by, or even influences, transcriptional status.

1.2.7 Nuclear Localisation

The movement of chromosome regions was visualised in living cells using real time fluorescence microscopy by detecting Lac-repressor-NLS (LacI) as a fluorescent fusion protein, LacI-EGFP, bound to a 10 kb repeat of the *lac* operator (*lac*O) array integrated at specific chromosomal loci (Robinett *et al.* 1996). There was little support for the directed movement of chromosome regions during the normal cell cycle suggesting instead that movements occurred randomly within a restricted region of the chromosome territory (Vazquez, Belmont & Sedat, 2001). Chromosomes were typically confined to

compact territories with each territory exhibiting obvious internal substructure. The direct visualisation of extended chromatin fibres showed they changed in length by less than 15% over several hours suggesting that the large-scale chromatin structure seen within a chromosome territory is relatively stable (Robinett *et al.* 1996). However, small local conformational changes in the fibres were detected and appear to have been brought about by the bending and coiling of existing structures. It could be hypothesised that these local changes in chromatin architecture were contributing to alterations in the expression of the genes present on the area studied.

There are numerous examples of a change in nuclear localisation serving as a prerequisite for modification of the underlying chromatin structure and therefore activation or inactivation of the locus. Heterochromatin can be spread in *trans* and so the recruitment of loci to heterochromatic regions can induce silencing. In T-lymphocytes, the heritable silencing of several developmentally important genes results in their recruitment to heterochromatin-containing nuclear domains. When naïve CD4 parent cells differentiate into either Th1 or Th2 helper cells the transcriptionally silent IL-4 (Th1) and IFNγ (Th2) genes were found associated with the centromeric heterochromatin, a configuration not seen in the parental cells (Baxter, Merkenschlager & Fisher, 2002). Ikaros, a protein found in non-lymphocyte precursor cells, silences lymphocyte-specific genes by recruiting them to centromeric heterochromatin (Cockell & Gasser, 1999). Lymphocytes do not have the Ikaros protein, do not show a colocalisation of the cell-specific genes with centromeric heterochromatin and so are able to actively transcribe these genes leading to differentiation into T, B and natural killer cells. There is also evidence that loci need to be relocated away from heterochromatin before they can adopt an open chromatin structure. The VP16 acidic activation domain when targeted to an engineered chromosome site activated transcription and induced a relocalisation of the locus from a predominantly peripheral to a more interior nuclear localisation (Tumbar & Belmont, 2001) and in human fibroblasts the MHC class II gene cluster is relocated away from the main body of chromosome 6 following gene activation by interferon-γ (Volpi *et al.* 2000). Studies on the β-globin locus have shown that relocation of the locus away from centromeric heterochromatin and

general histone H3/H4 acetylation are linked and occur even when the 5'HS2-5 elements of the LCR are deleted (Schubeler *et al.* 2000). The domain-wide H3/H4 acetylation might be a consequence of the relocation of the β-globin locus to a nuclear compartment rich in HAT activity. These examples show that nuclear localisation and an open or closed chromatin structure are interdependent with some instances of relocation acting as a prerequisite for chromatin activation or inactivation and other instances where activation of a locus drives the relocation. In actual fact there is probably a dynamic balance between the two; if a small change in localisation moves a locus to an area with a higher concentration of HATs then there might be some increase in histone acetylation and chromatin opening. A more open chromatin structure may support the maintenance of this new nuclear location and even promote further movements towards higher concentrations of HATs until the domain is basally acetylated and the chromatin open and 'poised'. The converse would also be true and would explain why transgenes are subject to gradual silencing; once a gene moves to an area of the nucleus with a lower HAT activity the locus will begin to become deacetylated. A less open chromatin structure would be less able to bind to factors involved in the maintenance of an open chromatin structure in an active domain and so it would be more likely to move to an area with even lower HAT activity and lose more histone acetylation. At each of these downward 'steps' there would be a possibility of an increase in histone acetylation and therefore a return to a more open chromatin, but the overall drive would be towards the generation of inert heterochromatin. The tethering of certain loci to the nuclear matrix in active domains of the nucleus may prevent this loss transcriptional activity by maintaining a nuclear address that is rich in HAT activity.

1.2.8 The Nuclear Matrix and S/MARs

The matrix and scaffold are residual structures remaining after high salt or LIS extractions, respectively (Craig *et al.* 1997). Chromatin released using 'physiological' buffers shows a residual skeleton upon which transcription and replication continue (Jackson, Yuan & Cook, 1988). The nuclear matrix is a fibogranular ribonucleoprotein network composed of the nuclear lamina and an

internal network of branched 10nm filaments (Nickerson, 2001). S/MARs define the bases of the DNA loops that are visible as a 'halo' around extracted nuclei and show an attachment pattern strikingly similar to that detected in living cells (Hart & Laemmli, 1998).

S/MARs are an incredibly evolutionarily conserved feature as S/MARs from one organism are able to bind components of the nuclear matrix from another organism (Cooper *et al.* 1997). Also, S/MARs from *Drosophila* or plants could functionally substitute for an endogenous S/MAR at the immunoglobulin μ locus (Hart & Laemmli, 1998).

S/MARs do not have a precise recognition sequence and binding proteins probably recognise their structural features; they are generally 70% AT-rich, composed of runs of As and Ts (A-tracts), generating non-B form DNA structure (Bode & Maass, 1988). Typically S/MAR DNA has a narrow minor groove, bends, and has a propensity to unwind, particularly in the base unpairing region (BUR). The unwinding property of the BUR within the S/MAR is invaluable for augmenting gene expression in stable transformations (Bode *et al.* 1992). Several BUR-recognising proteins have been identified to date and these include SAF-A (Romig *et al.* 1992; Kipp *et al.* 2000), SATB1 (Nakagomi *et al.* 1994), H1 and HMG-I/Y (Fukuda & Nishikawa, 2003). S/MARs may be involved in the condensation and decondensation of chromatin by the mutually exclusive binding of histone H1 and AT-hook proteins. The binding of H1 cooperatively stimulates loading of H1 onto adjacent nucleosomes promoting a condensed chromatin structure whereas AT-hook proteins such as HMGI/Y binding with a higher affinity to the S/MAR sequences would displace H1 and locally open the chromatin (Fukuda & Nishikawa, 2003). *In vitro* competition experiments show that HMG-I/Y binds selectively to the minor groove of A-tracts and displaces histone H1 to make S/MAR DNA accessible and open.

S/MARs augment transcription and there are a number of ways in which they might achieve this in addition to their chromatin opening activities. One way is by creating loops of chromatin where each loop, with S/MARs at its base, signifies a single domain of transcriptional potential i.e. within a single domain the chromatin is either 'open' and potentially transcribable or 'closed' and inert. The formation of a different set of loops in each cell type could help specify

unique transcriptional patterns by setting up the appropriate active domains around tissue-specific genes and by limiting the effects of enhancers to genes within a loop. The tethering of certain sequences to the nuclear matrix might also augment transcription by bringing genes closer to the transcriptional machinery as it has been shown that various transcription components are enriched at the nuclear matrix (Bode *et al.* 2000). Whatever their mode of action, they appear to require the presence of other regulatory elements to function correctly; S/MARs are not able to augment transcription of transgene when used with a construct without an enhancer (Fukuda & Nishikawa, 2003).

There appears to be two different functions of matrix attachments, a structural one that is mediated by constitutively expressed matrix proteins binding to the S/MARs and a regulatory one that anchors the S/MARs in a precise pattern of loops to a nuclear cage or network formed by tissue-specific proteins. The frequent colocalisation of S/MAR sequences with other regulatory elements reinforces the argument that they are involved in gene regulation (Phi-Van & Stratling, 1988). An example of the involvement S/MARs in the regulation of tissue-specific gene expression is in thymocytes where a set of S/MARs are bound by a thymocyte specific protein (at the BURs), SATB1, to form a precise network of loops that direct tissue-specific gene expression. SATB1 is expressed primarily in the T-cell lineage and is essential for proper thymus development. SATB1-null mice are small, have disproportionately small thymi and spleens, have multiple defects in T-cell development and die at 3 weeks (Alvarez *et al.* 2000). Examination of 589 genes revealed that at least 2% were derepressed at inappropriate stages of T-cell development in the thymus whereas gene expression patterns appeared normal in the liver, an organ not normally expressing SATB1. When FISH was used against SATB1 binding sites the signals were found at the bases of the loops in normal thymocytes and in the loops in SATB1-null cells providing compelling evidence that the chromatin architecture formed when SATB1 anchors sequences to the nuclear matrix directs thymocyte specific gene expression patterns (Cai, Han & Kohwi-Shigematsu, 2003). SATB1 when bound to the S/MAR targets CHRAC and ACF nucleosome remodelling complexes to a specific site within the IL2ra locus to modify nucleosome positioning across a region of some 7 kb. This chromatin remodelling activity is precisely linked to the proper regulation of IL2ra during

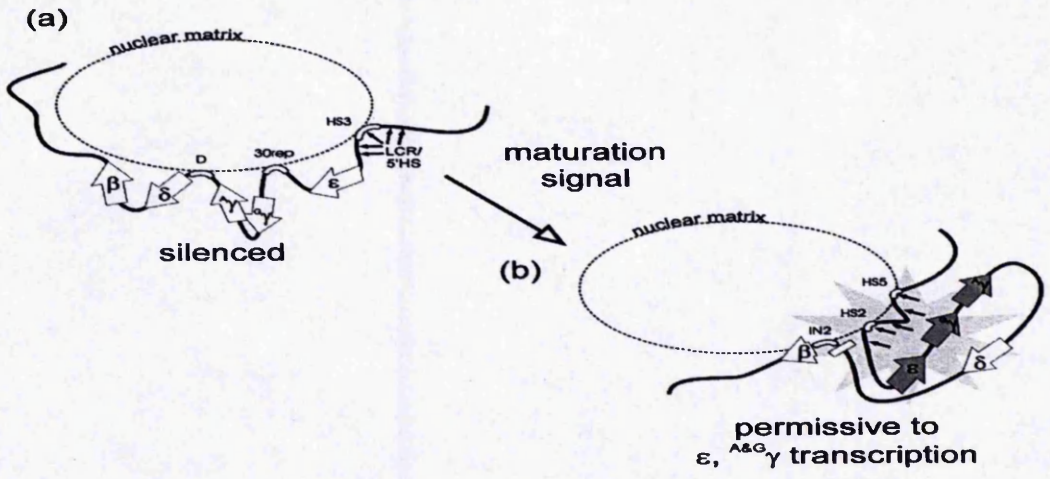


Figure 1.1. Involvement of S/MARs in the regulation of the β-globin locus. (a) Cells not expressing β-globin. Attachment of the 5'HS3 in the LCR is thought play a role in preventing globin gene expression. (b) Cells permissive for β-globin expression. As the cells develop they direct the remodelling of the whole domain and release of the previously tethered regions. The new set of matrix attachment regions facilitates β-globin gene expression. Taken from (Ostermeier *et al.* 2003).

thymocyte differentiation (Yasui *et al.* 2002). Investigations using quantitative real-time PCR has shown that nuclear matrix association dynamically mediates the looping of the β-globin locus to regulate transcription (Ostermeier *et al.* 2003). Figure 1.1 shows how the matrix attachments alter as cells remodel their chromatin domains to confer developmentally regulated gene expression patterns.

Heng *et al.* (2004) have shown that nuclear matrix associations of S/MARs are dynamic. Multiple-copy S/MARs were introduced in tandem arrays and it was shown that, even though they all contained identical sequences, not all S/MARs were matrix attached. Sometimes even a single-copy integration could be detected in the chromatin loop suggesting they were selected and used as nuclear matrix anchors in a discriminatory manner.

1.3 Gene Therapy and S/MARs

Current gene therapy vectors can be separated into two main classes, those that rely on integration into the host genome, such as retroviruses (reviewed in Walther & Stein, 2000), and those that function as extrachromosomal units, such as adenoviruses (reviewed in Ghosh, Gopinath & Ramesh, 2006) or

chromosome based ones (reviewed in Lipps *et al.* 2003). In general, the current generations of these systems do not work terribly well; the transgene is usually subject to fairly rapid silencing by the host cell as there is simply not enough regulatory information delivered with the gene. Chromosomes will have evolved so that they are context dependent and their expression program will be the result of numerous factors including precise chromatin architecture. The insertion of a gene into a random chromosomal location will not have its endogenous architecture and so will be without levels of control of gene expression, in addition, the insertion of the gene itself might disrupt the existing chromatin structure, potentially leading to the activation of oncogenes (Calos, 1996).

1.3.1 Human Artificial Chromosomes

The use of episomal vectors is appealing as it would bypass some of the problems associated with random integration into the genome. Human artificial chromosomes (HACs) can be formed *de novo* in human HT1080 cells following transfection of alpha satellite DNA arrays on bacterial-, yeast- or P1- artificial chromosomes and are generally circular, present at 1-2 copies per cell, are estimated to be 1-10 megabases in size and have a mitotic stability of ~98% (Monaco & Moralli). Immunofluorescence studies on HACs revealed the presence of regions of euchromatin (H3 dimethylated Lys4) and heterochromatin (H3 methylated Lys9) suggesting that independent domains were able to be formed across the construct ((Grimes & Monaco, 2005). Several studies have demonstrated the efficacy of HACs as vectors for the expression of transgenes. One example is the generation of HACs that contained multiple copies of the gene *guanosine triphosphate cyclohydrolase* were it was demonstrated that the transgenes on the HACs were responsive to interferon-γ induction (Ikeno *et al.* 2002). Their advantages are the capacity to carry large DNA segments and the formation of domains with different sets of chromatin modifications creating a much more 'natural' genomic environment; their disadvantages are the tendency to form multiple copies of (often

rearranged) input DNA (Grimes & Monaco, 2005) and issues with the delivery of such large DNA constructs to the target human cells.

1.3.2 Epstein-Barr Virus-Based Vectors

Another potential episomal expression system is relies on the properties of Epstein-Barr viruses (EBV); it is a human herpesvirus with a double-stranded DNA genome (~172 kb) (Mecsas & Sugden, 1987) that establishes itself latently as an extrachromosomal unit that replicates once per cell cycle. This controlled replication ensures vectors do not have an elevated mutation rate and do not readily rearrange (Sclimenti & Calos, 1998). Establishment of EBV as an episome requires just two components, the origin of replication, *oriP*, and the viral gene product EBV binding nuclear antigen-1 (EBNA-1) (Yates & Guan, 1991). Much work has been done in developing vectors based on these two components for use as gene therapy vectors such as the inclusion of inducible gene expression elements. An EBV vector carrying the gene defective in cystic fibrosis conferred gene expression for more than two months in transformed dividing cultured human respiratory epithelial cells (Lei *et al.* 1996). The Calos group constructed an EBV-based vector carrying the whole 19 kb genomic fragment containing with the gene (*serpina1*) encoding α_1-antitrypsin (Stoll *et al.* 2001). The vector underwent efficient extrachromosomal replication in dividing mammalian tissue culture cells and provided long-term expression of the protein (more than nine months). The advantage of this system is the ability to transfect whole genomic sequences (up to 100 kb) allowing the delivery of the gene and its associated regulatory sequences. The drawbacks to such systems, however, are the reliance on a virally encoded *trans*-acting factor (EBNA-1) which leads to cell transformation (Yates & Guan, 1991) and the fact that some EBV-based episomes show reduced mitotic stability, being lost at a rate of 4% per cell generation in the absence of antibiotic selection (Wade-Martins *et al.* 2000).

1.3.3 S/MAR-Based Vectors

In order to avoid the issues discussed above, transgenes were flanked with S/MAR sequences from the 5'-region of CHN50 and were found to be

expressed at a 10-fold higher level than constructs without S/MARs (Fukuda & Nishikawa, 2003). The chicken lysozyme S/MAR confers position-independent regulation of genes in transgenic mice with integrants showing 'normal' regulation of gene expression and copy number independence (McKnight *et al.* 1992). A construct containing a binding site for the thymocyte-specific protein SATB1 upstream of Myc is anchored to the nuclear matrix and confers inducibility to the Myc gene but only in thymocytes (Cai *et al.* 2003). This again shows that not only do S/MARS augment transcription, but that they are also involved in a level of gene regulation.

The use of S/MAR sequences flanking transgenes appears to solve many of the problems associated with the variability of expression of transgenes, but they still rely on integration of the construct into the genome with the potential for the activation of oncogenes. An episomally replicating vector that does not express viral proteins could avoid the problem of cell transformation. To this end a vector, pEPI-1, with S/MARs from the 5'-region of the interferon β-gene (endogenous chromosomal location: 9p21) in place of the gene coding for the SV40 large T-antigen was developed (Figure 1.2). The S/MAR, in association with the SV40 origin of replication, causes the plasmid to be stably maintained without selection for more than 100 generations at very low copy numbers in CHO cells (Piechaczek *et al.* 1999). This is the first instance of a plasmid-based vector not expressing any viral *trans*-acting factors with the capacity for episomal replication in eukaryotic cells.

There have been exciting developments in the quest to elucidate the mechanisms of establishment and maintenance of the episomal state. The key to the success of the vector appears to be an association with components of the nuclear matrix. Baiker *et al* (2000) demonstrated a specific interaction of the vector with a 32 kDa protein that was later found to be a histone H1 variant. FISH analyses suggested there was a non-covalent association of the vector and host chromosomes. Substitution of the S/MAR with other AT-rich sequences such as NTS- and NTS-2 (associated with an endogenous origin of replication in mouse cells) did not bind to the nuclear matrix and did not support the episomal state suggesting that the S/MAR association with the matrix is essential for function. SAF-A was identified as the main protein bound to pEPI-1

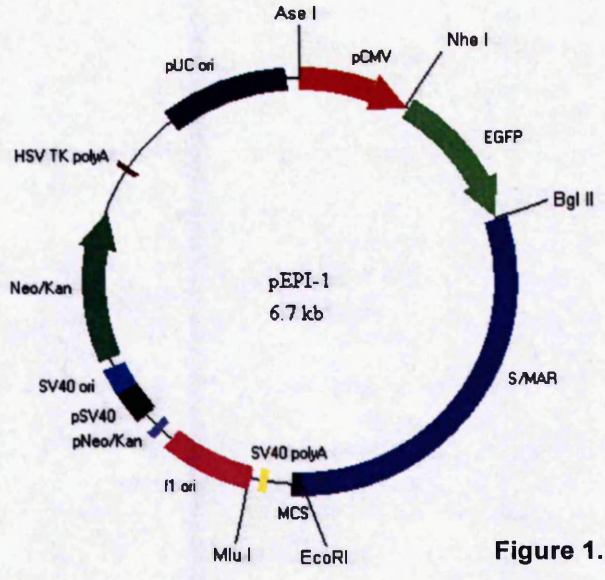


Figure 1.2. Vector map of pEPI-1

S/MARs by the use of cross-linking agents (Jenke *et al.* 2002). Treatment of cells with *cis*-diamminedichloroplatinum II allowed copurification of SAF-A with pEPI-1 (via the S/MAR) and showed that it did not bind any other matrix components. The immunoprecipitation of these cross-linked DNA-protein complexes was the first demonstration of the interaction *in vivo*. The association of the episome with the nuclear matrix is key to its success and probably functions by positioning the episomes close to the matrix bound transcription and replication factories. The stable association with the matrix (or 'mitotic scaffold') would also mediate efficient segregation at mitosis as it could 'piggy-back' with the host cells chromosomes. This is not a novel mechanism as many other episomally retained replicons rely on such a process, for example, EBV uses the virally encoded *trans*-acting factor EBNA-1 to tether the plasmid to the nuclear scaffold at mitosis and this association with the mitotic chromosomes ensures efficient segregation and a place in the newly reformed nucleus (Mackey & Sugden, 1999).

A series of deletion studies were carried out to discover the minimal elements required for episomal behaviour (Stehle *et al.* 2003). It was shown that the only requirements were an active transcription unit upstream of the S/MAR and that this transcription ran on into this region. Insertion of an SV40 poly adenylation site between the EGFP gene and the S/MAR resulted in a loss of function, as did deletion of the gene. The deletion of the SV40 poly adenylation signal did

not affect the function and northern analysis demonstrated that a cryptic termination site within the S/MAR is used normally. This requirement was further modified by the demonstration that a tetramer of a 155 bp minimal S/MAR DNA module can be substituted for the 2 kb S/MAR sequence (Jenke *et al.* 2004b). The 155 bp minimal element comprises of the core unwinding element and it is the mass binding to the SAF-box in SAF-A that supports episomal function; a dimer of the 155 bp element cannot bind SAF-A and integrates into the genome as this length of binding sites is not sufficient to mediate the association.

The persistence of the vector as an episome is essential for its use as a gene therapy vector, but there are also questions regarding the properties of gene expression and replication. The vector is maintained at a low copy number (5-15 copies per cell) as estimated by FISH and quantitative Southern analysis. To achieve this, the vector must replicate efficiently and in a controlled way. Over replication would result in an accumulation of vector, a lack of replication would mean it was gradually lost from the population over time. DNA melting at an origin could be supported by retrieving the energy stored in a nearby base unpaired region, this is possibly why S/MARs have frequently mapped adjacent to origins of replication and have been shown to mediate their function (Pemov, Bavykin & Hamlin, 1998). It has been shown that pEPI-1 replicates semi-conservatively in a once-per-cell-cycle manner (Schaarschmidt *et al.* 2004). This was shown by chromatin immunoprecipitation and quantitative PCR and demonstrated that during G1, Orc1p, Orc2p and Mcm3p (components of the pre-replication complex) are bound to multiple sites on pEPI-1. The sites are functional as there is an S-phase dependent dissociation of Orc1 and Mcm3 (this prevents re-replication of the cellular genome) and nascent strands corresponding to many regions of the vector were identified. The insertion of a well established ORC binding region into the vector did not alter this dispersed pattern of replication initiation. These are important observations as it shows that an origin of replication is defined by epigenetic mechanisms rather than sequence requirements and that any future modifications of pEPI-1 should seek to preserve them otherwise the vector may lose its replicational capabilities.

The vector is stably expressed in some cell types and this is due to the protection either the episomal status or the S/MAR affords to the pCMV. The

promoter used in pEPI-1 is usually silenced by cytosine methylation several days after transfection when linked to an integrated transgene (Prosch *et al.* 1996; Collas, 1998). However, it was shown in CHO and HaCat cells that the promoter remains unmethylated and expresses the gene product stably for over 16 weeks (Jenke *et al.* 2004a). Unfortunately, transfection of pEPI-1 into other cells of mammalian origin did not produce such convincing results (Papapetrou *et al.* 2006). The vector was episomally maintained and conferred sustained EGFP expression under non-selective conditions in K562 and primary fibroblast-like cells. The vector was retained episomally in the murine erythroleukemia cell line (MEL) but expression was apparently silenced as soon as 7 days after transfection. This silencing has been shown to be due to histone deacetylation as treatment with sodium butyrate (a HDAC inhibitor) resulted in EGFP expression. Further analysis showed that in fact there was a low level of expression in these cells and this is consistent with the belief that ongoing transcription into the S/MAR is required for episomal maintenance. This suggests that the interaction of SAF-A with cell and species specific factors means that the episome will behave differently in different cell types. This has huge implications for its intended use as a gene therapy vector; further research is required to determine which cell types are suitable for use with this vector and which are not.

The most exciting development with this system is the report of the production of genetically modified pigs (Manzini *et al.* 2006). The vector was delivered to embryos by sperm-mediated gene transfer and positive animals showed EGFP expression in all tissues analysed (skeleton, muscle, heart, liver, kidney and lung). The vector was expressed in an average of 79% of cells studied in the tissues and was shown to be retained episomally. This demonstrates the potential applications of this system for the delivery of transgenes to whole organisms. The complex interplay between transcription, replication and mitotic segregation with respect to the function of pEPI-1 as an episome is poorly understood. It is clear that a matrix-S/MAR association is essential for the establishment of the episomal state. Abolishing transcription running into the S/MAR also prevents the establishment of the episomal state suggesting that chromatin modifications mediated by the transcriptional activity are required in some way to allow the matrix association to be formed.

1.4 Aims and Objectives

The future use of this vector system as a human gene therapy vector requires a solid understanding of the mechanisms of establishment and maintenance in human cells. It is important to appreciate how and why pEPI-1 plasmids are converted into stable episomes so that improvements can be made to the process and any future modifications could have predictable results.

The initial aim of this study was to determine whether the vector had the ability to be stably maintained as an episome without selection pressure in cells of human origin as the literature at the time had only demonstrated its effectiveness in CHO cells. If this was successful the study aimed to address the possible underlying factors involved in pEPI-1 establishment with a view to making improvements to this step, increasing the efficiency with which stable pEPI-1^{+} clones were generated. The effect of transcription and replication (through studying the cell cycle) on the establishment phase would be studied as these processes appear to be intimately linked to the episomal behaviour of pEPI-1. Once established, the behaviour of the episome would be studied with respect to transcription (using EGFP as a reporter), copy number and mitotic segregation to try to elucidate the nature of their influence on the episomal nature of the vector and any interdependence they may exhibit. Modification of the basic vector to incorporate a *lac*O array (as in section 1.2.7) would allow the tracking of individual episomes in live cells giving information about their nuclear localisation, mitotic segregation and how copy number is linked to levels of reporter protein. It was hoped that this study would add to the understanding of the factors governing the ability of pEPI-1 to be established as an episome and that it would also serve as a benchmark for the behaviour of the vector in human cells so that the effect that any future modifications may have on its performance could be clearly determined.

Chapter Two: Materials and Methods

2.1 Molecular Biology Techniques

2.1.1 Plasmid preparation

2.1.1.1 Plasmids

pEPI-1 was a kind gift from Professor Hans Lipps (Witten/Herdecke University, Germany). pDsRed1-LamC was a kind gift from Dr. Chi Tang (University of Manchester, UK). pUC21-128lac and p3'SS-GFP-lac-NLS were kind gifts from Dr. Wendy Bickmore (MRC Human Genetics Unit, Edinburgh, UK). pEGFP-C1 and pDsRed2-C1 from were purchased from Clontech.

2.1.1.2 *Escherichia coli* strains

All plasmid amplification was carried out using DH5α *E. coli* (Invitrogen) except plasmids containing the Lac operator array, which were prepared using MAX Efficiency Stbl2 cells (Invitrogen).

2.1.1.3 Preparation of LB medium and agar plates

E. coli culture was performed in Luria-Bertani (LB) medium or on LB agar plates. LB medium was prepared by addition of 10 g of bacto-tryptone (Melford Laboratories), 5 g yeast extract (Melford Laboratories) and 10 g sodium chloride (Sigma) to 800 ml of water, the pH adjusted to 7.5. The final volume was made up to 1 L prior to autoclaving. LB agar was made by the addition of 7.5 g of bacto-agar (Melford Laboratories) to 500 ml of LB prior to autoclaving.

2.1.1.4 Transformation of DH5α competent cells

50 µl of competent cells were thawed gently on wet ice prior to addition of 1 µg of plasmid DNA and incubation on ice for 10 minutes. Cells were heat-shocked at 42°C for 40 seconds then returned to the ice for a further 2 minutes. 100 µl of pre-warmed SOC medium (Invitrogen) was added to the cells, they were then incubated at 37°C for 1 hour. Cells were spread on LB agar plates containing either 30 µg ml^{-1} kanamycin or 50 µg ml^{-1} ampicillin (Melford Laboratories), as

appropriate, and incubated at 37°C overnight. Individual colonies were picked using a pipette tip and transferred to LB containing either 30 μg ml^{-1} kanamycin or 50 μg ml^{-1} ampicillin, as appropriate, and cultured in a shaker incubator at 37°C, 225 revolutions per minute (rpm) for 16 hours. 2 ml LB cultures were used for testing the products of ligation reactions and 100 ml LB cultures were used for the production of plasmid DNA used experimentally.

2.1.1.5 Transformation of Stbl2 competent cells

50 μl of competent cells were thawed gently on wet ice prior to addition of 10 ng of plasmid DNA and incubation on ice for 30 minutes. Cells were heat-shocked at 42°C for 25 seconds then returned to the ice for a further 2 minutes. 900 μl of room temperature SOC medium was added to the cells, they were then incubated at 30°C, 225 rpm for 90 minutes. Cells were spread on LB agar plates containing 30 μg ml^{-1} kanamycin and incubated at 30°C overnight. Individual colonies were picked using a pipette tip and transferred to LB containing 30 μg ml^{-1} kanamycin and cultured in a shaker incubator at 30°C, 225 rpm for 16 hours. 2 ml LB cultures were used for testing the products of ligation reactions and 100 ml LB cultures were used for the production of plasmid DNA used experimentally.

2.1.1.6 Plasmid DNA purification

Plasmid DNA was isolated from 2 ml LB cultures using the QIAprep spin miniprep kit (Qiagen) according to manufacturer's instructions. Plasmid DNA was isolated from 100 ml LB cultures using Genopure plasmid midi kit (Roche) according to manufacturer's instructions.

2.1.1.7 Preparation of glycerol stocks

Glycerol stocks were prepared by addition of 0.15 ml sterile glycerol to 0.85 ml of mid-logarithmic phase bacterial cells. Glycerol stocks were stored at -80°C. Cells were recovered by stabbing a sterile pipette tip into the glycerol stock and streaking across an agar plate containing either 30 μg ml^{-1} kanamycin or 50 μg ml^{-1} ampicillin, as appropriate. Plates were incubated overnight at 37°C (DH5α) or 30°C (Stbl2) then individual colonies were picked and grown in 100 ml LB cultures.

2.1.1.8 Determination of DNA concentration

DNA concentrations were determined by measurement of the optical density (OD) at λ = 260 nm. DNA was diluted in molecular biology grade water (Sigma) where required. The DNA concentration (μg μl^{-1}) = (OD_{260} x dilution factor x volume) / 1000.

2.1.2 Sub cloning

2.1.2.1 Restriction digests

20 μl restriction digests were set up containing 1 μl enzyme (10 units), 2 μl 10x restriction buffer, 2 μg plasmid DNA and molecular biology grade water. 50 μl digests were performed when the products were to be separated on an agarose gel to allow excision of a specific fragment. 50 μl digests were carried out using 2.5 μl enzyme, 5 μl 10x restriction buffer, 5 μg plasmid and molecular biology grade water. Restriction enzymes and buffers were from Roche or New England Biolabs. Digests were carried out at 37°C for 1 hour. Double digests were performed in the buffer recommended by the appropriate company. Where the activity of one enzyme fell below 75% in a double digest, restriction reactions were performed sequentially. A QIAquick spinprep kit (Qiagen) was used to remove enzyme and buffer after the first digest.

2.1.2.2 *Eco*RI partial digests

The *lac*O array in pUC21-128lac was shortened by *Eco*RI partial digestion of the array. In order to establish the optimum time, temperature and enzyme concentration to produce a partial digest product with a length of 3-5 kb, a series of conditions were trialled. Once the appropriate conditions had been established by analysing the products on a 0.8% agarose gel, the reaction was scaled up to generate a sufficient quantity of fragments of the desired length. Reactions were carried out in 10 x 20 μl reactions prepared from a 200 μl 'mastermix' to reproduce the molecular kinetics of the trial reaction. The conditions used for the successful partial digest were 10 μg DNA (*lac*O array only), 20 μl 10x restriction buffer, 1.75 μl *Eco*RI, up to 200 μl with molecular biology grade water at 30°C for 5 minutes. The first strategy involved running

the digestion products on a gel and excising the region corresponding to 3-5 kb. The DNA was purified using a QIAquick gel extraction kit (Qiagen) and the DNA concentration determined. This method proved unsuccessful as there was too little DNA produced this way for a ligation reaction. It was decided to attempt a 'shotgun' approach, where the whole of the restricted DNA was added to the ligation; it was hoped that under the correct digestion conditions, there would be an excess of fragments of the desired length thus favouring the ligation of a fragment of 3-5 kb into pEGFP-C1. Partial digestion was carried out as before, the reactions were stopped by using a QIAquick spinprep kit (Qiagen) to remove enzyme and buffer, ethanol precipitated (2.1.2.5) to concentrate the sample, the DNA concentration determined and ligation reactions (2.1.2.8) performed. Ligations were carried out at vector: insert ratios ranging from 2:1 up to 1:9 (calculated using 3 kb as the insert length) to maximise the probability of producing a product with an insert of the desired length. A ratio of 1:5 produced the plasmid described.

2.1.2.3 Alkaline phosphatase treatment

Alkaline phosphatase (Roche) was performed on vectors digested with a single enzyme to prevent re-circularisation during ligation. 1 µl alkaline phosphatase was added to restricted DNA (buffered by the restriction buffer) and incubated at 37°C for 1 hour.

2.1.2.4 Purification of DNA

Restriction enzymes and buffers were removed from the restricted DNA using a QIAquick PCR purification kit (Qiagen) as per manufacturer's instructions.

2.1.2.5 Ethanol precipitation

In some instances the concentration of restricted DNA was too low for efficient ligation so an ethanol precipitation was performed to concentrate and purify the sample. Ice cold absolute ethanol (3x volume) and 3 M sodium acetate, pH 5.4 (1/9 volume) was added to DNA solutions, vortexed to mix and incubated at -20°C for at least 20 minutes. The solution was centrifuged at 15,000 relative centrifugal force (RCF) for 15 minutes, the supernatant carefully discarded and 1 ml 70% ethanol added. The microfuge tube was vortexed for 1 minute then

centrifuged at 15,000 RCF for 1 minute. The supernatant was carefully discarded and the pellet air dried. The DNA pellet was resuspended in an appropriate volume elution buffer (from a QIAquick DNA cleanup system kit).

2.1.2.6 Agarose gel electrophoresis

Restricted DNA was separated on 0.8% (w/v) agarose gels and visualised by staining with 0.5 μg ml^{-1} ethidium bromide in 1x tris-borate-EDTA (TBE) buffer for 10 minutes then using a transilluminator. Agarose gels were prepared by dissolving agarose (Sigma) in 1x TBE buffer. Gels were also run in 1x TBE. Markers used were the GeneRuler 1kb DNA Ladder (Fermentas) or the GeneRuler 100bp DNA Ladder Plus (Fermentas) as indicated. A stock of 10x TBE buffer, pH 8.3, was prepared using 108 g Tris Base (Sigma), 55 g boric acid (Sigma) and 7.44 g EDTA-Na_2-salt and adjusting the volume to 1 L with dH_2O.

2.1.2.7 Excision of restricted DNA

DNA of the required size was excised from the gel using a clean, sharp razor blade taking care to trim away any excess gel. Restricted fragments were recovered using a QIAquick gel extraction kit (Qiagen) according to manufacturer's protocol.

2.1.2.8 Ligation reactions

Ligations were performed using a vector: insert ratio of 1:3 and 1:1 in 10 μl reactions containing 100 ng vector DNA, the appropriate quantity of insert DNA, 1 μl T4 DNA ligase (Roche) and 1 μl ligase buffer (10x). Ligations were incubated at 16°C for 16 hours prior to transformation into *E. coli*.

2.1.2.9 Polymerase chain reaction

Oligonucleotide primers (Sigma) were designed to be complementary to the ends of the S/MAR sequence in pEPI-1 and to include *Ksp*I and *Bam*HI restriction sites at the 5' and 3' ends respectively. The primers were solubilised to 100 μM stock solutions in molecular grade water; working solutions of 10 μM were prepared to avoid repeated freeze-thaw cycles.

The primer sequence (5' to 3') used was

Forward primer:

CTTTGTACCGCGGCTAAATAAACTTATAAATTGTGAG

*Ksp*I

Reverse primer:

GCGCGGGGATCCCTATCAAGATATTTAAAGAAAAAAAATTG

*Bam*HI

The optimum annealing temperature for the primer set was determined by gradient PCR on a Mastercycler gradient machine (Eppendorf). Gradient PCR was performed in 20 µl reactions aliquoted from a 100 µl 'mastermix' shown in Table 2.1. The dNTP mix was made as a 10 mM stock using equal quantities of dATP, dCTP, dTTP and dGTP. The cycle parameters for the gradient PCR are shown in Table 2.2.

Table 2.1 PCR Mastermix

Reagent	Volume (µl)
dH_2O	74
10x High Fidelity *Taq* buffer (+$MgCl_2$)	10
DMSO	5
Forward primer (10 µM stock)	3
Reverse primer (10 µM stock)	3
dNTP mix	2
pEPI-1 plasmid (100 ng μl^{-1})	1
High Fidelity *Taq* (Roche)	1.5

Table 2.2 Gradient PCR cycle parameters

Step	Number of Cycles	Time (minutes)	Temperature (°C)
A	1	5	95
B	30	1 1 4	94 60-68 72
C	1	10	72
D	1	1	20

The gradient PCR products were resolved on an agarose gel (2.1.2.6) and the most suitable annealing temperature was determined to be 62°C. Amplification of the S/MAR sequence containing the altered restriction sites was carried out

using a TC-312 thermal cycler machine (Techne) according to the parameters in Table 2.2 except it was run at 62°C in Step B rather than using the gradient. Amplification reactions were performed in 20 µl reactions aliquoted from the 100 µl master mix shown in Table 2.1. The PCR product was purified from the reaction mix using a QIAquick PCR purification kit (Qiagen) according to manufacturer's instructions and then resuspended in 30 µl molecular biology grade water prior to restriction with *Ksp*I and *Bam*HI (2.1.2.1). The PCR product was purified (2.1.2.4) and subsequently used in ligation reactions (2.1.2.8).

2.2 Cell Culture Techniques

2.2.1 Cell lines and culture medium

HeLa cells were from the European Collection of Cell Cultures (ECACC). The cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM), product number D5671, with 4500 mg L^{-1} glucose and sodium bicarbonate (Sigma) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) (Sigma), 1% (v/v) 200 mM L-glutamine (Sigma), 1% (v/v) penicillin-streptomycin (100x, Sigma), 1.5% (v/v) sodium pyruvate (Sigma) and 1.5% (v/v) MEM non-essential amino acid solution (100x, Sigma).

2.2.2 Maintenance of cell lines

HeLa cells were cultured at 37°C with 5% carbon dioxide (CO_2) and 95% humidity in T25 and T75 angle-necked, screw-top flasks (Nunc or Corning). This adherent cell line was passaged by removal of medium, washing with phosphate buffered saline (PBS) then incubation with Trypsin-EDTA (Sigma) for 5 minutes at 37°C, 5% CO_2, 95% humidity. Gentle agitation of the flask was employed to release the cells from the flask, at which point the trypsin was inactivated by the addition of fresh DMEM. Cells were sub-cultured every 3-4 days and transferred to fresh medium at a 1:10 ratio to maintain log-phase growth. PBS was prepared by dissolving PBS tablets (Sigma) in water prior to autoclaving to sterilise.

2.2.3 Cryopreservation of cell lines

Cells from a confluent T75 flask were passaged as in section 2.2.2 to the point of trypsin inactivation then centrifuged at 500 RCF for 5 minutes at 4°C. The supernatant was discarded and the cell pellet resuspended in 3 ml of FBS supplemented with 5% (v/v) sterile DMSO (Sigma). Cells were aliquoted into cryovials (Nunc or Corning), 1 ml per vial, and frozen at -80°C overnight, then transferred to liquid nitrogen for cryopreservation.

2.2.4 Viable cell count

Viable cell numbers were determined by diluting 50 µl of cell suspension in 50 µl trypan blue (Sigma) and using a haemocytometer to count the viable cells.

2.2.5 Transfection of HeLa cells

HeLa cells were transfected using PolyFect (Qiagen) according to manufacturer's protocols. Briefly, $4x10^5$ cells were seeded into a 35 mm dish 24 hours prior to transfection. Cells were transfected with 1.5 µg DNA (for double transfections the ratio of plasmids was determined empirically but the total DNA remained a maximum of 1.5 µg) diluted in OptiMEM (Invitrogen) and 12 µl PolyFect. FACS analysis was employed to determine the transfection efficiency 48 hours later.

2.2.6 Generation of long-term stable cultures

2.2.6.1 G418 selection

Cells were transfected with pEPI-1 (or derivatives) as in section 2.2.5 and cultured for 48 hours, after this time they were seeded into T25 flasks. G418 (500 µg ml^{-1}) drug selection was started at this point. Medium was refreshed every 3-4 days to remove dead cells and replace deactivated G418; this was done carefully to avoid detaching mitotic cells and seeding secondary colonies. Drug selection was performed for a period of 3 weeks after which time colonies were counted and the heterogeneous population was transferred to a fresh T25 and cultured in the absence of selection.

2.2.6.2 Isolation of single cell clones

Single cell clones were produced in a similar way to 2.2.6.1, except cells were seeded into four 96 well plates 48 hours post transfection. Cells were seeded at 2-3 per well, this was found to be the optimum density for the generation of approximately 30 wells containing single colonies at the end of the selection period. G418 (500 μg ml^{-1}) selection was added at the same time as seeding. The medium and drug was refreshed every 3-4 days using a multi-channel pipette (though tips were discarded after every row of wells to prevent cross-contamination). Cells were cultured in 96 well plates for approximately two weeks then they were transferred to 24 well plates to allow expansion of the colonies and cultured for a further week in G418 supplemented medium. Cells were transferred from 24 well plates into T25 flasks for long-term culture. A sample of cells from each clone was expanded and stored in liquid nitrogen to allow repeats of experiments to be performed on clones with the same passage number.

2.2.7 Cell synchronisation prior to transfection (for use in section 3.2.5)

HeLa cells were synchronised using 2.5 μg ml^{-1} aphidicolin treatment for 12 hours, released by three changes of medium in 30 minutes and cultured for 24 hours then aphidicolin treated for a further 12 hours to achieve ~90% synchronisation. Cells were transfected with pEPI-1 DNA 1 hour after release for the early S phase time point. A mitotic shake off was performed 12-14 hours after release and half the cells transfected for the mitosis time point. The remaining cells recovered from the mitotic shake off were cultured for 2½ hours prior to transfection giving the early G1 time point. The mitotic shake off was performed by gently rinsing cells in the flask with 2 ml of medium 5-10 times to disturb the mitotic cells. These were counted using a haemocytometer and transferred to a 35 mm dish for transfection (2.2.5).

2.2.8 Cell culture for live cell imaging

Live cell dishes were prepared by coating with 15 μl poly-lysine (Sigma) for 10 minutes. Cells were seeded into poly-lysine-coated, gridded, glass bottomed

culture dishes at least 24 hours prior to microscopy. Once the cells had settled, they were treated with 2.5 μg ml^{-1} aphidicolin for 12 hours and were released by washing three times with medium over a 30 minute period. Cells were cultured in colourless medium (Sigma) supplemented with 25 mM HEPES (Sigma) to help buffer the solution. Cells were cultured in a standard incubator and were only moved to the microscope briefly at each time point for image capture. Image capture was performed on a heated stage in a heated, humidified chamber with 5% CO_2 to minimise the stress to the cells.

2.3 Analysis of pEPI-1$^+$ Long-Term Cultures

2.3.1 Vector rescue

2.3.1.1 Hirt extraction of extrachromosomal DNA

Cells were grown in 100 mm dishes (Corning) and the extraction (Hirt, 1967) was performed on approximately $1x10^6$ cells. Cells were washed twice with 10 ml TEN buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 150 mM NaCl,). 1.5 ml TEN buffer was combined with 1.5 ml of 2x Hirt solution (2x Hirt solution = 20 mM Tris-HCl pH 7.5, 20 mM EDTA, 1.2% (w/v) SDS), spread over the dish and incubated at room temperature for 20 minutes to allow cell lysis. The lysed cells were scraped from the dish using a cell scraper (Nunc) and transferred to a 15 ml Falcon tube where 750 μl 5 M NaCl (to a final concentration of 1 M) was added and the tube was inverted gently ten times. This solution was incubated overnight at 4°C then centrifuged at 40,000 RCF for 1 hour, also at 4°C to remove the precipitates. Proteinase K (Sigma) was added to the solution to a final concentration of 100 μg ml^{-1} and incubated for 3 hours at 37°C. A phenol-chloroform extraction and ethanol precipitation (2.1.2.5) was performed and the DNA/RNA pellet was resuspended in 100 μl Tris-EDTA (pH8.0). The RNA was digested by addition of RNase A (Sigma) to a final concentration of 50 μg ml^{-1} for 1 hour at 37°C. A second ethanol precipitation was performed and the DNA was resuspended in 10 μl molecular biology grade water (Sigma).

2.3.1.2 Retransformation of *E. coli*

To confirm the vector recovered by Hirt extraction had replicated in HeLa cells, the extracts were 'digested' with *Dpn*I (2.1.2.1). Only bacterially-derived DNA can be cleaved by this enzyme therefore the vector should remain circular if it has replicated in the HeLa cells. The digestion mix was transformed into *E. coli* as in section 2.1.1.4 and positive transformants were selected on agar plates using 30 $\mu g\ ml^{-1}$ kanamycin.

2.3.1.3 Analysis of recovered vectors

Individual bacterial colonies were picked at random and the DNA purified (2.1.1.6). This recovered DNA was subjected to control restriction and compared with the original untransfected vector to ensue no genetic rearrangements had occurred. A sample of this recovered plasmid was also used to retransfect wild type HeLa cells to ensure similar behaviour.

2.3.2 Fluorescence in situ hybridisation

2.3.2.1 Chromosome preparation

Cells were passaged 24 hours prior to the preparation to ensure they were mitotically active and were 70% confluent for the preparation. Cells were transferred to a 15 ml Falcon tube and 0.2 $\mu g\ ml^{-1}$ colcemid (Sigma) was added, the cells were then incubated in a 37°C water bath for 2 hours. Cells were pelleted by centrifugation at 180 RCF for 5 minutes, the supernatant was removed (the final drop was left) and the cells resuspended in 10 ml pre-warmed 0.075 M KCl. The suspension was incubated in a 37°C water bath for 20 minutes prior to centrifugation at 180 RCF for 5 minutes. The pellet was resuspended drop wise in 5 ml of 3:1 (v/v) ice-cold methanol:acetic acid (Sigma, molecular biology grade) and this was left at room temperature for 1 hour. The cells were pelleted and resuspended in methanol:acetic acid as before. The chromosome preparations were 'dropped' (close to the slide to prevent shear forces detaching episomes from mitotic chromosomes) onto glass slides and dried on a 70°C heat-block.

2.3.2.2 Preparation of hybridisation mix

The 'BioNick Labelling System' was used (Invitrogen) as per manufacturer's protocol. All reagents were supplied in the kit except 1 μg of pEGFP-C1 plasmid was added to act as the probe. The pEPI-1 vector was not used as the S/MAR may hybridise to endogenous S/MARs in addition to the episomes. The probe was purified using a QIAquick PCR cleanup kit and eluted as 50 μl in EB. The hybridisation mix was prepared as a master mix by adding 50 μl formamide (Sigma), 10 μl 20x SSC, 10 μl of 500 μg ml^{-1} salmon sperm competitor DNA (Invitrogen), 12 μl of 0.2-2.0 μg ml^{-1} biotinylated probe, 8 μl molecular water, 10 μl dextran sulphate (Sigma). The mix was vortexed briefly to mix then aliquoted into microfuge tubes, 10 μl each. The microfuge tubes were incubated in a 70°C water bath for 5 minutes then were immediately placed in a 37°C water bath for 15 minutes.

2.3.2.3 Labelling

The slides were incubated with 100 μg ml^{-1} RNAase A in 2x SSC (from 20x SSC which is 175.3 g NaCl, 88.2 g Sodium Citrate to pH 7 in 1 L) in a slide box at 37°C in a water bath for 1 hour. The slides were washed in 2x SSC at room temperature then dehydrated through 70%, 90% and 100% ethanol soaks for 2 minutes each then thoroughly air-dried. Slides were then warmed to 37°C on a metal tray in a water bath for 5 minutes to coincide with the final incubation of the probe from section 2.3.2.2. The hybridisation mix was dropped onto a clean coverslip then the slide was mounted and sealed ensuring it was centred and free of bubbles. This was done one slide at a time as it was imperative that the slides and the probes were kept at 37°C. The mounted slides were placed on a metal tray in a 37°C water bath and left overnight to hybridise.

A solution 1% (w/v) Bovine Serum Albumin (BSA) in 20x SSC was prepared and aliquoted into four 100 μl portions into microfuge tubes and the following was added; to the first, 2 μl mouse anti-biotin (Invitrogen), the second, 2 μl Alexa Fluor 488 conjugated goat anti mouse (Invitrogen), and the third, 2 μl FITC conjugated donkey anti goat (Invitrogen), the fourth was left plain (the 'block'). These were centrifuged at 15,000 RCF for 15 minutes at 4°C. The pellet was discarded.

The slides were removed from the metal tray and washed four times in 2x SSC at 45°C for three minutes each wash. The coverslips should float off during the first wash. They were then immersed in 4x SSC/Tw (200 ml 20x SSC, 1 ml Tween 20 up to 1 L with water) for 2 minutes at room temperature. The slides were then treated with the blocking buffer by adding 40 μl to a strip of Parafilm and lowering the slide onto it and this was left at room temperature for 5 minutes. The Parafilm was removed and the first antibody applied as with the block. The slide was placed in a box containing dampened paper to prevent drying out and this box was placed in a 37°C incubator for 30 minutes. The box was removed from the incubator and the Parafilm taken off the slides, which were then washed three times in 4x SSC/Tw for 2 minutes each time at 37°C. The second antibody was applied in the same way as the first, and again incubated in the damp box at 37°C for 30 minutes. This was repeated for the application of the third and final antibody. The slides were washed finally in 4x SSC/Tw and then mounted with VectorShield containing propidium iodide (PI) and sealed using nail polish.

2.4 Immunostaining

Cells were transfected with p3'SS-GFP-lac-NLS and cultured for 24 hours prior to seeding onto coverslips. 2.5 μg ml^{-1} Aphidicolin was used 12 hours after seeding to synchronise the cells; they were released 12 hours later by thorough washing with 3 medium changes over a 30 minute period. Cells were monitored for the onset of mitosis and were fixed shortly afterwards, during early G1. Fixation was by treatment with methanol pre-chilled to -20°C for three minutes then cells were allowed to air dry. All immunofluorescence steps were carried out on ice in a humidified chamber in the dark. Samples were blocked by incubation with PBS^+ (2% (w/v) BSA in PBS, prepared fresh) for 1 hour then washed once using PBS. The primary antibody (rabbit anti-GFP polyclonal Av peptide antibody; B.D. Living Colours,) was diluted 1:200 in PBS^+ and incubated with the coverslips for 1 hour. The primary antibody was removed by three times, 5 minute wash steps with PBS. The second block step was carried out using 2% (w/v) goat serum in PBS^+ for 30 minutes. The secondary antibody

(Alexa Fluor 488 goat anti-rabbit IgG; Invitrogen Molecular Probes) was diluted to 5 μg ml^{-1} in PBS^{+} and incubated with the cells for 30 minutes. The secondary antibody was removed by washing five times with PBS, 5 minutes per wash, after this the cells were mounted on glass slides using VectaShield and sealed with nail polish.

2.5 Image Analysis

2.5.1 Slide preparation

HeLa cells were seeded onto sterile glass coverslips in 24 well plates and cultured for 24 hours prior to fixation. Coverslips were removed from the medium and washed with PBS. Cells were fixed using 4% (v/v) paraformaldehyde (Sigma), 0.1% (v/v) Triton X-100 (Sigma) in PBS for 10 minutes and then washed 5 times with PBS. Nuclei were stained using DAPI 1:2000 (v/v) in PBS for 10 minutes before washing 5 times with PBS. Coverslips were mounted onto glass slides in VectorShield (VectorLabs) and were sealed using clear nail varnish.

2.5.2 Confocal laser scanning microscopy

2.5.2.1 Fixed samples

Images were analysed by CLSM using a Zeiss LSM 510 Meta confocal laser scanning microscope equipped with oil immersion 63x and 100x objective lenses. Fluorochromes were visualised using an argon laser with excitation wavelengths of 488 nm for EGFP and Alexa Fluor 488, and using the HeNE1 laser at 543 nm for Cy3 (DsRed). For acquisition of the DAPI, the mercury lamp was used to produce the image (the system does not have a UV laser). Image resolution was 1020x1024 pixels and signal-to-noise ratio was improved by taking an average from four successive scans for each image or slice. Z-stacks were generated by scanning slices sequentially with an interval of 0.2 μm.

2.5.2.2 Live cell imaging

Live cell imaging was performed using the above microscope, lasers and objectives (2.3.2.1). Cells were maintained in the culture incubator for the

majority of the experiment period. Cells were transferred from the incubator to the microscope chamber quickly but care was taken not to agitate the culture dish and disturb any loosely attached mitotic cells. The microscope chamber and stage were preheated and maintained at 37°C and the atmosphere was humidified and kept at 5% CO_2. Image resolution was 1024x1024 pixels and the signal-to-noise ratio was improved by taking an average from two successive scans. The number of scans was decreased from four to reduce the stress and damage to the cells.

2.5.3 Image processing

2.5.3.1 Intensity analysis

Intensity analysis was performed using the WCIF Image J program (ImageJ 1.34m). The intensity of EGFP expression in each clone was determined by taking the mean intensity of 10 cells in each clone. The intensity of each cell was calculated by averaging the intensity readout along a straight line across the whole cell. Mean (n=10) background intensity was subtracted from each value.

2.5.3.2 Analysis of immunostained cells

Images were processed using the Imaris program (Bitplane AG Scientific Solutions v4.2.0) and are presented either as a 'snapshot' of the three dimensional projection formed using the 'surpass' function, or as a gallery of the Z-stack.

2.6 Flow Cytometric Analysis and Cell Sorting

Flow cytometric analysis and FACS were performed on a FACS Vantage (Becton Dickinson) flow cytometer and the data was analysed using CellQuest (Becton Dickinson). Transfection efficiency was determined 48 hours post transfection. Cells were treated with trypsin to put them into suspension, then centrifuged at 500 RCF for 5 minutes at 4°C. The supernatant was discarded and the cell pellet resuspended in 1 ml of HBSS (Gibco) supplemented with 10% (v/v) FBS. For analysis or sorting of cells expressing EGFP, 5 μg ml^{-1} PI

(Molecular Probes) was also added to allow identification of live cells. PI was omitted from suspensions of cells expressing DsRed as the emission spectra overlapped. Cells were sorted into 6 well plates and cultured as normal. Where G418 selection was required, cells were allowed to recover in drug-free medium for up to 24 hours first.

2.7 Statistical Analysis

A student t-test was used to determine any significance between sets of data; the degree of significance corresponded to $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***). All error bars were calculated as the standard error of the mean as they were determined from independent experiments.

Chapter Three: Investigating the Establishment of Stable Clones of pEPI-1^{+} HeLa Cells

3.1 Introduction

3.1.1 Establishing stable long-term cultures with episomal pEPI-1 in HeLa cells

When this chapter was researched, published studies concerning the episomal maintenance of pEPI-1 had been performed using Chinese Hamster Ovary (CHO) cells (Piechaczek *et al.* 1999; Baiker *et al.* 2000). The development of this vector for applications in human gene therapy, by its definition, requires an understanding and enhancement of the system in human cells rather than rodent. For this reason, HeLa cells were chosen as the cell line with which to carry out these experiments. It was essential to initially ascertain whether pEPI-1 was able to be maintained stably as an episome in HeLa cells with characteristics similar to those published for CHO cells. To this end, HeLa cells were transfected with pEPI-1 vector and the expression of the reporter gene, EGFP, was investigated over a period of several months following selection with G418 to produce stable clones. Vector rescue analyses were performed to confirm the presence of the vector as an episome in the populations.

3.1.2 Efficiency of stable clone formation

The first aspect of this investigation focussed on the relatively low efficiency of establishment of stable episomal replicons in mammalian cells. Plasmids containing *ori*P, the latent origin of replication for Epstein-Barr virus (EBV) give rise to drug-resistant clones in which the *ori*P replicon is established with an efficiency of 1 to 10% (Leight & Sugden, 2001b). It would be feasible to assume the efficiency of establishment of stable episomal behaviour of pEPI-1 in mammalian cells was similar to that observed for *ori*P plasmids as there are many similarities in their manner of maintenance (association with the nuclear

matrix and 'piggy-backing' with mitotic chromosomes to ensure segregation). This low efficiency of stable clone formation represents a serious problem if a modified version on pEPI-1 is to be used for gene therapy applications; an understanding of the factors contributing to stable clone formation would be essential as research could be directed towards improving this inefficient step. Leight and Sugden (2001b) demonstrated that the establishment of an *ori*P replicon was dependent on the acquisition of a particular (as yet unknown) epigenetic state and it is likely the same would apply to the establishment of pEPI-1 episomes. Under the conditions used, the formation of this privileged epigenetic state is a rare event and so occurs only in a minority of cells. There may be many factors that influence the probability of this event occurring that can be manipulated to increase its frequency, thereby creating more positive stable clones from a transfection event.

3.1.3 Transcriptional interference

The use of G418 to select for positive cells could impact on the efficiency of stable clone formation as there may be competition between the SV40 early promoter associated with the *Kan/Neo*^{R} gene and the CMV promoter upstream of EGFP and the S/MAR. This effect is termed transcriptional interference (TI) and refers to the direct negative impact of one transcriptional activity on a second transcriptional activity in *cis*. It has been postulated that TI provides another level of gene regulation, adding yet more layers to the complexity of the control of transcription. TI has been observed in the genomes of higher eukaryotes within their natural chromosomal context. N-ras (a small GTPase) is downregulated by transcription of the *unr* gene upstream; deletion of the *unr* promoter in a mouse model results in an increase in transcription of N-ras and is embryonic lethal in the homozygous state (Boussadia *et al.* 1997). Another example of TI is found at the mammalian β-globin locus (described in 1.2.5) and concerns the influence of the transcription of 2 'embryonic' β-like globin genes (Ey and βh1) on the expression of 2 'adult' β-type globin genes (β-major and β-minor). In the wild-type situation, Ey and βh1 are transcribed at high levels only during primitive erythropoiesis while β-major and β-minor are expressed at low

levels. The developmental upregulation of the adult β-like globin genes is coincident with the silencing of the embryonic β-like globin genes and allows high levels of β-major and β-minor in foetal and adult erythropoiesis. In accordance with the role of the LCR, high level expression of all the genes in the locus requires its presence, however, it has been shown that the LCR has no impact on the developmental regulation of the individual genes. Deletion of the LCR results in the expression of all the genes at a very low level but still in a tissue-specific and developmental-specific pattern (Hu *et al.* 2007), this would agree with the role of the LCR as an enhancer in that it has the ability to increase the transcription of those genes marked as 'on' by the specific transcription factors. The correct pattern of expression at this locus can be explained by a set of deletion experiments to show TI as the mode of regulation (Hu *et al.* 2007). (i) Deletion of β-major increased expression of the downstream β-minor gene, (ii) deletion of Ey, βh1 or both increased β-major and -minor expression in the embryo, (iii) deletion of Ey did not affect the transcription of βh1, (iv) deletion of Ey greatly increased transcription of βh0 (a weakly expressed embryonic gene) without affecting βh1, (v) deletion of βh1 had no effect on either Ey or βh0. These results are consistent with a direct transcriptional interference of Ey on βh0 and of Ey and βh1 on the β-major and -minor genes in the embryo and of β-major on β-minor in later developmental stages rather than a previously held model of LCR competition. In this model, all the β-like globin genes compete for a rate-limiting interaction with the LCR and proximity to the LCR would favour the LCR-promoter interaction. The autonomous silencing of the LCR-proximal embryonic genes would prevent them interacting with the LCR and thereby stimulate transcription of the adult genes.

TI has also been demonstrated with artificial promoter arrangements (Eszterhas *et al.* 2002). This study investigated the expression of two reporter genes each flanked by a pCMV and an SV40 large T antigen poly adenylation signal in various arrangements (convergent, divergent and tandem) in both orientations integrated at two different chromosomal loci. It was concluded that "the two transcriptional units interfere with each other in ways that cannot be fully explained by simple models". TI is also a common means of gene regulation in extrachromosomal elements such as bacteriophages (Dodd & Egan, 2002;

Saha, Haggardljungquist & Nordstrom, 1987; van Rijn *et al.* 1989; Xu & Koudelka, 2000), transposable elements (Kimura & Yamaguchi, 1998), insertion sequences (Simons *et al.* 1983) and plasmids (Jagura-Burdzy & Thomas, 1997; Brantl & Wagner, 1997). Whereas TI is employed as a means of regulation in these instances, it is possible that the arrangement of transcription units in pEPI-1 is such that there are unwanted interactions between the two promoters resulting in 'interference'.

If there was promoter interference at work then the *Kan/Neo*^{r} gene would need to be the dominant promoter to confer resistance to G418 meaning that the EGFP gene becomes much less active, this could allow the gradual formation of a more repressive chromatin environment to be established around the S/MAR. Previous work (Stehle *et al.* 2003) has shown that active transcription from a gene upstream of the S/MAR is essential the episomal behaviour of pEPI-1. If this reduction in transcription occurred during the critical period then it might influence the frequency with which vectors formed stable associations with the nuclear matrix resulting in a loss of vector molecules at each cell division due to the obvious negative impact on segregation. For this reason it was postulated that addition of G418 at a time point later than 48 hours post transfection might prevent any interference between promoters during the 'crisis period' and therefore increase the percentage of stable clones obtained.

3.1.4 Dependence of cell cycle phase

The efficiency of stable clone establishment depends upon each introduced vector attaining an epigenetic state permissive for long-term maintenance, however, the issue of efficient gene delivery into the nuclear environment is an essential prerequisite as without this there would be no chance of transcription or replication. Many widely used transfection techniques (including PolyFect) are able to successfully overcome the barrier imposed by the plasma membrane of a cell, it is once the plasmid is in the cytoplasm that problems are encountered. Firstly, the packaging particles need to be dismantled to release the DNA then the plasmid must tackle the obstacle of the nuclear membrane, all the while evading degradation by the cell's own defence mechanisms. Whilst

plasmids can be imported via the nuclear pore complex (NPC) (Dean, Strong & Zimmer, 2005), the most common form of nuclear entry is believed to occur following mitosis. The breakdown and reassembly of the nuclear membrane that occurs during mitosis offers an opportunity for cytoplasmic plasmids to be included in the newly formed nucleus. Studies measuring the expression of reporter genes following transfection into various cell types have shown a link between and the efficiency of transfection and the cell cycle phase, reaching a maximum when timed to exploit the reconstruction of the nuclear membrane following mitosis. A study using HeLa cells supports this theory that mitosis is required for efficient reporter gene (GFP) expression, and by extension, the delivery of the plasmid to the nucleus. They found that GFP expression was greatly enhanced in cells that had undergone a mitotic event compared with cells that had not divided. GFP expression could be detected in a few of the undivided cells, consistent with the theory of a limited amount of nuclear import via the NPC, but this was a rare event (Escriou *et al.* 2001).

The literature concerning the impact of cell cycle phase on transfection efficiency generally shows that transfection of cells during mitosis compared with during early G1 (a) increases the number of positive cells and (b) increases the amount of detected reporter gene product (GFP) (Brunner *et al.* 2000). This increase in GFP detected could be due to either an increase in transcription from the same number of plasmids, or the same transcription level from an increased number of plasmids per cell. In terms of increasing the efficiency of stable clone formation of cells transfected with pEPI-1, an increase in the number of $EGFP^+$ cells obtained from each transfection event would help to a certain degree with the absolute numbers of stable colonies generated at the end of the selection period but it would do nothing to improve the actual efficiency of the process. Situation (b) is slightly more complex, an increase in the number of vector molecules per cell might increase the probability of that cell possessing stable episomes at the end of the selection period if each vector has a fixed probability of establishing itself as an episome – the more vector copies per cell, the more chance there will be that some of them become established. Alternatively, if there is the same vector copy number per cell and the increase in detected EGFP is due to an increase in the transcription levels, this would bring the vector into contact with the transcription machinery at the

matrix more frequently and increase the opportunity for the S/MAR to form stable matrix associations.

It was noted that the percentage of cells able to establish stable clones was similar to the mitotic index of HeLa cells, that is, ~5%. It was possible that the introduction of pEPI-1 to cells at mitosis would facilitate the establishment of an appropriate epigenetic status by promoting an immediate interaction with components of the nuclear matrix. The newly reassembled nucleus would contain pEPI-1 vectors that were matrix tethered and so already with the ability to efficiently contact the transcription and replication machinery. Transfections at specific time points during the cell cycle should reveal whether the cell cycle phase has any impact on the establishment of pEPI-1 stable clones.

3.1.5 Stochasticity

It is possible that the process of establishing a epigenetic state on pPEI-1 vectors conducive to long term maintenance is due to a purely random, stochastic event (or events) and there is little that can be manipulated to alter the frequency of this event occurring. If this is true then at any one time a certain percentage of input vectors should achieve this favourable epigenetic state and establish stable colonies. It therefore follows that a greater number of input vectors in each and every cell should yield a greater number of stable clones across the transfected population. If this were the case then investigations could be directed at transfecting greater numbers of plasmid per cell to boost the number of positive clones after selection. Tseng *et al.* (1997) investigated the expression of EGFP as a reporter gene with respect to plasmid number in HeLa cells following transfection. FACS analysis was employed to determine the number of EGFP molecules and plasmids in single cells. 30% of transfected cells were positive for EGFP expression but of this subpopulation, the vector demonstrated 100% efficiency in transgene expression. This indicates that sorting cells following transfection with pEPI-1 into fractions corresponding to EGFP expression (and by extension, pEPI-1 copy number) should reveal whether there is a correlation between input number and the

probability of achieving stable clone status as would be predicted by a purely stochastic mechanism for the epigenetic modifications.

The aims of this chapter were:

1) To investigate whether pEPI-1 was maintained episomally in stable long-term cultures in a human derived cell line.
2) Characterise the process of stable clone formation to gain a better understanding of the factors governing the switch from transient expression to stable long-term maintenance.
3) Improve the efficiency of stable clone formation.

3.2 Results

3.2.1 The vector pEPI-1 is maintained episomally in HeLa cells in long-term cultures

HeLa cells were transfected with pEPI-1 DNA and investigated for expression of the reporter gene, EGFP, using the confocal laser scanning microscope (CLSM) and the fluorescence activated cell sorter (FACS). Transient transfections were examined 48 hours post-transfection. Long-term cultures were produced by selection with 500 μg ml^{-1} G418 for 3 weeks after which time the cells were cultured in the absence of selection pressure. Analysis of transiently transfected HeLa cells shows a broad range of EGFP expression (Figure 3.1 (b)) with a proportion of cells possessing very high levels of EGFP expression. This is in contrast to stable long-term cultures, which show a much narrower range of EGFP expression, even in a heterogeneous population (Figure 3.1 (d)). It should also be noted that the absolute levels of EGFP measured are much lower in the stable cultures than the transiently transfected cells. This would be consistent with the transiently transfected population possessing many more copies of the vector per cell compared with the stable population.

Cells transfected with a lipid-based delivery system receive a random number of plasmids. If it is assumed that total EGFP expression in an individual cell is proportional to the number of vector molecules then Figure 3.1 (b) suggests that there is a broad range of initial vector input number. Selection of these pEPI-1 positive cells with G418 would produce a mixed stable population as the cells within it would have originated from many different progenitors, all with variable numbers of input vector. To standardise any future experiments, all stable long-term populations were cloned from single cells; this should ensure that any future analyses performed on a clonal population should be standardised with respect to episome number in cells derived from the same clone.

HeLa clones (n=6) were cultured in the absence of selection pressure for up to 10 months (a passage number of approximately 100) and it was observed that expression of EGFP remained stable and the vector retained its episomal status. Figure 3.2 shows CLSM images of a representative clone, B4, taken at various time points and illustrates the maintenance and stability of EGFP

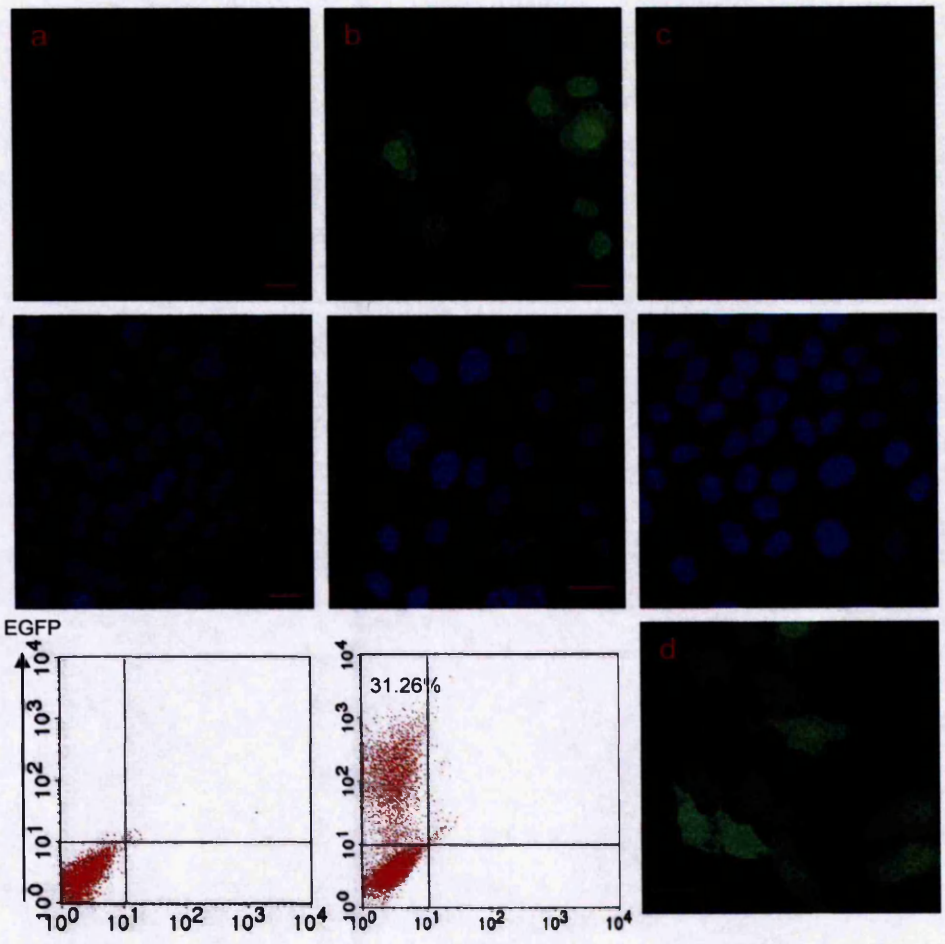


Figure 3.1. HeLa cells transfected with pEPI-1 express EGFP in transient and in long-term cultures following drug selection. HeLa cells were transfected with pEPI-1 and cultured for 48 hours, at which point cells were either analysed for transient expression or cultured in media supplemented with 500 μg ml^{-1} G418 for 3 weeks to generate a stable population. Cells were analysed by CLSM and FACS. (a) Mock transfected HeLa cells as a control. With corresponding DAPI stain and FACS plot. (b) Transiently transfecting HeLa cells with pEPI-1results in EGFP expression when analysed 48 hours post transfection. With corresponding DAPI stain and FACS plot. (c) Untransfected HeLa cells as a control with corresponding DAPI stain. Prepared at the same time as (d). Heterogeneous population of HeLa cells stably expressing EGFP in the absence of selection pressure. Cells were cultured in drug-free media for 5 weeks prior to image capture. Green = EGFP, blue = DAPI DNA stain. Scale bar represents 20 μm.

expression. The episomal status of the plasmid was established by extraction of extrachromosomal DNA followed by vector rescue. A modified Hirt protocol (2.3.1.1) was used to isolate extrachromosomal DNA from approximately $1x10^6$ stably transfected HeLa cells and used to transform *E. coli*. These transformants were selected for using 30 μg ml^{-1} kanamycin, and individual colonies were amplified and the plasmid DNA isolated. Restriction analysis was performed on the recovered plasmid and compared with a control restriction digest of the original pEPI-1 plasmid. Figure 3.2 (f) is a restriction digest of Hirt

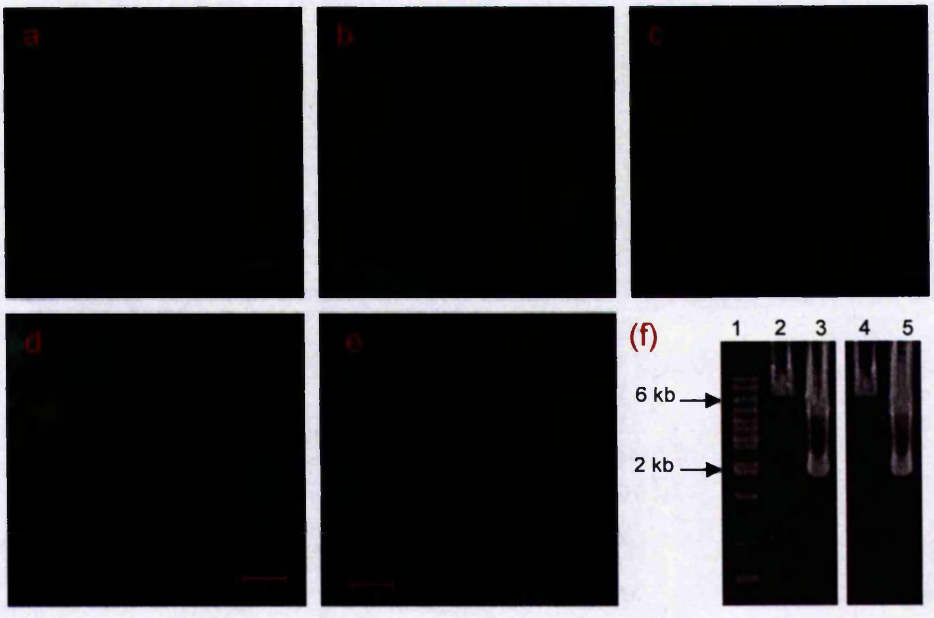


Figure 3.2. EGFP expression is maintained in long-term cultures of HeLa cells transfected with pEPI-1. CLSM images are of clone B4. HeLa cells were transfected with pEPI-1 and positive clones were selected using 500 μg ml^{-1} G418 added 48 hours after transfection and continued for 3 weeks. Single cell clones were obtained by seeding cells into 96 well plates 48 hours after transfection; after 3 weeks, wells containing a single colony were harvested and the clones amplified. (a) – (e) Are images taken at 2, 4, 6, 8, 10 months post transfection respectively. Green = EGFP.
(f) 0.8% agarose gel of digested pEPI-1 recovered from *E. coli* as a result of vector rescue after 6 months of non-selective culture. Lane 1, 1 kb ladder. Lane 2, original pEPI-1 plasmid digested with *Eco*RI produces a fragment of 6.7 kb. Lane 3, original pEPI-1 plasmid digested with *Eco*RI and *Bgl*II produces 2 fragments, the 4.7 kb vector backbone and the 2 kb S/MAR. Lane 4, recovered plasmid linearised with *Eco*RI produces a fragment of 6.7 kb. Lane 5, recovered plasmid digested with *Eco*RI and *Bgl*II produces 2 fragments, the 4.7 kb vector backbone and the 2 kb S/MAR.

extracted DNA from clone HPA1 at 6 months posttransfection and indicates that pEPI-1 remained extrachromosomal in HPA1 since the DNA recovered by Hirt extraction has the same restriction pattern as the original, untransfected, vector. In fact, the plasmid DNA isolated from *E. coli* can be successfully used to retransfect HeLa cells. Transfected cells still undergo the 'crisis period' when cultured in the presence of G418 and stable clone formation is still a rare event. This provides evidence that the recovered vector has not been genetically altered in any way during the establishment of stable clones and that the factors involved are epigenetic.

3.2.2 Stable clone formation is an inefficient event

FACS analysis of HeLa cells 48 hours after transfection with pEPI-1 showed the transfection efficiency to be reliably between 30 and 45% (Table 3.1). Stable long-term clones were produced using G418 to select for those cells containing the vector. Given the high transfection rate, as established by FACS analysis, it might be assumed that of the 10,000 cells initially seeded at least a third would contain copies of pEPI-1 and therefore survive drug selection. However, it was observed that the number of colonies, and therefore stable clones, produced after a 3 week selection period was much lower. The efficiency of stable clone formation was calculated with this unmodified vector and standard transfection protocol so that it would be clear if any alterations to the plasmid or the protocol subsequently improved this efficiency. HeLa cells were transfected with pEPI-1 using PolyFect and cultured for 48 hours then a sample was prepared for FACS analysis to determine the transfection efficiency; this was done by measuring the number of cells expressing EGFP. On each occasion 10,000 cells were seeded and grown in media supplemented with G418 for 3 weeks. After this time the number of discrete colonies in each flask was counted. Any colonies significantly smaller than the majority were discounted as it was likely they were a secondary colony produced by a mitotic cell becoming detached from a primary colony during media changes. Table 3.1 shows that the mean efficiency of stable clone formation of 6 independent transfection and selection events is 2.61% and is therefore an extremely inefficient event.

Transfection number	1	2	3	4	5	6
eGFP positive cells	3500	4100	4200	3700	3200	4500
Number of colonies	81	105	98	115	85	121
Efficiency of stable clone formation	2.31	2.56	2.33	3.11	2.66	2.69

Mean efficiency = 2.61% ±0.3

Table 3.1. A low percentage of positively transfected HeLa cells are able to establish stable colonies. FACS analysis was used to determine the transfection efficiency 48 hours after transfection with pEPI-1. 10,000 cells were cultured in medium supplemented with 500 $\mu g\ ml^{-1}$ G418 for 3 weeks. Colonies were counted after this selection period. The number of $EGFP^+$ cells seeded was estimated from the transfection efficiency (e.g. if 10,000 cells were seeded at a transfection efficiency of 35% it was estimated that 3,500 $pEPI^+$ cells had been present at the onset of drug selection). It was assumed that each colony obtained had originated from a single $EGFP^+$ cell.

3.2.3 Stable clone formation is not dependent on the timing of G418 selection pressure

With HeLa, less than 3% of pEPI-1 transfected cells are able to make the transition from a transient state to a stable long-term state. Increasing the efficiency of this process would be an ideal starting point for improving the vector. HeLa cells were transfected with pEPI-1 and grown for 48 hours. A sample was prepared for FACS analysis to determine transfection efficiency; 10,000 cells were seeded and grown in medium supplemented with G418 and 10,000 cells were grown in drug-free medium. Care was taken not to allow the cells to become more than 80% confluent to reduce the possibility that $pEPI^-$ cells would have a selective advantage and outgrow the $pEPI^+$ cells. The drug-free cells were passaged 2 days later, 4 days after transfection, and 10,000 were transferred to medium containing G418, while 10,000 were again grown in drug-free medium. This was repeated so that G418 was added to cells 2, 4, 6, 10 and 14 days after transfection with pEPI-1. In each case the cells were cultured for 3 weeks and then the number of colonies was counted. Figure 3.3 (a) shows that the mean efficiency of stable clone formation, resulting from 3 independent transfection-selection events is not altered by a delay in the addition of drug selection. CLSM analysis revealed no difference in the appearance of the cells produced with a delay in drug selection compared with those produced when G418 was added at the normal time (Figure 3.3 (b-f)) suggesting that all the stable populations formed were the same and 'normal'.

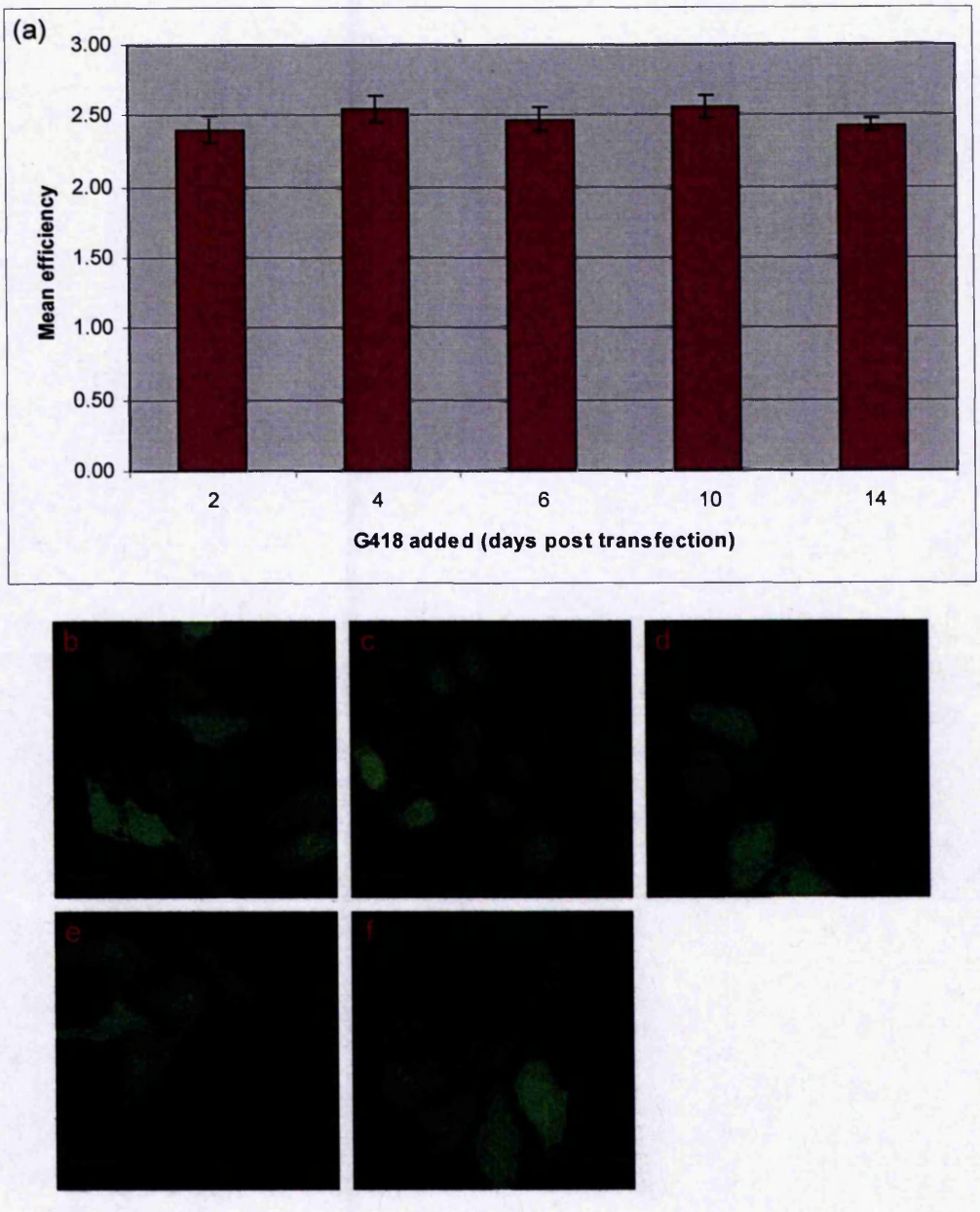
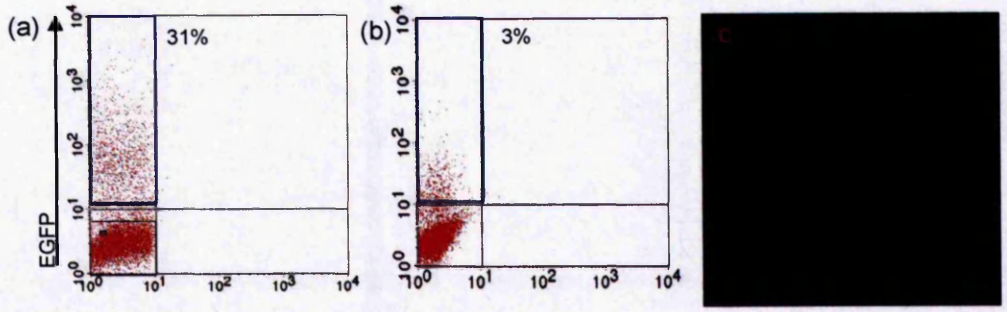


Figure 3.3. The timing of G418 selection does not affect the efficiency of stable clone formation. (a) FACS analysis was used to determine the transfection efficiency 48 hours after transfection with pEPI-1. 500 μg ml^{-1} G418 was added to the culture medium of 10,000 cells at 2, 4, 6, 10 and 14 days following transfection. Colonies were counted after 3 weeks of selection pressure in each case. The efficiency at each time point is the mean of 3 independent transfections (and therefore selections) and variations are represented by the standard error of the mean. (b-f) CLSM images of the stable populations generated by addition of G418 selection at 2, 4, 6, 10 and 14 days posttransfection. Green = EGFP.

3.2.4 G418 drug selection is not necessary for stable clone formation

The observation that delaying drug selection by over 2 weeks appears to have no effect on the percentage of cells able to form stable clones suggests that the presence of G418 is neither detrimental nor actually required for the evolution from a transient to stable state. It was decided to test whether selective pressure to maintain a transcriptionally active plasmid is ever required in the establishment of stable populations. To test this hypothesis, FACSorting was utilised to exclude $EGFP^-$ cells following transfection with pEPI-1, eliminating the need for drug selection and any effect it may have on the establishment of stable populations. HeLa cells were transfected with pEPI-1 and grown for 48 hours as previously described then $EGFP^+$ cells were sorted using the FACS. Positive cells were cultured in drug-free medium for 2 weeks then $EGFP^+$ cells were re-sorted and again cultured in drug-free medium for 2-4 weeks before analysing the resulting populations using the CLSM. Figure 3.4 (c) shows that this population of cells produced without drug selection are expressing EGFP (this population is representative of independent experiments, where n=3). This indicates that a stable long-term population positive for EGFP expression can be established in HeLa cells without the use of selection pressure. It was noticed that the population produced by FACSorting appeared to be much more homogeneous for EGFP expression than any of those produced using G418 selection; intensity analysis of these cells showed they had a mean AU of 35±14 compared with 74±38 for the cells in Figure 3.3 (b) produced by G418 selection. The mean AU and the highest value (corresponding to the brightest cell in the field) is lower for FACSorted cells, but most interesting is the greatly reduced standard deviation indicating that the levels of EGFP expression vary less from cell to cell in the population produced by FACS than those produced by G418 selection. This raises the possibility that even though stable pEPI-1 clones can be established without the addition of G418, the selective pressure creates a state that is permissive for high EGFP expression.



(d) Cell Number	Cell Intensity (AU)	
	No Drug	Drug
1	16	135
2	19	123
3	22	96
4	45	104
5	49	26
6	53	53
7	23	62
8	35	68
9	47	31
10	44	46
Mean Intensity (AU)	35±14	74±38
Range	37	109

Figure 3.4. G418 drug selection is not necessary for stable clone formation. (a) & (b) FACSorting of HeLa cells transfected with pEPI-1 gated for EGFP, the blue box indicates the region sorted. Cells were cultured in the absence of drug selection. (a) HeLa cells 48 hours post transfection (b) HeLa cells 16 days post transfection. (c) LSM image representative of a stable population produced from the sorts. (d) Intensity measurements of drug treated and drug-free EGFP expressing cells. These HeLa cells were grown in drug-free media for 3 weeks after the second sort. Green = EGFP.

3.2.5 The efficiency of stable clone formation is not significantly affected by the timing of pEPI-1 delivery

The above findings suggest that a factor other than drug selection may influence the ability of a cell to establish stable clone status. The correct chromatin structure and nuclear localisation of each pEPI-1 vector is likely to be essential for its maintenance as an episome. It is possible that the introduction of plasmid at a specific time in the cell cycle would facilitate the achievement of the appropriate chromatin structure and/ or nuclear localisation. To test whether plasmid delivery at a specific phase in the cell cycle (early S, mitosis, or early G1) affects the efficiency of stable clone formation, HeLa cells were

synchronised at the G1-S phase boundary using 2.5 μg ml^{-1} aphidicolin for 12 hours, released, then synchronised again to maximise the number of cells in synchrony (2.2.7). Cells were transfected with pEPI-1 DNA 1 hour after release for the early S phase time point. A mitotic shake-off was performed 12-14 hours after release and half of those cells were transfected to give the mitosis time point. The remaining cells recovered from the mitotic shake-off were cultured for 2½ hours prior to transfection resulting in the early G1 time point. FACS analysis was used 48 hours posttransfection on a portion of each of the samples to determine the transfection efficiency. Viable cell number was calculated using a haemocytometer and trypan blue and a known number of cells for each fraction were cultured in G418, as standard, for 3 weeks. The number of colonies obtained for each transfection at the end of this selection period was recorded.

Transfection of cells during early S phase results in a slight reduction in the number of cells able to achieve long-term maintenance of pPEI-1 compared with transfection of cells during mitosis (Figure 3.5). However, the variation in the number of stable clones produced following transfection at the different

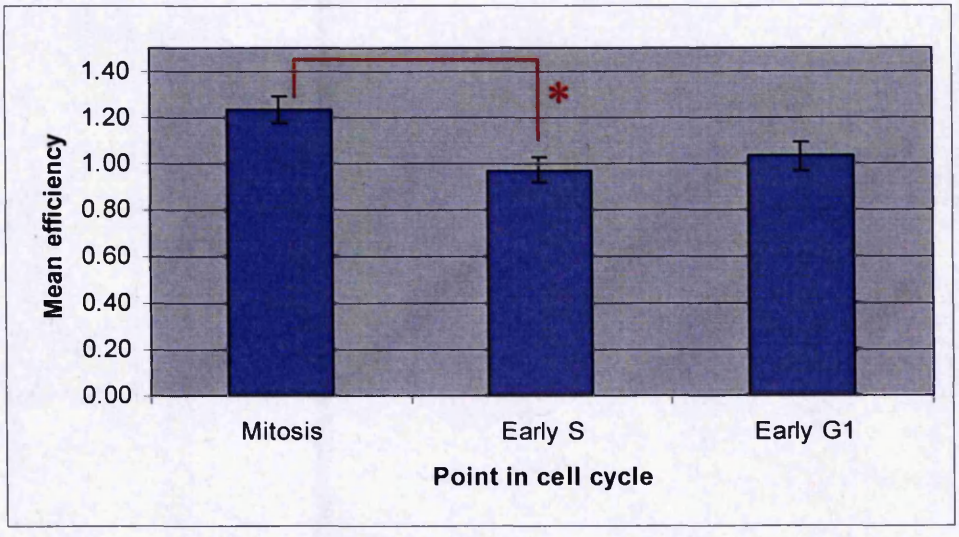


Figure 3.5. The efficiency of stable clone formation is not significantly affected by the timing of pEPI-1 delivery. HeLa cells were synchronised using two rounds of 2.5 μg ml^{-1} aphidicolin to achieve ~90% synchronisation and transfected with pEPI-1 at intervals after release to correspond to mitosis, early S phase and early G1. Cells were cultured for 48 hours then FACS analysis was used to determine the transfection efficiency. Cells were cultured in media supplemented with 500 μg ml^{-1} G418 for 3 weeks prior to colony counting. The efficiency in each case is the mean of 3 independent transfections and variations are represented by the standard error of the mean.

phases in the cell cycle is small making it unlikely that the mechanism for switching between a transient and stably transfected vector state is dependent on the timing of pEPI-1 delivery. The efficiency of stable clone formation is below the expected figure of ~2.5% and this is likely due to the process of cell synchronisation adding extra stress to the cells.

3.2.6 Plasmid input number does not influence the probability of stable clone formation

Considering the modifications in transfection and selection protocols described above do not affect the efficiency with which HeLa cells transfected with pEPI-1 are able to convert from a transient to a stable state, it is possible that this event is a purely random event. Each plasmid may have a fixed probability of attaining the required epigenetic modifications or nuclear localisation required for the long-term expression and replication as an episome due to the stochastic nature of the events involved. If this were the case then those cells receiving more copies of pEPI-1 during transfection would be more likely to form stable colonies. To investigate this, HeLa cells were transfected with pEPI-1 and sorted after 48 hours into 4 fractions according to EGFP expression – high, medium, low and negative (Figure 3.6 (a)). The subpopulations were cultured for 3 weeks in media containing G418; the resulting colonies were counted and used to calculate the efficiency of stable clone formation. Figure 3.6 (b) shows that the efficiency of stable clone formation is similar for the medium and low populations (2.20% ±0.2 and 2.17% ±0.2 respectively). No colonies were obtained from the $EGFP^-$ population. The slight bias against the high fraction (1.83% ±0.1) could be explained by the fact that the very high levels of EGFP produced would probably be cytotoxic. Thus, increasing plasmid input number, to an upper limit, does not increase the chance of a cell progressing to form a stable colony. Interestingly, the stable populations generated from each of the initial 3 $EGFP^+$ subpopulations appear phenotypically similar to each other and to populations generated from other heterogeneous, unsorted transfections. Those colonies produced from cells that received many copies (the 'high' subpopulation) do not have significantly higher levels of EGFP expression then

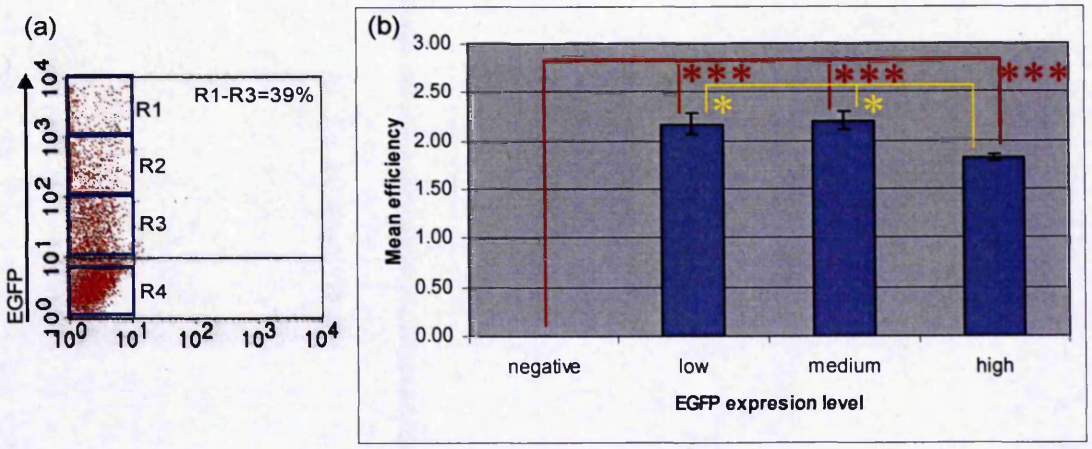


Figure 3.6. Effect of plasmid number on the efficiency of stable clone formation. (a) FACS plot of a sort based on EGFP expression 48 hours post transfection of HeLa cells with pEPI-1. R1-4 correspond to high, medium, low and negative EGFP expression levels. (b) Shows the efficiency of stable clone formation for each population. Cells were cultured in media supplemented with 500 μg ml^{-1} G418 for 3 weeks prior to colony counting. The efficiency in each case is the mean of 3 independent transfections and variations are represented by the standard error of the mean.

those colonies generated from fewer initial vector copies (the 'low' subpopulation) suggesting that the stable colonies established from each of the 3 subpopulations have a similar episome copy number. The efficiency of stable clone formation is slightly reduced in all 3 positive fractions compared with the 'standard' efficiency of 2.61% established previously (see Table 3.1). This reduction in overall efficiency is likely due to the stress exerted on the cells during sorting. As the cells in each fraction were subjected to the same conditions, comparison between fractions should still be valid.

3.3 Discussion

3.3.1 The vector pEPI-1 is maintained episomally in HeLa cells in long-term culture

At the time when these investigations were initiated, there were no published data on the ability of pEPI-1 to establish and maintain episomal status in HeLa cells. The future use of subsequent generations of this prototype gene therapy, by definition, would require use in human cells. The results here demonstrate that pEPI-1 behaves in HeLa cells in a similar manner to published observations using CHO cells (Piechaczek *et al.* 1999). HeLa cells transfected with pEPI-1 DNA and selected using G418 for 3 weeks show maintenance of EGFP expression in long-term culture without the use of further drug selection. Figure 3.2 shows CLSM images of stable clone B4, cultured in drug-free medium, taken at various intervals over a period of 10 months, or ~100 passages. B4 is representative of observations gathered for 6 clones cultured for extended periods of time in non-selective media. The 6 clones monitored were from 3 independent transfection and selection events (2 from each). The similarity in the levels of EGFP expression observed at each time point is striking and indicates that transcription of this reporter gene is maintained at a stable level without the need for selection pressure in long-term cultures.

To exclude the possibility that the continuation of EGFP expression is due to integrated vector rather than persistence of episomes, Hirt extracts were performed to recover extrachromosomal DNA from these clones at various time points during long-term culture. The ability of these extracts to transform DH5α *E. coli* provides evidence for the recovery of episomes from these cultures. It was possible that the stable clones generated contained inert, residual copies from the initial transfection and EGFP expression even though there was a replication defect that would eventually lead to cells without the episome. To exclude this possibility, the Hirt extracts were incubated with *Dpn*I prior to retransformation of *E. coli*. *Dpn*I specifically cleaves at G^{m}ATC motifs which are generated in dam^{+} *E. coli* such as DH5α, therefore unreplicated, inert input plasmid would be linearised and unable to successfully retransform the bacteria. The ability of the Hirt extracts to successfully transform *E. coli*

demonstrates the episomes recovered are resistant to *Dpn*I digestion indicating they have lost their prokaryotic 6-methylcytosine and so have been replicated in these cultures. Digestion of the original prepared vector with *Dpn*I shows it is linearised as expected and this preparation is unable to retransform DH5α cells (data not shown). These results provide evidence that pEPI-1 is maintained episomally in HeLa cells in long-term culture and is not subject to silencing as evidenced by the continuing EGFP expression. Publications have since confirmed the persistence of pEPI-1 as an episome in long-term cultures in cells of human origin (HeLa cells Schaarschmidt *et al.* 2004; HaCat, Jenke *et al.* 2004a; and K562 cells Papapetrou *et al.* 2006).

Unfortunately, observations from other members of the group suggest that a small percentage of established clones experience attenuation in EGFP expression levels in long-term culture implying that stable transgene expression is not efficiently maintained in 100% of the stable clones produced. Reintroduction of 500 μg ml^{-1} G418 reverses this gene inactivation and EGFP expression is readily detected again in only a couple of days. Continuously culturing in cells in 250 μg ml^{-1} has been proposed as this should prevent this downregulation of EGFP expression from occurring. However, the continuous presence of selective pressure would mean that some studies into the behaviour of derivatives of the episome may not be representative of the actual behaviour in unselected cells. This represents a serious problem for the future prospects of this system as a first generation gene therapy vector; one of the advantages of this vector system over others currently in development was the steady, persistent level of transgene expression without the need for ongoing selection pressure. On the other hand, the use of this approach offers an excellent opportunity to study a minimal system and could answer some fundamental questions about the complex interplay between transcription, replication, nuclear localisation and chromatin modifications.

3.3.2 Stable clone formation is an inefficient event

It was essential to definitively ascertain the efficiency of stable clone formation, to determine whether this figure was reliably reproducible and to document the

variation between independent transfection-selection events in order to be clear whether (a) improvements to either the transfection or the selection protocols affected this efficiency in a meaningful way or (b) modifications to the pEPI-1 vector affected this process in a positive or negative way. The insertion or deletion of DNA to the original pEPI-1 might have unpredictable effects on what appears to be a delicately balanced vector arrangement. It has been shown that preventing transcription from an upstream gene from running into the S/MAR can ablate the episomal behaviour (Stehle *et al.* 2003), while deleting the SV40 *ori* appears to have no effect on autonomous replication at all (Jenke *et al.* 2004b). However, these studies did not report on the efficiency of stable clone establishment, only on the ability of these modified constructs to be stably maintained as episomes once the selection process has been completed. It may be interesting to investigate whether any of the modified clones, for example pMARS (Jenke *et al.* 2004b) lacking the SV40 *ori*, are able to establish stable clones with a greater or lesser efficiency than the original vector. It is clear that until we more fully comprehend the mechanisms involved in the transition of pEPI-1 from a transiently transfected plasmid to a stably maintained episome then we must be clear what effects any alteration in the arrangement of elements in pEPI-1 might have as they may have an unanticipated impact on this process.

The best characterised mammalian episomal replicon, *ori*P containing plasmids, establishes stable clones with a low efficiency. The efficiency of 2.61% (Table 3.1) observed for the establishment of pEPI-1 stable clones in HeLa cells, therefore, is not unexpected. The apparent similarities in the method of maintenance (nuclear matrix association during interphase and an association with mitotic chromosomes during cell division) of pEPI-1 episomes and *ori*P replicons suggested that the efficiencies of stable clone establishment would be comparable. The figure 2.61% was the mean of 6 independent transfections, the standard deviation (±0.3) is small and suggests that this observed efficiency is due to factors affecting the system in a predicable, or at least quantifiable manner.

If the mechanistic comparisons between pEPI-1 and *ori*P plasmids are extended further we might hypothesise that the establishment of pEPI-1 stable clones also requires an (unidentified) inefficient epigenetic event. The restriction

of recovered and amplified extrachromosomal DNA from established pEPI-1 HeLa cells with either *Bam*HI or *Bgl*II and *Eco*RI together, reproduces the corresponding restriction pattern of original pEPI-1 plasmid (Figure 3.2 (f)) demonstrating that stable episomes have not been genetically altered to promote their maintenance. Further confirmation of the unaltered genetic state of recovered pEPI-1 vectors is shown by the inefficiency with which they are able to retransfect HeLa cells. These cells have to survive the same 'crisis point' and have a similarly low efficiency of stable clone formation (data not shown) as transfection-selections using the original pEPI-1 plasmid. Epigenetic modifications of pEPI-1 such as an association with hyperacetylated histones would be lost during the extraction and amplification processes and so purified pEPI-1 plasmid from the retransfected *E. coli* would be epigenetically the same as the original vector thereby negating any positive impact these modifications might have on the efficiency of stable clone formation.

An interesting way to address this issue of what precisely is the epigenetic status conducive for stable clone establishment would be to form chromatin with the pEPI-1 DNA *in vitro*, possibly by incubation with *Xenopus laevis* oocyte nuclear extract (Sessa & Ruberti, 1990), prior to transfection. Formation of 'active' chromatin on the vector may assist in survival of the establishment phase. Alternatively, if it were possible to extract intact episomes from stable clones in large enough quantities, these could be used to retransfect HeLa cells. If the correct 'code' of epigenetic modifications is the key requirement for establishing stable clones then it would be expected that these recovered plasmids show virtually one hundred percent efficiency.

Transfected *ori*P plasmids have the ability to express transgenes and have been shown to replicate during the transient phase (Leight & Sugden, 2001b) but replicants are subsequently lost at more than 25% per cell generation. The amount of replicated plasmid detected at 14 days posttransfection were less than 1% of the level detected 4 days posttransfection indicating that the critical point in this system is the mitotic stability. Inefficient segregation of plasmids at mitosis leading to an unstable copy number and the risk of cytoplasmic degradation of those vectors not included in the reformed nucleus would explain the rapid loss of vectors during this period. How far does the similarity between pEPI-1 and *ori*P plasmids extend? It would be interesting to investigate whether

transfected pEPI-1 plasmids are also replicated efficiently in the transient phase. If the critical step is the formation of a stable association between matrix and vector then actions could be taken to promote this. One avenue to explore could be the cotransfection of pEPI-1 pre-complexed with SAF-A. SAF-A has been shown to organise S/MAR DNA from the *topoisomerase I* gene into loops *in vitro* (Romig *et al.* 1992), if this could be achieved with pEPI-1 DNA then this complex could even be directly microinjected into the nucleus, bypassing any nuclear import issues.

3.3.3 Stable clone formation is not dependent on the timing of G418 selection pressure

G418 selection pressure is employed in the generation of a population of pEPI-1$^+$ stable HeLa cells. It is normally added 48 hours after transfection and continued for 3 weeks to eliminate any cells not possessing functional copies of the vector. It was hypothesised that the inefficiency of stable clone formation may be due to TI of the *Kan/Neo^r* gene region with the pCMV and *EGFP* gene. It has been shown that active transcription upstream of and running into the S/MAR is an absolute requirement for the establishment of stable episomes (Stehle *et al.* 2003). If the presence of G418 in the culture medium is driving very high transcription of the *Kan/Neo^r* gene to provide resistance to the drug, then this may disrupt the formation or initiation of transcription complexes at the pCMV and result in a lower frequency of gene activation during the crisis period. If this region is not actively transcribed during a critical period then an alteration in topological status will not be conferred to the S/MAR region and no matrix association will be established (with obvious consequences). The addition of selection pressure may alter the pattern of gene expression from the plasmid and downregulate EGFP expression in favour of the *Kan/Neo^r* gene and addition at 48 hours posttransfection may be within this critical period. Delaying the addition of G418 to transfected cells might allow the establishment of the episomal state before a high level of transcription from the (possibly interfering) *Kan/Neo^r* gene region. Addition of G418 has no adverse effects on stably established populations so it is unlikely that the delayed addition of G418 would

disrupt any S/MAR-matrix associations already formed in the newly transfected cells.

G418 was added to HeLa cells at 2, 4, 6, 10 and 14 days posttransfection with pEPI-1 DNA. The cells were selected for 3 weeks as normal and the number of stable colonies generated was recorded. Figure 3.3 (a) shows the mean efficiency of stable clone formation (n=3) for drug added at each of the time points. There is no statistical difference in the numbers demonstrating that a delay in the addition of selective pressure to these cells does not increase the efficiency. By extension this also suggests that the *Kan/Neo*^{r} gene region is not interfering with transcription from the pCMV in to the S/MAR and that the TI hypothesis is null.

It is possible that even though TI appears not to be at work here, there may be some other form of interference. The efficiency of *ori*P replicon establishment appears to have been improved by making alterations to the plasmid (Leight & Sugden, 2001a). *Ori*P is composed of two *cis*-acting elements, the family of repeats (FR) and the dyad symmetry element (DS). The FR contains 20 imperfect repeats for EBNA-1 dimers and the DS possesses 4 binding sites of lower affinity and is the site near to the origin of replication. It was shown that plasmids containing the DS on its own supported replication as efficiently as *ori*P (DS+FR) plasmids 2-4 days posttransfection in HeLa and 293 cells. The DS plasmids were stable for 2 weeks posttransfection in 293 cells and present at >100 copies per cell (compared with ~2 per cell for *ori*P plasmids) indicating that these DS replicons are established efficiently in some cell lines with the ability to be replicated and partitioned correctly. It was suggested that the FR might be inhibitory to the epigenetic event required for establishment of the replicon possibly by promoting intramolecular interactions between EBNA-1 dimers bound to FR and DS rather than intermolecular associations between bound EBNA-1 and mitotic chromosomes. There doesn't appear to be any direct analogies to be drawn with respect to pEPI-1; there is only one matrix attachment region in pEPI-1 and it is unlikely it forms intramolecular interactions with other elements elsewhere in the plasmid so it would be difficult to suggest what should be removed to improve the establishment efficiency. However, the improvement observed for DS plasmids reveals there are elements within a

replicon that can inhibit as well as promote the efficiency of stable clone formation and this might be true for pEPI-1 as well.

3.3.4 G418 drug selection is not necessary for stable clone formation

The observation that drug selection could be delayed to up to 14 days posttransfection and yet stable clones were still generated posed the question of the requirement of drug selection in the process of stable clone formation. It could be expected that all the transiently transfected plasmids would have been lost from the population by 14 days posttransfection. The addition of G418 at this point was only necessary to eliminate pEPI-1$^-$ HeLa cells rather than drive transcription from the *Kan/Neo*^{r} gene which would somehow contribute to the establishment of the episomal status. If the function of drug selection was purely as a means of eradicating cells not containing stably established episomes then it follows that another method of removing negative cells would also produce stable populations. Transfected cells were grown for 48 hours to reach maximum EGFP expression then sorted by FACS to separate the positive cells from the untransfected cells. The EGFP$^+$ cells were grown in drug-free medium for two weeks then resorted. There was a massive loss of EGFP$^+$ cells during this two week period so that only ~5% remained positive by the 2nd sort. This is consistent with observations during this time period of establishment of an *ori*P replicon where more than 25% of plasmids are lost each cell generation for two weeks (Leight & Sugden, 2001b). The resorted EGFP$^+$ HeLa were cultured in drug-free medium for a further 2-4 weeks before fixation and CLSM analysis. Vector rescue validated the presence of pEPI-1 as an episome in these populations. Figure 3.4 (c) demonstrates that pEPI-1 is able to establish stable EGFP expression episomally in HeLa cells in the absence of selection pressure. Comparison of the CLSM images in Figure 3.4 (c) with those of a population produced by G418 (Figure 3.3 (b-f)) selection suggests there may be differences in EGFP expression levels in the clones produced with no selection. There are no high EGFP expressing cells produced by FACS sorting. The whole population has a much lower level of expression when compared with a population produced by G418 selection; a mean of 35 AU for FACSorted cells

and 74 AU for G418 cells. This indicates that the presence of selective pressure at some stage in the establishment process may be required for high level EGFP expression. Alternatively, the presence of a second transcription unit may be sufficient to promote high level EGFP expression if it is the process of transcription that enhances EGFP expression rather than the introduction of selection pressure. This issue could be readily resolved by exchanging the *Kan/Neo*^{r} gene for another reporter gene and monitoring the expression levels of both prior to and following the establishment of a stable population.

These results appear on the surface to be in disagreement with observations recently published (Nehlsen, Broll & Bode, 2006) regarding the establishment of a pEPI-1 derivative, the 'maxicircle', and a 'minicircle' version of pEPI-1. This study investigated the affect of bacterial sequences in pEPI-1 on the establishment and maintenance of episomal behaviour. The introduction of FRT sites between the pCMV and the *EGFP* gene and between the SV40 poly A site (the 'maxicircle'), produced a 'minicircle' version of pEPI-1 when Flp-mediated recombination was induced between the FRT sites. This investigation also used FACSorting to separate EGFP$^+$ and EGFP$^-$ cells following transfection with either the mini- or maxi-circle as the removal of the *Kan/Neo*^{r} eliminated the use of G418 as a means of selection. Nehlsen *et al.* demonstrated that transfection of CHO, HEK293 or NIH3T3 cells with the minicircle followed by FACSorting was able to produce stable, EGFP expressing episomes in culture without any selection pressure. The EGFP expression was stable for more than 50 population doublings (PDs, 24 hours for CHO and HEK293 cells) as monitored by FACS analysis. The same sets of experiments were also performed on the maxicircle to assess the ability of this construct to establish and maintain EGFP expression as an episome. CHO cells were transfected with either the mini- or the maxi-circle and grown for 5 PDs (in non-selective media) "to allow for the establishment of episomes" before recovering EGFP$^+$ cells by FACSorting. The cells were examined 12 days later and it was found that 70% of the minicircle but only 10% of the maxicircle sorted cells were expressing EGFP. The cells were then incubated in drug-free media for a further 40 PDs before FACS analysing them again. Minicircle cells still maintained EGFP expression in over 60% of the cells but in none of the maxicircle cells, in fact, the maxicircle itself was not detected in Southern blots performed after 26 PDs indicating that it had

been lost from the population. It was claimed that incubation of the maxicircle in G418 prevented this loss of expression and so it was concluded that G418 was required for long-term reporter gene expression from the episomal construct containing the bacterial sequence elements (the maxicircle) but not from the wholly eukaryotic version (the minicircle). These data seem to disagree with my findings that pEPI-1 was established as an episome and actively transcribed EGFP in HeLa cells as a result of FACSorting and without the need for G418 selection pressure. However, their results using FACSorting of the maxicircle in HEK293 and NIH3T3 cells paint a much muddier picture. NIH3T3 cells showed a complete loss of EGFP expression in the population by 32 PDs but the HEK293 cells appeared to have comparable EGFP expression to the minicircle after 32 PDs. It is likely the difference in cell types used in the two investigations is a major factor any possible disparity of findings since there appears to be 3 different outcomes in the Nehlsen study using the maxicircle for FACSorting for the 3 cell types used.

Aside from cell-specific behaviour with regard to EGFP expression from pEPI-1 episomes, there are also some experimental differences that might explain a difference in outcome. This investigation sorted cells 48 hours posttransfection as would be standard selection behaviour if G418 was employed. Cells were allowed to progress through the 'crisis phase' for 2 weeks before resorting cells to remove those cells that had not retained pEPI-1 as episomes. The loss of $EGFP^+$ cells during the 2 week period (~95% of cells) between the 1^{st} and 2^{nd} sorts would agree with the massive loss of *ori*P^+ cells over the same 2 week establishment phase (Leight & Sugden, 2001b) and so was not seen as an indication of loss of EGFP expression in episome containing cells. The fact that there was no more appreciable loss in EGFP expression in the 2-4 weeks following the second sort indicated that stable EGFP expressing episomes were established in the culture. In contrast, Nehlsen *et al.* only did one sort, 5 days posttransfection and it is possible that the loss of EGFP expression seen 12 days later when examined by FACS is due to the fact that stable replicons had not been established by day 5 and so unsuccessful maxicircle plasmids were lost over this period. Strikingly, examination of the percentage of CHO cells still $EGFP^+$ 12 PDs after transfection shows a value of 3% of the cells sorted after 5 PDs (which were all $EGFP^+$). This figure of 3% would be in absolute agreement

with the efficiency of stable clone formation calculated in 3.2.2 for HeLa cells transfected with pEPI-1. The finding by Southern analysis that all the maxicircle vectors had been lost by 26 PDs in CHO cells may not be entirely the case. The maxicircle is maintained at a low copy number and if it is only in a very low proportion of the cells (only present in 3% of cells after only 12 PDs) used for the blot; the technique would not be sensitive enough to detect such small quantities of episomal DNA. The results obtained by treatment of the sorted maxicircle cells at day 5 with G418 can also be explained by this theory. The G418 would eliminate any of those cells that had received maxicircle plasmid (and were therefore $EGFP^+$ at PD=5) that was unable to establish episomal status over the crisis phase and thus boost the overall proportion of EGFP expressing cells in the population that was examined by FACS at day 12. This explanation of events would agree with observations from the well characterised *ori*P episomal system; at day 14 posttransfection, only 1% of the *ori*P plasmids detected at day 4 remain in cells establishing *ori*P replicons - they are lost during the intervening days as they are unable to establish a stable replicon. However, if this explanation of the data from Nehlsen *et al.* is sound then it produces an exciting result. The loss of EGFP expression and by extension, loss of maxicircle, from transfected CHO cells between the sort after 5 PDs and analysis after 12 PDs could be taken to show that the efficiency of stable clone formation is 3%. There are 65% of the minicircle cells left in the population that are still expressing EGFP suggesting that the efficiency of stable clone formation for minicircles is 65%. This would be a significant improvement in efficiency and steps should be taken to more fully investigate if this is truly the case.

3.3.5 The efficiency of stable clone formation is not significantly affected by the timing of pEPI-1 delivery

Many published studies have shown that transfection of cells during mitosis can greatly improve the efficiency of transgene expression compared with transfection at other phases of the cell cycle, particularly G1 (for example, Brunner *et al.* 2000; Escriou *et al.* 2001). This investigation sought to address

whether, in addition to any effects that may be seen with respect to the efficiency of transgene expression, the transfection of HeLa cells with pEPI-1 at defined phases of the cell cycle had any impact on the efficiency of stable clone formation. As can be seen in Figure 3.5, there is very little difference in the efficiency of stable clone formation between the different transfection time points. There was a slight improvement in efficiency when cells were transfected during mitosis (1.23% ±0.1) compared with transfection during early S phase (0.97% ±0.1) but there was no statistical significance between mitosis and early G1 (1.03% ±0.1) as would have been predicted if the efficiency of transgene expression had a bearing on the efficiency of stable clone establishment. The observed efficiency of stable clone formation for all three transfection timings is much lower than that documented previously (2.61% in section 3.2.2). There were some technical issues with the methodology that were never fully resolved that could account for the observed drop in efficiency. Control cells that were synchronised and subjected to the same conditions as each of the transfection time points but only mock transfected showed similar levels of mortality over the ensuing 48 hours to the corresponding cells transfected with pEPI-1. An increased incidence of cell death in all samples was assumed to be due to the aphidicolin treatment and the mitotic shake-off. Treatment with nocodazole, a mitotic block, was substituted for the aphidicolin and the procedure modified accordingly but it was found that cells recovered from this treatment even less readily than the aphidicolin (data not shown). The use of PI during the FACS analysis allowed the transfection efficiency to be calculated as a percentage of the viable cells only so that the efficiencies of stable clone formation documented should still be 'real' data as they still rely on the number of colonies obtained from a known number of viable, transfected cells.

The synchronisation of cells prior to transfection might be perturbing their transcription and replication programs so that the nuclear environment is not so permissive for the establishment of stable clones. The use of counterflow centrifugal elutriation to fraction the different cell phases could be a less disruptive means of transfecting cells at specific points in their cell cycle (Brunner *et al.* 2000). HeLa cells transfected with a GFP reporter showed a 170-fold induction of expression (G1 cells compared with late S/G2 cells) when

fractioned by elutriation. Using a thymidine block to synchronise cells only resulted in a 30-fold difference suggesting that the process of synchronising cells alters their response to transgenes. It might be interesting to use counterflow centrifugal elutriation to obtain cells at specific points in the cell cycle prior to pEPI-1 transfection would have any affect on the efficiency of stable clone formation as it would avoid any possible side effects of the treatments necessary for cell synchronisation.

There may be no actual effect of cell cycle phase on the efficiency with which pEPI-1 vectors are able to be established as episomes. It has been reported that pEPI-1 is able to establish stable clones in non-dividing cells as efficiently as in cycling cells (Papapetrou *et al.* 2006). $CD34^+$ enriched cells isolated from umbilical cord blood cultured in serum-free media supplemented with a combination of cytokines for up to 72 hours prior to pEPI-1 transfection had the same transfection efficiency as unstimulated $CD34^+$ cells. This shows an apparent lack of requirement for cells to even be in the cell cycle in order to establish stable episomes. This may explain why there was no improvement in efficiency when HeLa cells were transfected at various phases of the cell cycle – the mechanism is cell cycle independent.

These results show that there is no significant difference in the efficiency of stable clone formation when cells are transfected during specific phases of the cell cycle. Whilst transfections during mitosis may improve the efficiency of transfection of pEPI-1 plasmid shown as an increase in the proportion of $EGFP^+$ cells, or an increase in the amount of EGFP per cell, they do not promote the establishment of a greater number of cells with stable episomes.

3.3.6 Plasmid input number does not influence the probability of stable clone formation

The theory behind this experiment was to investigate whether the probability of a pEPI-1 transfected HeLa cell establishing a stable clone correlates with the input number of vector molecules. Put another way, does each vector molecule have a definable chance (albeit very low) of being established as an episome, so if more vectors are introduced to a cell will there be a greater chance that

this cell will be successful in achieving established clone status? This was addressed by FACSorting transfected cells 48 hours posttransfection into 4 subpopulations corresponding to EGFP intensity (which was used as a measure of vector input number) and then assessing the number of stable colonies generated following drug selection. Sorting cells in this way relies on the assumption that EGFP expression is proportional to vector copy number. This assumption was based on work by Tseng *et al.* (1997) who showed that transiently transfected HeLa cells expressed the reporter gene, EGFP, with 100% efficiency from the input plasmids. However, this assumption may not hold absolutely true for pEPI-1 vectors and it should be considered that there may be different levels of EGFP expression from each vector molecule in this transient transfection investigation.

If there was a correlation between vector input number and stable clone establishment efficiency then there should have been the most colonies obtained from the high subpopulation and the least from the low. This was not the case, in fact, there was a slight bias against HeLa cells with very high initial levels of EGFP (1.83% ±0.1) compared with the medium (2.20% ±0.2) and low (2.17% ±0.2) subpopulations (Figure 3.6 (b)). This suggests that there is some disadvantage to an increased input vector number and this could be for a variety of reasons. The extremely high levels of EGFP produced could have an adverse effect on the long-term health of the cell, perturbing normal cellular process in a way that negatively impacts of the mechanism employed by pEPI-1 vectors to establish episomal status. Huge amounts of exogenous DNA and protein present in the HeLa cells may trigger a defence response that targets the foreign material for destruction or silencing. The presence of bacterial sequences would be ready targets cellular antiviral defences. In addition, high levels of transcription of the transgenes may disrupt the 'normal' transcription pattern which in turn could impact on the replication program (Schwaiger & Schubeler, 2006) of the native genome. This disruption of typical cell behaviour could produce an environment less permissive for the establishment of stable episomes. Additionally, extremely high levels of EGFP would most likely be cytotoxic.

A further observation was that no colonies were obtained from the negative population suggesting that there were no clones generated with episomes that

expressed G418 resistance but not EGFP. It was also noted that the medium and low fractions established stable clones with equal efficiency.

Possibly the most interesting observation was the similarity in appearance of all three of the established populations when analysed using CLSM. There was the same distribution of EGFP expression levels in each of the heterogeneous populations that were generated following G418 selection of each of the 3 FACSorted fractions of pEPI-1 transfected HeLa cells. If the mechanism for establishing stable episome status was purely due to a vector randomly achieving the correct set of epigenetic modifications then if would be expected that in addition to the 'high' fraction having more positive colonies after G418 selection, more vectors per cell would also have made the transition from transient to stable status. It would be expected that there would be a higher average level of EGFP expression in the population generated from the high sort compared with the 'medium' or 'low' as each of these would have fewer episomes per cell. This does not appear to be the case and suggests that there may be a critical upper limit for episome copy number able to be stably maintained by HeLa cells. The odds of each vector establishing episome status are such that there would be enough input vectors in the 'low' fraction able to reach the limit of 6 per cell. Extra vector copies in the 'medium' or 'high' sorts would be unable to be tolerated by the cell and become lost, even if they had the potential to evolve episomal epigenetic modifications.

It would be interesting to discover whether an average of 6 episomes per cell is the limit of the system. The use of a pEPI-1 derivative with a different reporter gene and different antibiotic resistance cassette (e.g. hygromycin or puromycin) could help to answer this. What would be the outcome of transfecting and selecting established pEPI-1$^+$ HeLa cells with the pEPI-1 derivative? Would there be ~6 pEPI-1 and ~6 pEPI-derivative episomes per cell? Would any new episomes be able to be established? Does treatment with the new antibiotic select for the replacement of pEPI-1 episomes with pEPI-derivatives or are they stably matrix associated?

Put another way, if the establishment event was purely stochastic then why do we see ~6 episomes per cell for 3% of the transfected population? Why do we not see 3 episomes per cell in 6%, or 1 episome per cell in 18% of the initial transfected population? The fact that successful cells have similar average

episome copy number indicates that there may be something definably different about a proportion of either the input vector molecules or the cells. What this might be remains a mystery but it leaves the possibility of improving the efficiency of clone formation once it is elucidated.

3.3.7 Conclusions

The attachment of *ori*P plasmids to the nuclear matrix via EBNA-1 and the association of the pEPI-1 S/MAR to SAF-A indicates that the ability to connect to the matrix is integral to the success of these two classes of episomes. Many other episomal constructs such as BPV via E1 and E2 (Lehman & Botchan, 1998), KSHV via LANA (Ballestas, Chatis & Kaye, 1999) and YACs containing S/MARs (Cossons *et al.* 1997) also rely on a matrix association for their preservation. With this is mind it is tempting to speculate that it is the ability of the vector to bind the nuclear matrix that is the rate-limiting step in this process, and that failure to establish this association would result in a failure of the vector to transform from a transient plasmid to a stable episome. The continuous evolution of the chromatin status during this crisis phase would either facilitate or inhibit this process depending on the nature of the modifications by impacting on the level of transcription into the S/MAR and consequently the topological status of the region. Histone hyperacetylation would promote high levels of transcription across the *EGFP* gene and into the S/MAR, possibly providing a favourable topological state for base-unpairing and hence matrix association. It is possible that vectors need to undergo multiple rounds of replication before they gain the correct set of epigenetic modifications to allow a stable maintenance and propagation of this active, matrix-bound state to subsequent cell cycles. Until this happens the vector would be susceptible to any defect in the transcription, replication or segregation process. A reduction in transcription would prevent the formation of the 'episomal' chromatin modifications and instead promote the formation of deacetylated, inactive chromatin (this would reduce the transcription further encouraging the formation of less active chromatin, reducing transcription further and so on). The equilibrium would be massively in favour of the formation of an inactive chromatin state and once the

process of deactivation begins it would be very difficult to reverse it. In addition, transcription is intimately linked to replication programming and a loss of transcriptional activity could lead to a failure to replicate, obviously this would lead to a reduction in copy number each cell generation eventually leading to a complete loss of vector.

Newly introduced vector molecules are able to be transcribed, as evidenced by the production of the reporter gene, EGFP. If one assumes a common mechanism of establishment of *ori*P and pEPI-1 episomes then it could be conjectured that newly introduced pEPI-1 is able to replicate the same as *ori*P plasmids (Leight & Sugden, 2001a). Even though the newly transfected vectors are permissive for transcription and replication, this can only occur if the vector is, *and remains*, in an active nuclear compartment. Without an established S/MAR-matrix association there would be a high probability that these vectors would drift out of the active areas, losing their proximity to transcription and replication factories and all the activating factors (such as HATs) and come into contact with deactivating factors (such as HDACs). Only a minority of vector molecules are able to avoid the pitfalls and acquire the epigenetic modifications permissive for episomal maintenance as reflected in the inefficiency of this process. However, once this stable state is achieved, it appears to be very stable and faithfully reproduced in each new generation.

Chapter Four: Characterisation of Stable pEPI-1$^+$ Clonal HeLa Cells

4.1 Introduction

The use of future generations of pEPI-1 and its derivatives for gene therapy applications would require stable, reliable expression of the transgene(s). Many medical conditions that would benefit from the use of pEPI-1 would necessitate the expression of a therapeutic protein for the lifetime of the patient. In addition to merely the persistence of gene expression, the level of expression would also be important. A defined, steady level of expression that did not fluctuate and was uniform across the population of transformed cells would be highly desirable. Gene expression of transiently transfected vectors is generally proportional to copy number – more vectors in, more gene product out. It was noticed that clonal pEPI-1$^+$ cells appeared to have different levels of EGFP expression even though they should be identical since they all originated from the same mother cell during the establishment phase. This raised the possibility that differences in gene expression of established clones is less a result of a variation in copy number and more a result of epigenetic modification of the episomes. The histone acetylation of the 'epi' domain would be permissive for high levels of gene expression, however, if the episome was mainly histone deacetylated, this would facilitate the formation of a more closed chromatin structure that would be refractory to transcription factor binding and result in much lower levels of reporter gene expression. It was also possible that the variation in gene expression between clonal cells was due to different vector copy number per cell. FISH analysis of stable pEPI-1$^+$ clones revealed that the copy number varied between 4 and 13 in stable CHO cells (Baiker *et al.* 2000). Whether this was also the case in HeLa cells was not clear. In order to better characterise the behaviour of pEPI-1 episomes in stably established HeLa clones it was decided to investigate the variation in EGFP expression and to determine if it is linked to vector copy number or epigenetic control. The aims of this chapter were to investigate the nature of different clonal populations of

stable pEPI-1 HeLa cells with respect to (a) episome number and (b) reporter (EGFP) gene expression.

4.2 Results

4.2.1 Clonal populations of HeLa cells stably transfected with pEPI-1 show 3 distinct phenotypes with respect to the level of EGFP expression

It was decided to investigate the level of EGFP expression in different clones of HeLa cells stably transfected with pEPI-1. Single cell clones were generated by seeding HeLa cells into 96 well plates 48 hours after transfection with pEPI-1 at a density of 2-3 cells per well. Positive cells were selected using G418 and wells containing a colony derived from single cell were identified, the colony recovered and expanded (2.2.6). Single cell clones were cultured in the absence of selection for 2-4 weeks before analysis. Hirt extraction of extrachromosomal DNA followed by vector rescue was employed to confirm the episomal status of vectors in the stable clones (Figure 4.1).

Visualising fixed cells using the CLSM revealed clear differences in the level of EGFP expression both within a single clone and between different clonal populations. To better characterise the distribution EGFP transcription levels between cells from a single clone (theoretically identical) and between cells from different clones, the intensity of EGFP fluorescence was measured. The mean intensity of signal for each clone was calculated and is represented in (Figure 4.2). An evaluation of these clones suggested they broadly fell into 3 distinct groups with respect to EGFP expression levels. The proportion of clones from a single transfection-selection falling into each phenotype was reliably reproduced in each independent transfection-selection event (n=6). Scrutiny of 20 clones stably transfected with pEPI-1 revealed 16 with a fairly uniform 'normal' level of EGFP expression (40-79 AU), 2 possessed a uniform 'low' level of expression (1-39 AU) and 2 clones had a mainly uniform 'medium' level of EGFP expression but also had a small percentage of cells with a high level of EGFP expression (above 80 AU). Clones with uniform 'high' levels of EGFP expression were never seen. These measurements show that the majority of clones obtained (80%), the 'normal' phenotype, have a mean intensity of 40-79 AU.

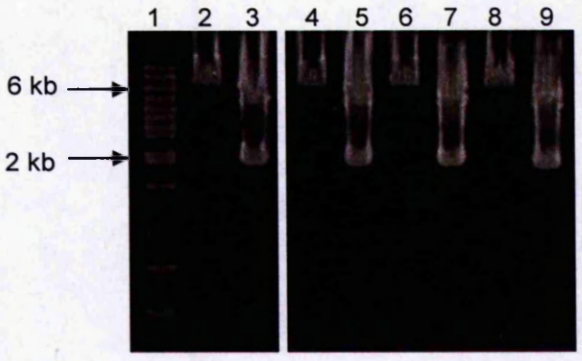


Figure 4.1. Restriction analysis of vector rescue plasmids. Hirt extracts were prepared from stable clones and used to transfect *E. coli*. The plasmid was recovered from discrete colonies of *E.* coli and was subject to restriction analysis and comparison with original, untransfected pEPI-1. Restriction products were separated on a 0.8% agarose gel and visualised by ethidium bromide staining. Lane 1, 1 kb ladder. Lane 2, *Eco*RI digestion of the original plasmid. Lane 3, *Eco*RI and *Bgl*II digestion of the original plasmid. Lanes 4 & 5 as 2 & 3 except with clone B4. Lanes 6 & 7 as 2 & 3 except with clone B7. Lanes 8 & 9 as 2 & 3 except with clone B8. Digestion with these enzymes produces an identical restriction pattern with recovered vector as it does with untransfected plasmid indicating that the Hirt extracted DNA was genetically unaltered pEPI-1 plasmid.

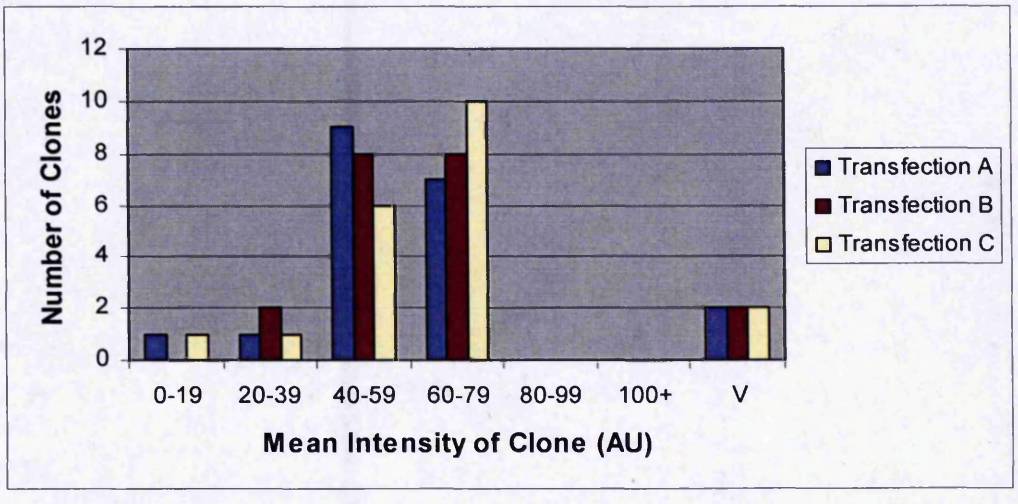


Figure 4.2. Analysis of the intensity of EGFP expression in stable long-term clones. HeLa cells were transfected with pEPI-1 and cultured for 48 hours prior to seeding into 96 well plates. Cells were cultured in media containing 500 μg ml^{-1} G418 for 3 weeks, after which time single colonies were harvested and expanded. Cells remained in media supplemented with G418 for 1 further week. Single cell clones were analysed using the CLSM 6-8 weeks after transfection. 20 single cell clones from 3 independent transfections (A, B and C) were scored by measuring the mean intensity of 10 cells per clone. 'V' corresponds to clones exhibiting a 'variegating' phenotype.

4.2.2 FISH analyses of clones to determine mean vector copy number

To further characterise and explain the nature of the 3 phenotypes observed when stable long-term pEPI-1$^+$ clones are established it was decided to perform FISH analysis. The aim of this was to determine the distribution of episomes in clonal cells and also to discover if the different phenotypes could be explained by the fact that they had different copy numbers. Figure 4.3 shows representative images of 4 different stable clones. Control pEPI-1$^-$ HeLa cells (Figure 4.3 (a)) show no signals whereas the vast majority of nuclei examined from each clone showed clear fluorescent signals. The analysis of 25 nuclei from 6 different clones (2 low, 2 normal and 2 variegating with respect to EGFP intensity) is represented in Table 4.1. Vector copy number ranges from 3 to 11 with standard deviations of ~2. Clones B7 and A11, low EGFP intensity phenotype, have mean copy numbers of 4.80±1.38 and 4.84±1.34 respectively and also have a much narrower range of copy number, 3-7 per cell, in contrast to the normal and variegating clones which range from 3 to 11. This mean copy number of just below 5 is slightly lower than the mean copy number of ~6 observed for the normal and variegating phenotype clones and may partly explain the lower levels of EGFP expression observed in the 'low' phenotype clones.

As was expected, the number of episomes detected in clonal cells is highly variable (3 to 11 copies) in most clones, but, most importantly, there appears to be far less variation in the mean copy number per cell in each of the 3 phenotypes. The standard deviation for the variegating clones is similar to that of the normal suggesting that the huge differences in EGFP intensity between clonal cells is not purely a result of variable copy number within the clone as by that reasoning there should be a percentage of bright cells in the normal clones as well. This indicates that the 3 phenotypes observed are a result of the combination of different vector copy numbers per cell and also epigenetic control.

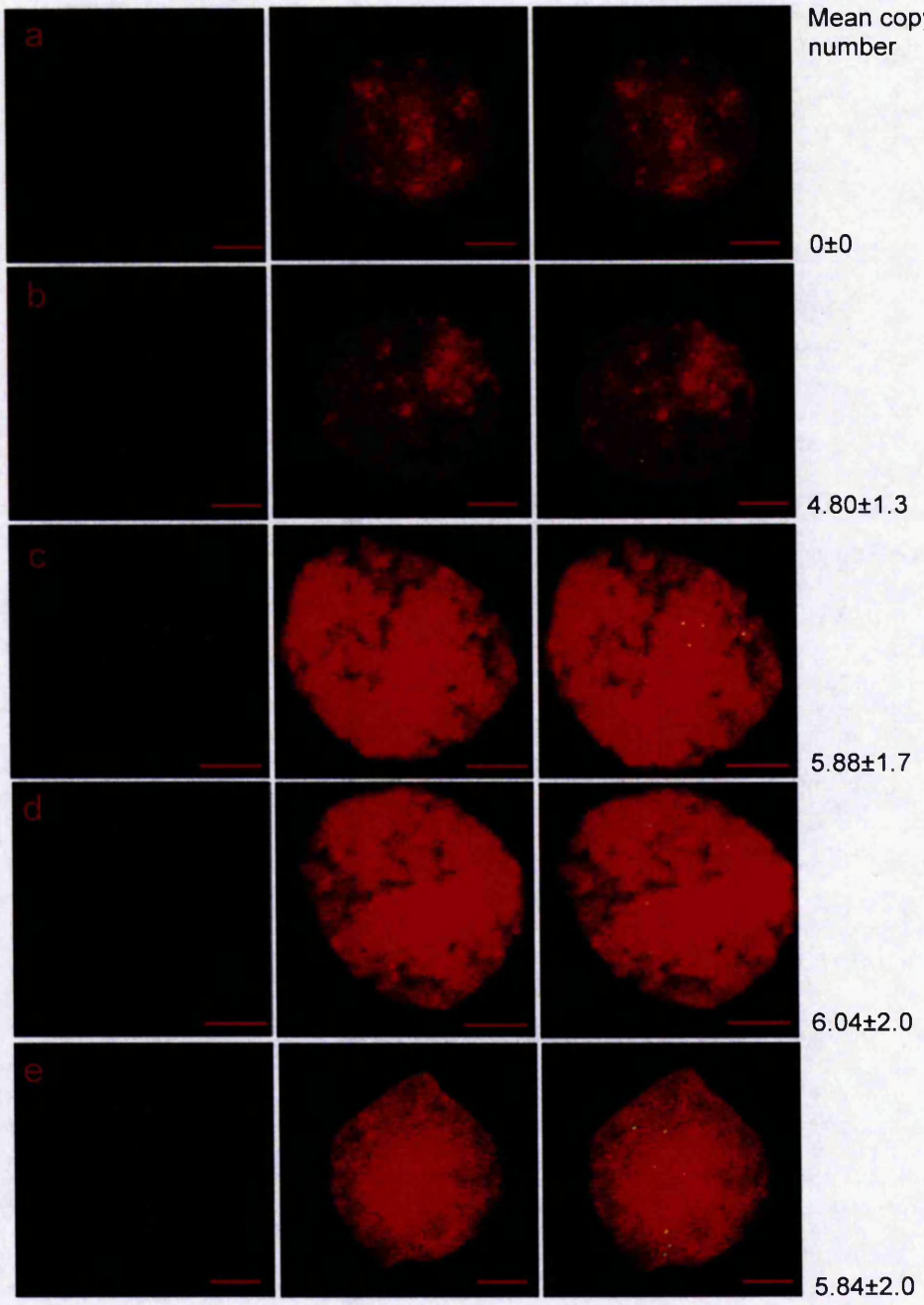


Figure 4.3. Investigating vector copy number in HeLa cells stably transfected with pEPI-1 by FISH analysis. Left to right: Alexa Fluor 488, propidium iodide and merged images. Representative images, mean copy number derived from 25 nuclei. (a) Untransfected HeLa as control shows no signals. (b) Clone B7 (low EGFP expression). (c) Clone B4 (normal EGFP expression). (d) Clone C21 (normal EGFP expression). (e) Clone B8 (variegating EGFP expression). (b-e) show signals corresponding to individual episomes in the stable clones whereas the control untransfected cells have no hybridisation. Scale bar represents 5 μm.

Nucleus	Copy Number					
	Clone B7	Clone A11	Clone B4	Clone C21	Clone B8	Clone C16
1	4	7	6	5	5	4
2	3	5	4	7	7	4
3	4	5	9	4	4	7
4	5	3	7	5	6	6
5	3	4	6	7	5	11
6	6	4	6	10	4	5
7	4	7	4	11	4	5
8	7	5	11	5	4	5
9	7	6	8	5	9	8
10	3	6	5	4	6	6
11	4	5	5	10	11	11
12	4	3	7	6	5	4
13	5	4	4	3	4	7
14	3	4	5	7	5	5
15	5	6	4	5	5	9
16	5	5	6	8	8	7
17	8	8	5	6	7	5
18	5	3	8	6	5	5
19	4	3	5	4	4	4
20	4	4	5	5	5	7
21	5	6	4	4	10	4
22	7	4	6	6	4	4
23	6	5	6	6	9	6
24	4	5	4	5	6	5
25	5	4	7	7	4	5
Mean	4.80 ±1.38	4.84 ±1.34	5.88 ±1.76	6.04 ±2.01	5.84 ±2.08	5.96 ±2.03

Table 4.1 Mean copy number of stable clones. Mean copy number was calculated by counting the FISH signals in 25 cells from six different clones (two with low expression, two with normal and two with variegating).

4.2.3 Analysis of EGFP expression in 'normal' clones

In light of the observations by FISH, that cells from the same clone have extremely variable vector copy number (between 3 and 11 for cells from clones B4 and C21), it was decided to investigate more fully the variation in EGFP expression levels both between clonal cells and between cells from different clones.

Stably established pEPI-1^{+} HeLa clones that had been cultured in non-selective media for at least 2 weeks following the selection period were fixed and analysed using CLSM to more fully assess the level of EGFP expression. Figure 4.4 (a-d) shows representative images of 4 different 'normal' clones. The fluorescence for a clone was calculated by taking the mean of the intensity measurements for 20 cells (minus the measurement for background) and the distribution of intensity levels for each clone is represented in the graphs (g-h) in Figure 4.4. The intensity measurements for 20 cells from clone B4 is shown in (f) and they range from 45-77 AU. These figures show that while there is a modest degree of variability in the EGFP levels between clonal cells, the differences are not great even though the copy number per cell ranges from 3 to 11 for clone B4. There is not a 3-fold difference in EGFP intensity measurements as would be expected if all episomes were transcribed equally in the cells – cells with 11 copies should be at least twice (and almost 3 times) as bright as those with 3 copies but differences of this magnitude are not seen. The mean intensity of fluorescence for the clones depicted is 62±9 AU for clone B4, 61±12 AU for C21, 43±4 AU for A5 and 66±4 AU for E2 indicating that overall, the levels of EGFP are similar across the different clones.

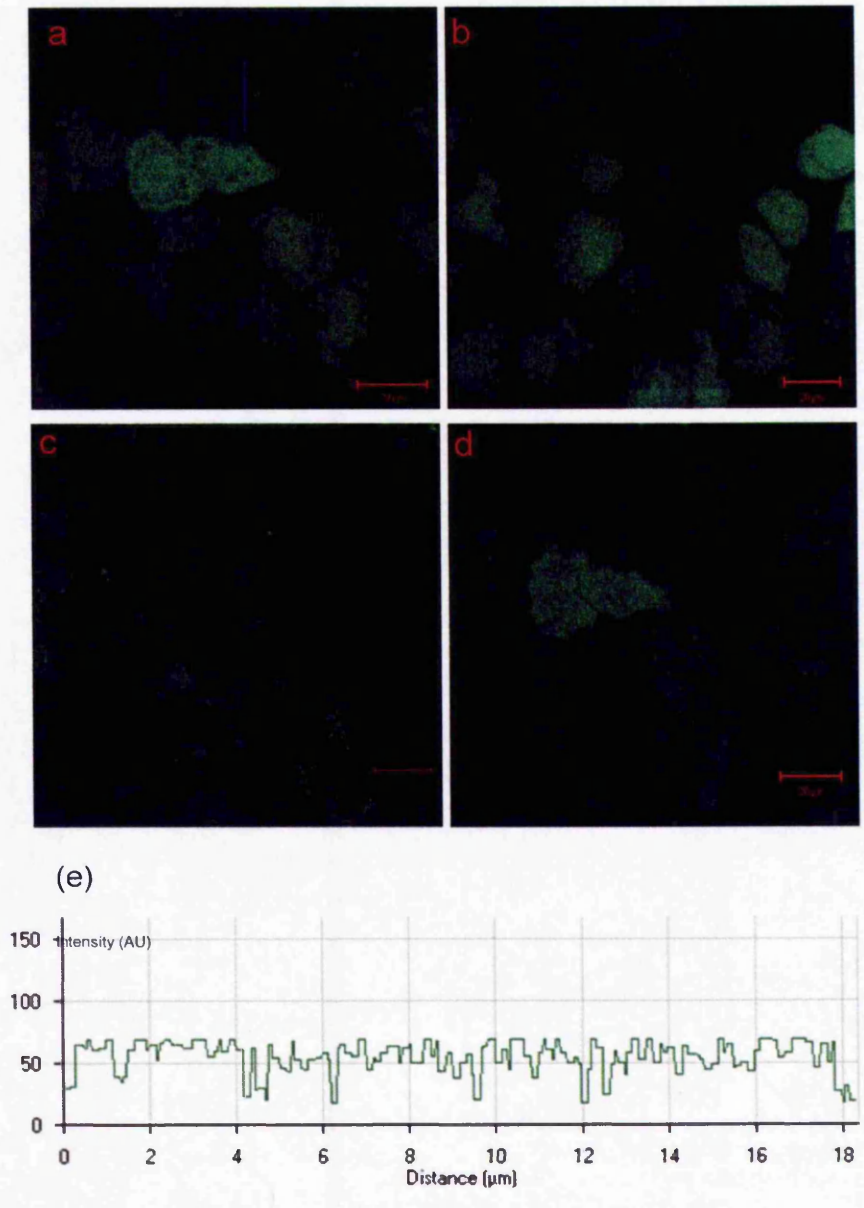
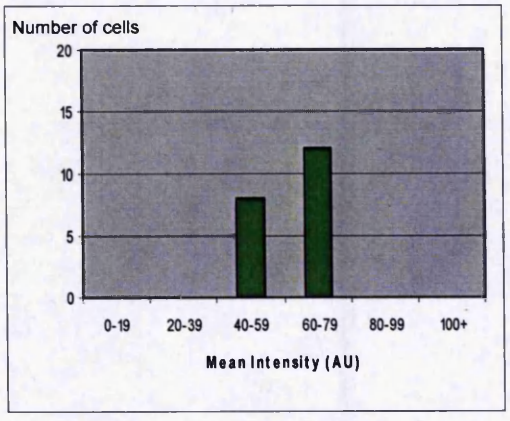


Figure 4.4. Intensity analysis of typical stable pEPI-1$^+$ HeLa cells, 'normal' phenotype. Cells were transfected with pEPI-1 and cultured in media supplemented with G418 for 3 weeks. Cells were cultured in drug-free media for 2-4 weeks prior to fixation and CLSM analysis. Mean intensity values for each cell were calculated and the mean intensity for a clone was taken as the mean of 20 cells in that clone. The background fluorescence was measured and subtracted from the intensity value measured for each cell. (a) Clone B4 has a mean intensity of 62 ±9 AU, (b) Clone C21 has a mean intensity of 61 ±12 AU, (c) Clone A5 has a mean intensity of 43 ±4 AU, (d) Clone E2 has a mean intensity of 66 ±4 AU. Blue bar corresponds to the intensity profile in (e). (f) Table of mean intensity readings for 20 cells from clone B4. (g)-(j) Distribution of mean intensity measurements for 20 cells from the clones in (a)-(d).

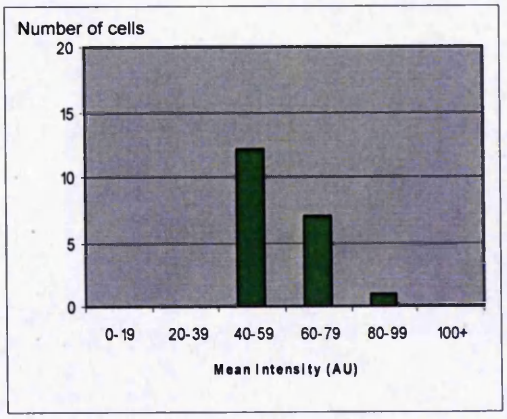
(f)

Cell Number	Mean Intensity (AU)
1	48
2	70
3	68
4	55
5	57
6	70
7	65
8	64
9	62
10	72
11	76
12	77
13	45
14	64
15	53
16	56
17	66
18	57
19	59
20	61
Mean	62 ±9

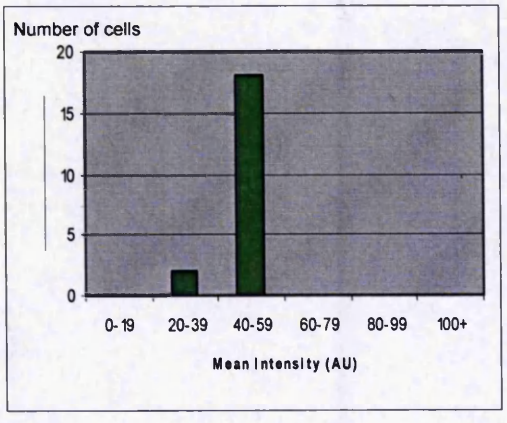
g



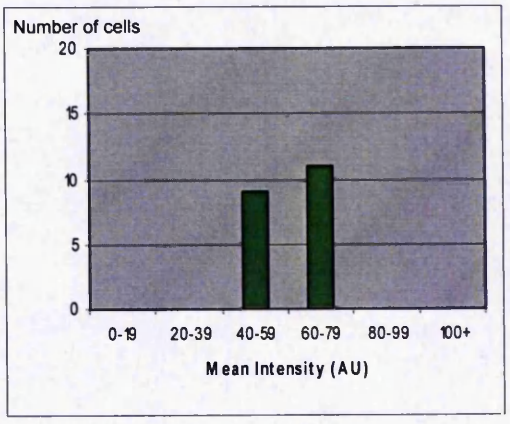
h



i

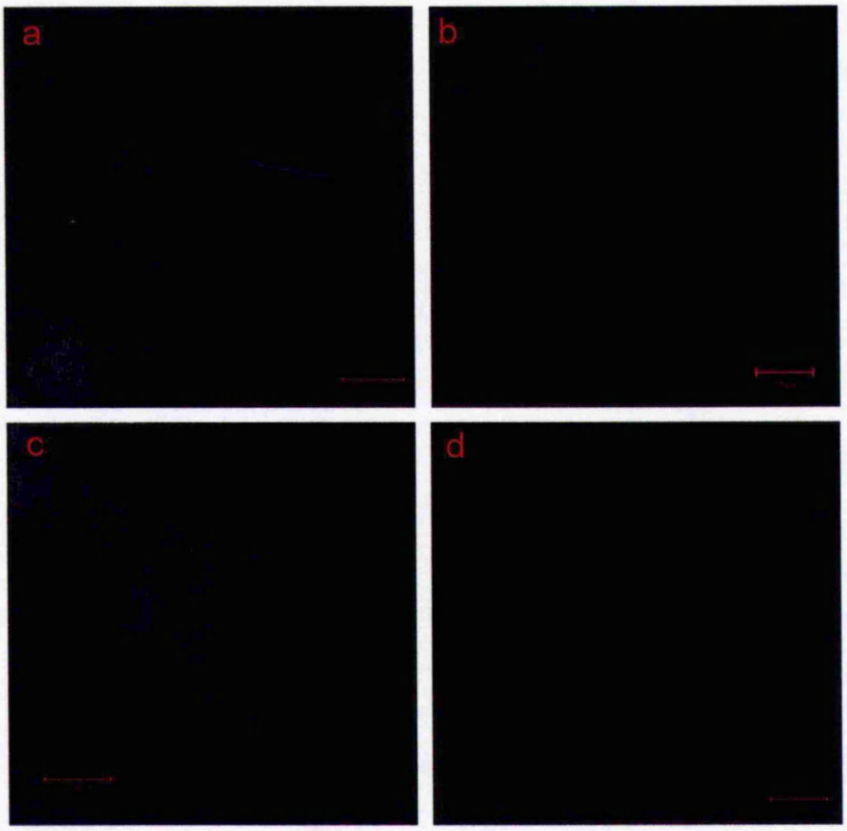


j



4.2.4 Analysis of EGFP expression in 'low' clones

Clones expressing low levels of EGFP were also analysed to determine the variation in expression levels between different cells. FISH analysis suggested that on average, these clones may have a slightly lower vector copy number and this could account for the lower levels of EGFP produced by cells. Figure 4.5 (a-d) shows representative images of 4 different low clones. The fluorescence for a clone was calculated by taking the mean of the intensity measurements for 20 cells and the distribution of intensity levels for each clone is represented in the graphs (g-h) in Figure 4.5. The intensity measurements for 20 cells from clone B7 is shown in (f) and they range from 17-39 AU. There is much less variation in the mean intensity between the 4 clones represented the figure than there is between clonal cells, but there is also a similar range of intensity in the 4 clones.



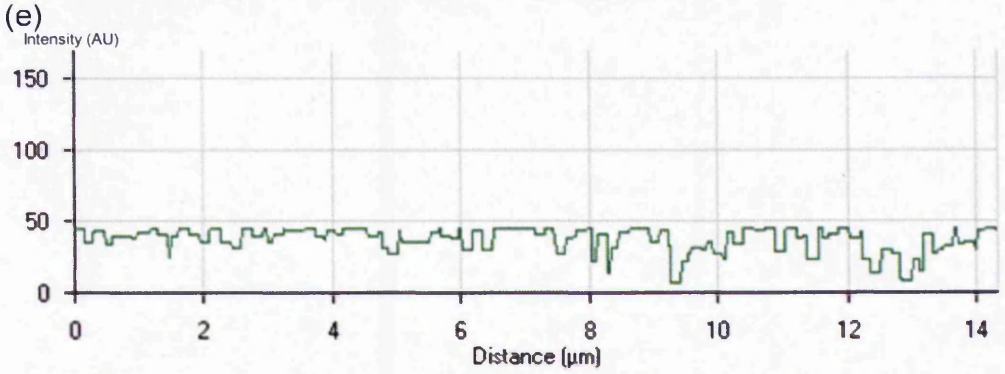
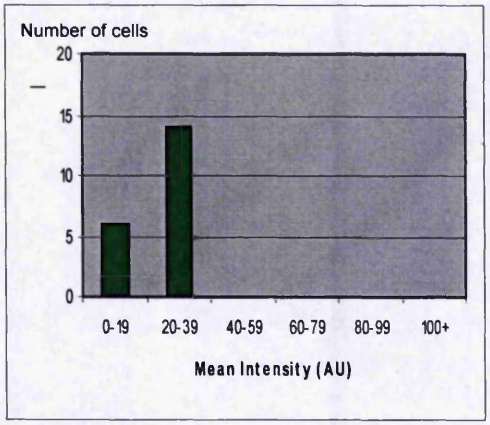


Figure 4.5. Intensity analysis of typical stable pEPI-1$^+$ HeLa cells, 'low' phenotype. Cells were transfected with pEPI-1 and cultured in media supplemented with G418 for 3 weeks. Cells were cultured in drug-free media for 2-4 weeks prior to fixation and CLSM analysis. Mean intensity values for each cell were calculated and the mean intensity for a clone was taken as the mean of 20 cells in that clone. The background fluorescence was measured and subtracted from the intensity value measured for each cell. (a) Clone B7 has a mean intensity of 29 ±8 AU, (b) Clone A11 has a mean intensity of 16 ±2 AU, (c) Clone D6 has a mean intensity of 23 ±4 AU, (d) Clone D13 has a mean intensity of 21 ±4 AU. Blue bar corresponds to the intensity profile in (e). (f) Table of mean intensity readings for 20 cells from clone B7. (g)-(j) Distribution of mean intensity measurements for 20 cells from the clones in (a)-(d).

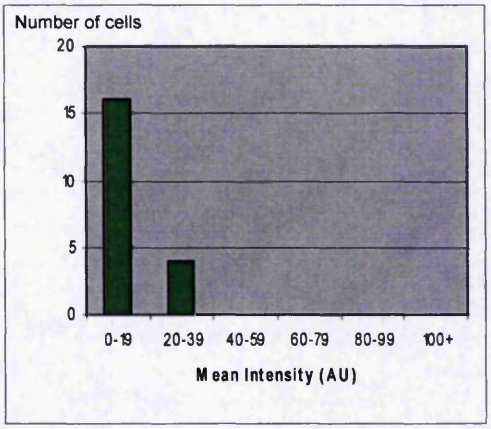
(f)

Cell Number	Mean Intensity (AU)
1	35
2	17
3	32
4	18
5	34
6	33
7	26
8	34
9	19
10	18
11	35
12	33
13	34
14	19
15	36
16	27
17	29
18	38
19	18
20	39
Mean	29 ±8

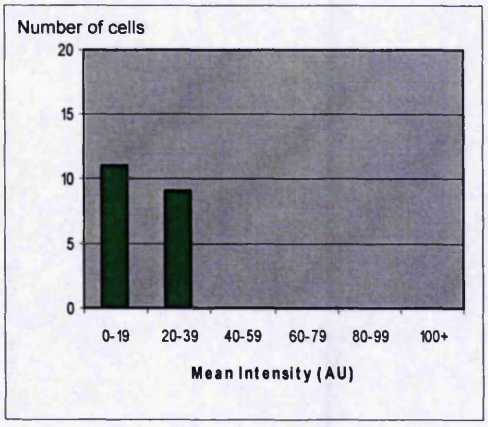
g



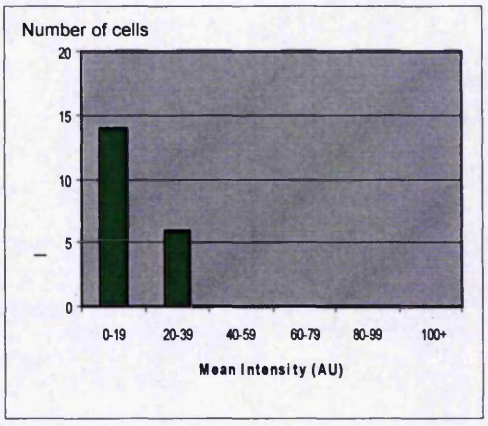
h



i



j



4.2.5 Analysis of EGFP expression in 'variegating' clones

The distribution of EGFP expression between cells within a variegating clone was investigated to establish whether there was a continuous variation or discrete levels. Figure 4.6 (a-d) shows representative images of 4 different variegating clones. The fluorescence for a clone was calculated by taking the mean of the intensity measurements for 20 cells and the distribution of intensity levels for each clone is represented in the graphs (g-h) in Figure 4.6. Analysis of these graphs suggests that there are in fact 2 distinct levels of EGFP expression within the clone rather than a continuous spread. There appears to be a background of 'normal' expression and then an approximate doubling of intensity in about 20% of the cells studied. The range of the 'normal' subpopulation of cells in clone B8 has a mean intensity of 63±7, which is similar to that measured for the cells in the normal phenotype. The 'high' subpopulation has a mean intensity of 123±17 which is by far the greatest standard deviation and it also has a greater range than the normal or low expression brackets. This possibly indicates that the 'high' element of the expression profile is more unstable than the others and so it produces a greater range of transcription levels.

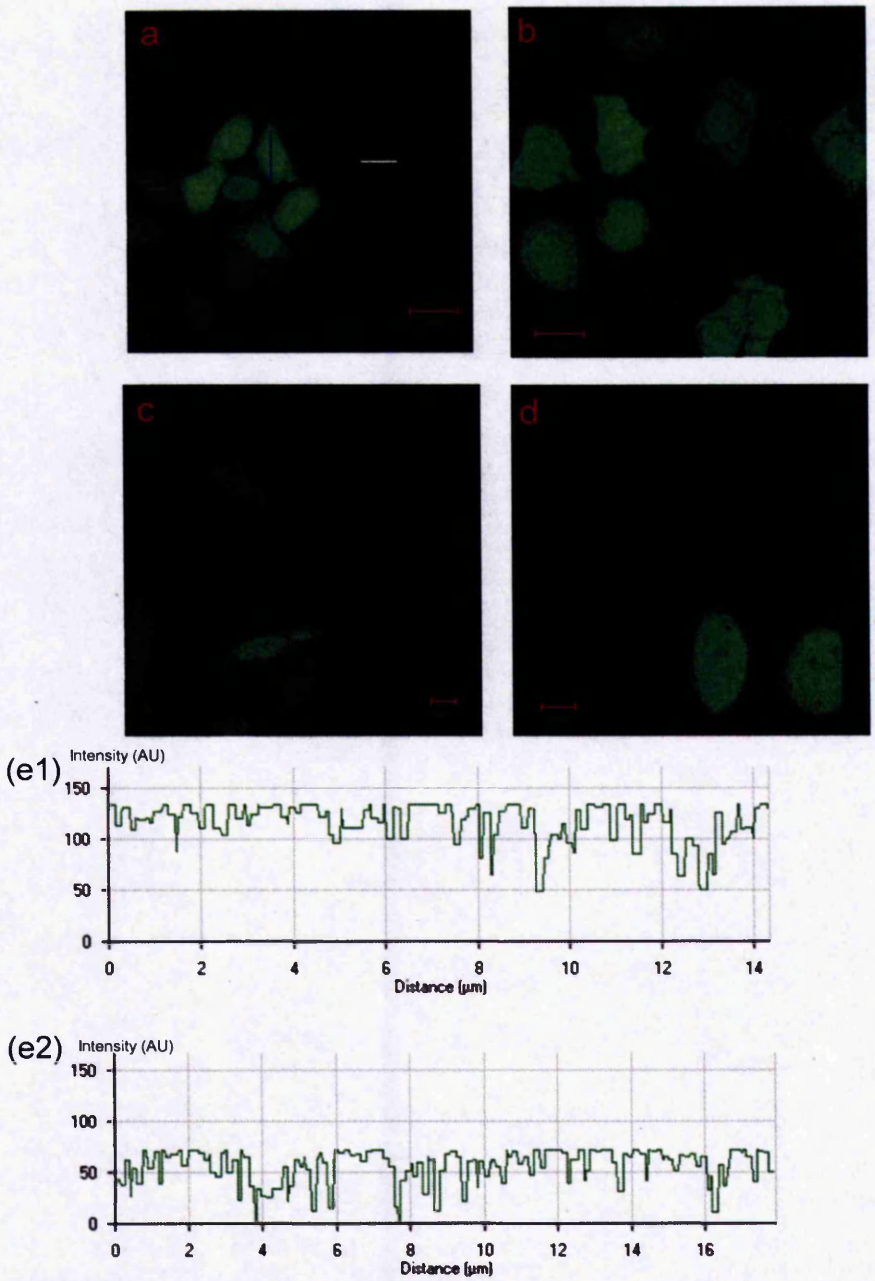
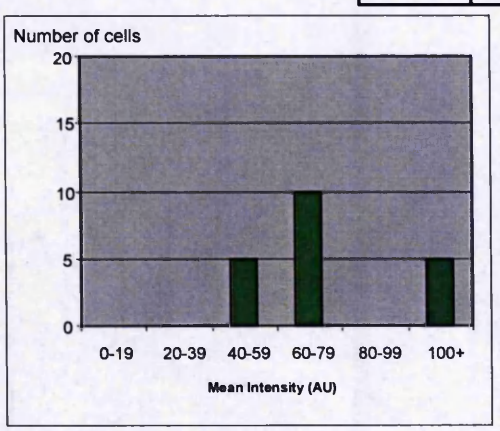


Figure 4.6. Intensity analysis of typical stable pEPI-1$^+$ HeLa cells, 'variegating' phenotype. Cells were transfected with pEPI-1 and cultured in media supplemented with G418 for 3 weeks. Cells were cultured in drug-free media for 2-4 weeks prior to fixation and CLSM analysis. Mean intensity values for each cell were calculated and the mean intensity for a clone was taken as the mean of 20 cells in that clone. The background fluorescence was measured and subtracted from the intensity value measured for each cell. (a) Clone B8 has a mean intensity of 78±28 AU, (b) Clone C16 has a mean intensity of 82±25 AU. (c) Clone A21 has a mean intensity 71±21 AU. (d) Clone E6 has a mean intensity of 74±26 AU. Blue bar corresponds to the intensity profile in (e1) White bar corresponds to the intensity profile in (e2). (f) Table of mean intensity readings for 20 cells from clone B8. (g)-(j) Distribution of mean intensity measurements for 20 cells from the clones in (a)-(d).

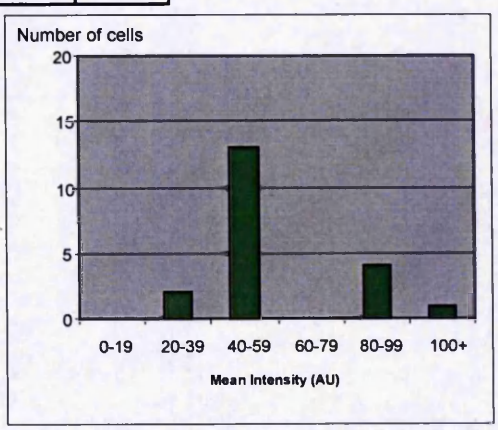
(f)

Cell Number	Mean Intensity (AU)	Mean
1	74	
2	55	
3	61	
4	66	
5	74	
6	58	
7	71	
8	54	
9	56	
10	69	
11	65	
12	61	
13	60	
14	56	
15	62	63±7
16	115	
17	113	
18	137	
19	145	
20	105	123±17
Mean	78±28	

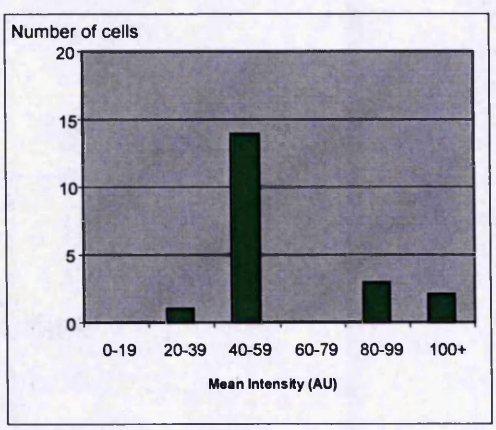
g



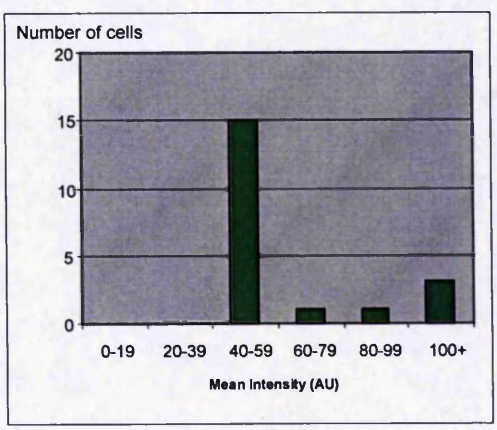
h



i

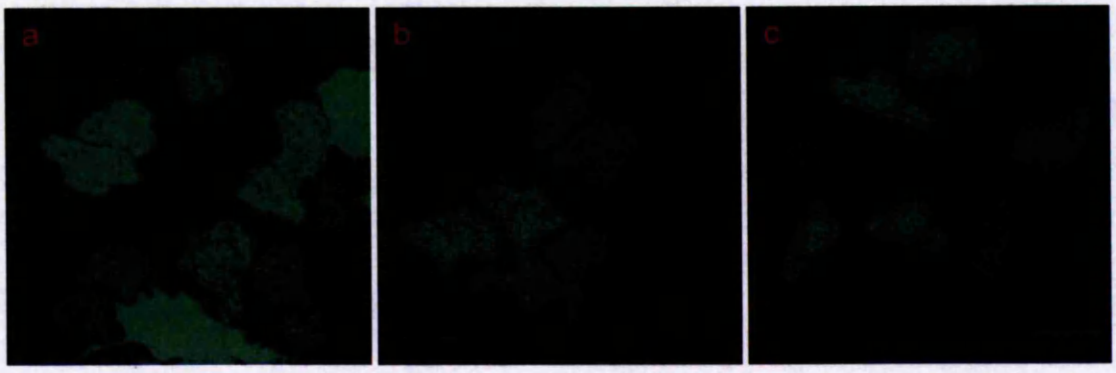


j



4.2.6 The variegating phenotype is stable

The observation by FISH that the variegating clones have a similar mean copy number and standard deviation to the normal clones suggests that the ~20% of cells expressing high amounts of EGFP are not this way because of an imbalance in copy number. Rather, it is more likely that there is higher gene expression in the bright cells. To help distinguish between these two possibilities, variegating clones were FACSorted to separate the two populations into a high and a normal fraction. The cells were cultured in the absence of selection and CLSM images were captured after 2 weeks. Figure 4.7 (a-c) shows that the variegating phenotype has reappeared in all the fractions. Figure 4.7 (d) shows the mean intensity of the cells from each fraction. This strongly suggests that the variation in EGFP expression within variegating clones is due to a difference in the level of transcription that is 'predictably unstable' so that at any one time there are a proportion of the episomes that can be more active for gene expression than is the norm and that these 'privileged' episomes can arise from any of the cells in a variegating clone.



(d)

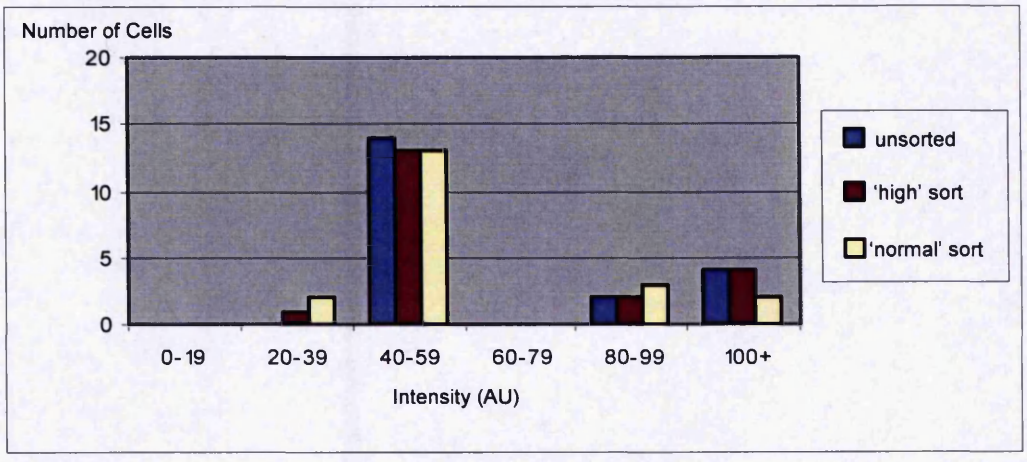


Figure 4.7. The variegating phenotype reappears in cells sorted by EGFP intensity. Cells from clone B8 were sorted by FACS into 2 populations with respect to the level of EGFP (and hence he level of fluorescence) present. These subpopulations were cultured for 2 weeks prior to fixation and CLSM analysis. (a) Unsorted control cells from clone B8. (b) Cells from the 'high' EGFP fraction. (c) Cells from the 'normal' EGFP fraction. (d) Distribution of the mean intensity measurements for 20 cells from (a-c). The 3 different cultures are reasonably similar with respect to EGFP levels indicating that all the cells in this clone have the potential to be 'variegating'.

4.3 Discussion

4.3.1 Clonal populations of HeLa cells stably transfected with pEPI-1 show 3 distinct phenotypes with respect to the level of EGFP expression

HeLa cells were transfected with pEPI-1 and selected with G418 to establish stable clones. The seeding of these transfected cells into 96 well plates concurrent with the drug-selection enabled the generation of populations derived from individual transfected cells. The cells were seeded at a density of 2-3 cells per well and it was found that this produced ample wells containing a single colony derived from a single cell. Hirt extraction and analysis of the extrachromosomal DNA recovered was performed on these clonal populations to confirm that they possessed episomal vector copies (Figure 4.1).

The analysis of fixed cells revealed clear variation in the level of EGFP expression between clonal cells and between different clones. The mean intensity of fluorescence was measured for 20 cells from 3 independent transfection-selection events and it showed three apparent phenotypes – a uniform low with an intensity of 1-39 AU, a uniform normal with an intensity of 40-79 and a variegating phenotype in which 20-30 % of the cells have a high level of EGFP intensity and the other cells have normal levels (Figure 4.2). Interestingly, there were no stable high clones. It is possible that higher EGFP transcription levels are disadvantageous to the cells and so any cells with vectors that have been fortunate enough to become localised to an extremely active nuclear region would be selected against.

4.3.2 FISH analysis of clones to determine vector copy number

Previous studies (Baiker *et al.* 2000) had provided information regarding the vector copy number in pEPI-1$^+$ clones in CHO cells. This study was focussing on the characterisation of the system in HeLa cells as it was deemed more in keeping with the potential future use of this system as a human gene therapy vector. Therefore it was essential to fully investigate all aspects of the system in HeLa cells in case there were species-specific differences. FISH analyses were

performed on stable HeLa clones and the range of copy number detected in different clonal cells was in agreement with the studies using CHO cells (Figure 4.3).

It is interesting to note that there is a far greater variation in copy number per cell within a clone, where EGFP levels appear to be fairly uniform, than there is between phenotypes, where there is a clear difference in the average EGFP level between phenotypes. This indicates that when pEPI-1 vectors are established as stable episomes, the level of EGFP expressed is no longer proportional to the copy number. In fact, it strongly suggests that the reason there are is a difference in EGFP expression between the normal and variegating phenotypes is because of differential epigenetic control of the same number of episomes per cell. The variegating phenotype might be such because the transcriptional program is not entirely stable on all episomes, for some reason. The variegating clones have, on average, the same vector copy number as the normal clones, but there are some cells with high levels of EGFP expression that could be the result of a temporary upregulation in transcription. This increased level of expression is not stably maintained and so after a period of time it is downregulated again, back to the level of the 'normal' clones. The 'low' phenotype has a mean copy number of slightly lower than the other two suggesting that copy number may play some part in the lower levels of EGFP and that the epigenetic control of the episomes is the same as that of the normal episomes. This further supports the supposition that variations in EGFP are a result of transcriptional activity rather than episome number.

4.3.3 Analysis of EGFP expression in normal and low clones

While there was some variation in the intensity of fluorescence measured between cells from the same clone, overall, there was a similar mean intensity, range and standard deviation recorded between cells of the same phenotype. The mean intensity measured for the clones here was compared with the mean copy number estimated in section 4.2.2 (Table 4.2). There appears to be some correlation between a lower mean copy number and a lower mean intensity of fluorescence, but the standard deviation is such that no definitive conclusions

Clone	**Mean copy number**	Range	Standard Deviation	**Mean intensity (AU)**	Range	Standard Deviation
B7	**4.80**	5	1.38	**29**	18	8
A11	**4.84**	5	1.34	**16**	13	2
B4	**5.88**	7	1.76	**62**	32	9
C21	**6.04**	8	2.01	**61**	37	12

Table 4.2. A comparison of copy number and intensity measurements for clone of normal and low phenotypes.

can be drawn from these figures. Without information linking copy number directly to reporter gene levels for individual cells it is not possible to determine the impact copy number has on the level of reporter gene expression.

4.3.4 Analysis of EGFP expression in variegating clones

A fuller analysis of the intensity of EGFP in clonal cells falling into this phenotype showed that there were effectively two subpopulations. Approximately 80% of cells had 'normal' intensity measurements but the remaining ~20% had a much higher mean intensity. Clone B8 had a mean intensity of 63 ±7 for the majority of the cells in this clone; this is similar to measurements obtained from the normal phenotype cells. The remaining cells had a mean intensity of 123 ±17 (Figure 4.6). This is almost double that of the bulk of the cells. This analysis also showed that the range and standard deviations are much more when dealing with the high intensity cells meaning there is less of a common gene expression level in these cells.

It is unclear how or why the variegating population exists; it could be suggested that the differences in EGFP expression result from a massive imbalance in vector copy number. Analysis by FISH certainly did not reveal any major differences in the copy numbers seen in the normal and the variegating phenotype; they both had a proportion of cells with a higher copy number. However, there is very little fluorescence variation between cells in the normal phenotype despite having a vector distribution similar to the variegating phenotype (at least as far as can be determined by FISH). Further to this

argument against a highly unstable segregation step as the cause for the variegating phenotype; the phenotype persisted in culture with no noticeable alteration in the 2 different expression subpopulations over many weeks. If mitotic instability at work here then it would be expected to see the episome gradually lost from the population resulting in an observable difference in the intensity of fluorescence than was measured here.

4.3.5 The variegating phenotype is stable

Cells from clone B8 were FACSorted into the two subpopulations according the EGFP fluorescence intensity. The cells were cultured for 2 weeks alongside a sample of unsorted clone B8 cells. If there were two subpopulations with different copy numbers (a high and a normal) then it would be expected that the 'high' fraction would continue to be very bright and the 'normal' fraction would be much more similar in appearance to the 'normal' phenotype cells. However, CLSM analysis showed that there is virtually no difference in all three of the cell cultures (Figure 4.7 (a-c)). This provides good evidence for the theory that all cells within the variegating clone have the ability to show increased EGFP expression at any time. This increase in expression is a relatively rare event as only ~20% of cells experience it at a time. These data also suggest that the switch to high expression (and off again) is reasonably rapid as the phenotype has reappeared only 14 days after sorting. This would also agree with the FISH analysis of variegating clones which indicated that the number of vector molecules in normal and variegating clones is similar. It did not, however, give any information about the epigenetic status of the episomes in each phenotype and it is likely this is where the differences will be found. Therefore, it is suggested that the variegating phenotype is not due to an instability of episome number in a subpopulation of cells that is a precursor to episome loss or a permanent reduction in gene expression, rather, the phenotype is 'stable' and persistent.

It would be tempting to speculate that not all of the episomes are responsible for the vast majority of gene expression observed. When the stable state is being established, vector molecules will be found in many different nuclear

environments – some may be in an active area and so would be permissive for efficient transcription and others could be found associated with inactive compartments. These vectors most probably would either become lost during the establishment phase as they are unable to evolve the correct set of epigenetic modifications required for long-term maintenance or they would become incorporated in heterochromatin and would be effectively silenced. The normal phenotype would have maybe half of the episomes in the highly active nuclear regions and the remainder would only support a very low level of transcription (enough to maintain the episomal state). The low phenotype has, on average, a lower copy number so it would be likely it would have less highly transcribed episomes resulting in an overall lower expression state. A third possibility could exist where vectors are balanced between these two states – they are able to evolve the epigenetic configuration necessary to mediate association with the S/MAR but they are localised to a much less active region of the nucleus. There would be a constant tussle between gene activation and deactivation. The variegating clones would have episomes that fell into this third group. They would have a similar number of episomes in the active regions to the normal phenotype and this would be the default expression level. But unlike the normal phenotype, the extra vectors would be in a region that is possibly less active and later replicating. The 'energy' needed to overcome this more compact chromatin would mean that there is not always a detectable level of transcription from these vectors. Occasionally enough activating factors will bind to open the chromatin and allow a 'normal' level of expression and this will be in addition to that stable level seen for the remaining 'normal' episomes. The likelihood of transcription is lower in these areas and that is why it is a rare event and happens in only 20% of the cells at any one time. This would explain the reappearance of the variegating phenotype after FACS sorting.

4.3.6 Conclusions

Since all cells in a clonal population were derived from a single parent cell it should follow that they all contain the same number of episomes. Different clonal populations would have different numbers of episomes but the copy

number should be the same for all cells within a population. If all the cells in a population have the same copy number then it would be expected that EGFP expression would be the same in every cell. These assumptions turned out to be incorrect in both instances; the copy number varies between cells in a single clone and the EGFP expression varies in a way that appears to be more linked to epigenetic control than copy number and indicates that not all vectors in a cell have the same transcriptional potential. The ability to directly compare episome number to gene expression using a LacI-EGFP fusion protein to tag an array of *lac* operator binding sites engineered into pEPI-1 would be vital in answering some of the outstanding questions here.

Chapter Five: Modification of pEPI-1

5.1 Introduction

The analysis of EGFP expression in established pEPI-1$^+$ HeLa clones revealed that there are significant differences in EGFP expression between cells within the same clone as well as from different clones. The reappearance of the variegating phenotype following FACSorting of a clone into fractions corresponding to EGFP levels indicated that there is a large epigenetic influence rather than a copy number effect driving the massive differences in EGFP expression observed. FISH analysis also suggested that variations in the level of EGFP expression measured are due to variations in transcription levels as the mean number of episomes per cell appears to be fairly similar between different clones. Importantly, the mean number of episomes per cell in a 'normal' clone is similar that observed for cells in a variegating clone suggesting that transcription in variegating clones is controlled by reversible epigenetic modifications to a much greater degree than in the normal clones.

FISH analysis also revealed that there is a massive difference in copy number in cells from the same clone in all phenotypes even though there is little difference in EGFP between cells in normal or low phenotypes. There remains the possibility that the different EGFP levels are a result of the variable number of episomes measured in clonal cells. The use of live cell imaging to monitor how the levels of EGFP differ between clonal cells and between cells from different phenotypes would provide a wealth of information if this could be combined with a method of determining episome copy number in the cell in question. In addition, the visualisation of individual episomes during the different phases of the cell cycle and during the crisis phase could give clues as to what the mechanisms for establishment and maintenance are. These experiments would provide the next level of investigation into the system as until now all observations have been made using fixed cells that either provide information about copy number or EGFP expression level but not both concurrently and in live cells.

It was decided to use the *lac*O-LacI system in an attempt to observe the dynamics of individual episomes in live cells. As mentioned above (section 1.2.7) the insertion of an array of *lac* operator repeats to a genome allows detection of this region due to the binding of LacI-EGFP fusion proteins to this array. The *lac*O-LacI system is probably used for such investigations because LacI binds *lac*O with a K_d of 10^{-13} M, six orders of magnitude lower than non-specific binding (Miller & Reznikoff, 1980), and has a high affinity for operator sequences within nucleosomes (Chao, Martinson & Gralla, 1980) providing a specific, high affinity 'label' for operator DNA in eukaryotic cells. This technique has been successfully used to monitor the movement of chromatin loci or domains in various circumstances and cell types. The first application of this system in eukaryotic cells was the visualisation of chromatin organisation at the level of the 100-130 nm fibre during interphase directly in living CHO and yeast cells (Robinett *et al.* 1996). Other studies showed a *lac* operator labelled late-replicating chromosome region normally located adjacent to the nuclear envelope in G1 and early S phase moved to the nuclear interior 4-6 hours after the onset of S phase. They showed that this movement was tightly correlated with the initiation of DNA replication in this labelled region (Li, Sudlow & Belmont, 1998; Belmont, 2001). There are many more examples of this system such as (Marshall *et al.* 1997; Tumbar, Sudlow & Belmont, 1999; Tsukamoto *et al.* 2000; Tumbar & Belmont, 2001; Vazquez *et al.* 2001) and is reviewed in Belmont (2001).

The initial studies integrated multiple vector copies of a 256-mer array of *lac* operator repeats into CHO and yeast cells which produced an easily detectable signal (Robinett *et al.* 1996). They also reported the detection of a clearly distinguishable signal from a single copy insert in CHO cells providing information on the limits of sensitivity of the system. It was concluded that the detection sensitivity appeared to be primarily limited by the background noise produced by unbound LacI-GFP molecules; the signal to noise ration for this 10 kb of repeats was estimated to be 12:1. Wendy Bickmore's group produced another construct, this time containing 128 *lac*O repeats and integrated 7-16 copies of this array into human HT-1080 cells which produced clear signals and allowed them to successfully perform live cell imaging (Chubb *et al.* 2002). These studies all used the *lac*O array as an integrant in the genome of the cell

line under investigation; the introduction of these amounts of foreign (prokaryotic) DNA to the cells was easily tolerated as it was only a small percentage of the total DNA. Introducing 10 kb of sequence to a 6.7 kb vector would be a completely different situation as it would virtually triple the size of the vector. This massive alteration in size could very well impact on the behaviour of the vector with respect to its establishment and/ or maintenance as an episome and would therefore negate many of the benefits of directly observing individual vectors in live cultures. It was decided reduce the size of the array to ~64 copies of the *lac*O repeat, approximately a quarter of the size of the original construct (Robinett *et al.* 1996). This was deemed long enough to be easily detected, providing ~128 binding sites (each mutated LacI binds as a dimer to one *lac*O site (Chen & Matthews, 1992) and therefore ~128 bound EGFP molecules; the limits of detection for spatially localised signals is estimated to be 10-20 molecules (Niswender *et al.* 1995). This would mean that the engineered pEPI-1 plasmid would contain at least 50% more binding sites, and hence bound EGFP than the threshold of detection. However, even though this would be reduced due to the background noise produced by unbound LacI-GFP molecules, there should still be enough bound LacI-EGFP to produce a signal bright enough to be clearly distinguishable above background.

The reporter gene in pEPI-1 is EGFP and production of this protein in the stable clones would contribute to the background noise problem generated by unbound LacI-EGFP and reduce the signal to noise ratio further. To avoid these issues, it was decided to replace the *EGFP* gene with that of another colour reporter gene. The lab was already using a vector expressing a DsRed1-LaminC reporter gene so it was decided to replace EGFP with the DsRed1 from this vector prior to insertion of the *lac*O repeats.

Aims of this chapter were:

1) Create a red fluorescent version of pEPI-1 to minimise EGFP background in investigations using the LacI-EGFP fusion protein.
2) Engineer an array of *lac*O sites into 'Red pEPI-1' to allow investigations into the behaviour of individual episomes in stable long-term clones.
3) Ensure all pEPI-1 derivatives behave in the same way as the original vector to validate any observations from the live cell imaging in (2).

5.2 Results

5.2.1 The use of DsRed as a reporter instead of EGFP in pEPI-1 is unsuitable as detection levels are too low in transiently transfected HeLa cells

It was intended to produce a modified version of pEPI-1 containing a number of *lac*O repeats to which a LacI-EGFP fusion protein encoded on a second plasmid, p3'SS-GFP-lac-NLS, would bind. Transiently transfecting p3'SS-GFP-lac-NLS into a clone stably maintaining the modified pEPI-1 would produce an excess of the protein meaning there would be a background of unbound LacI-EGFP. In order to maximise the signal-to-noise ratio it was decided to remove the additional EGFP background produced by the episome. A reporter gene would still be required verify which cells had transcriptionally active episomes so EGFP was replaced with DsRed1 to produce pEPI-DsRed1.

Briefly, the plasmid pEPI-1 was digested with *Nhe*I and *Bgl*II to remove the EGFP gene from the remainder of the plasmid (see Figure 1.2 – plasmid map). pDsRed1-LamC was also digested with *Nhe*I and *Bgl*II to release the DsRed1 gene from its backbone. The fragments were separated on an agarose gel from which they were excised and purified prior to a ligation reaction. DH5α *E. coli* were transformed with the ligation mix and cultured overnight at 37°C on 30 μg ml^{-1} kanamycin plates. Colonies were picked, amplified and plasmid recovered. Control digestions (Figure 5.1 (a)) were performed using *Nhe*I, *Bgl*II and *Eco*RI to confirm the ligation product was formed from the vector backbone (4 kb), the S/MAR (2 kb) and *DsRed1* (690 bp).

Transient transfections were performed in HeLa cells using pEPI-DsRed1 and cells were analysed by CLSM 48 hours later. The result of the analysis of these fixed cells was disappointing; the levels of DsRed1 detected were very low (Figure 5.1 (c)) and not readily detectable at levels above background. This was disappointing when compared the EGFP levels produced following transient transfections of pEPI-1 (Figure 3.1). Transient transfections introduce huge quantities of vector into cells so that the reporter is synthesised in great excess. When the pEPI-1 vectors are established in HeLa cells the copy number is estimated to be ~6 per cell and so it would be expected that the level of DsRed1 produced in stable cells would be significantly lower than that produced 48

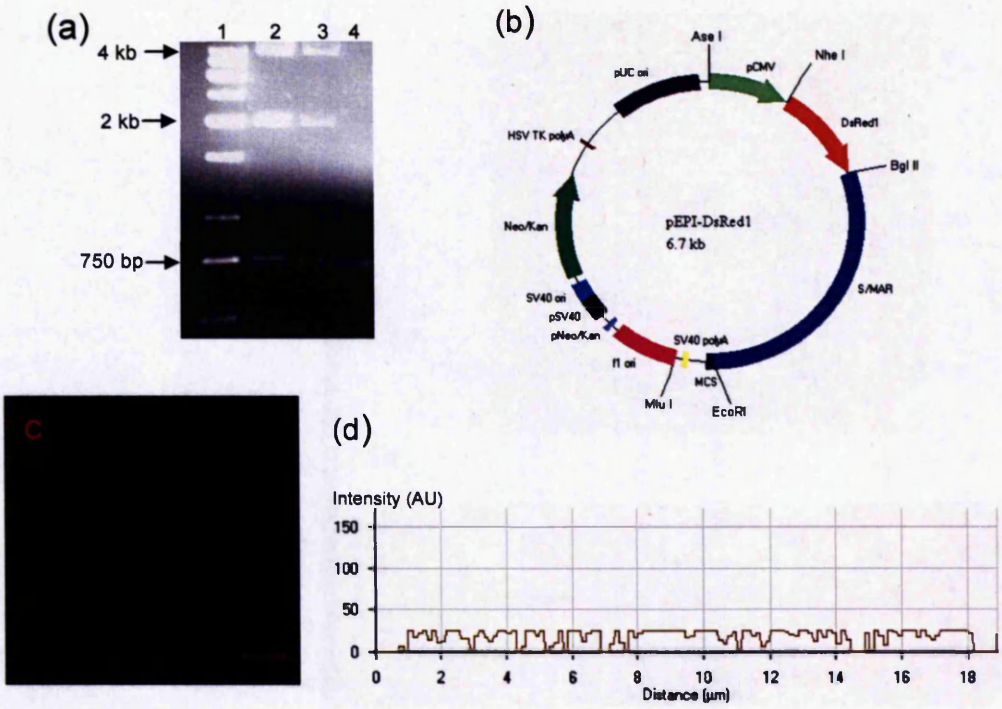


Figure 5.1. Construction and trial transfection of pEPI-DsRed1. (a) Restriction analysis of the ligation reaction between pEPI-1 (-EGFP) and *DsRed1*. Plasmid minipreps were digested with *Nhe*I, *Bgl*II and *Eco*RI then run on a 0.8% agarose gel to identify plasmids containing the 690 bp *DsRed1* insert. DNA was visualised by ethidium bromide staining. Lane 1, 1 kb ladder. Lane 2 shows a restriction digest of the original pEPI-1 plasmid and shows the 4 kb p-C1 backbone, the 2 kb S/MAR and the 720 bp *EGFP* fragment. Lane 3 is a digest of a miniprep prepared from the ligation reaction and shows the 4 kb p-C1 backbone, the 2 kb S/MAR and the 690 bp *DsRed1* insert. Lane 4 represents the *DsRed1* fragment used in the ligation reaction. (b) Vector map of pEPI-DsRed1. (c) CLSM image of HeLa cells transiently transfected with pEPI-DsRed1. Transfected HeLa cells were cultured for 48 hours prior to fixation and image capture. Red = DsRed1.

hours after a transient transfection. The level of DsRed detected in Figure 5.1 (c) was so low that any reduction in amount following the loss of vector copies during the generation of stable clones would result in it being undetectable. It was concluded that DsRed on its own would be unsuitable for use as the reporter gene in a modified version of pEPI-1 as stable long-term cultures.

5.2.2 Construction of pEPI-DsRed2-LamC and characterisation of behaviour in transfected HeLa cells

In view of the low detection of the DsRed1 produced by HeLa cells transiently transfected with pEPI-DsRed1, it was decided to produce a plasmid containing

a DsRed-LaminC fusion protein. It was hoped that the concentration of DsRed at the nuclear lamina would produce an easily detectable signal, in addition, it was hoped it would offer a clear visual boundary to the nuclear interior and provide extra information about the dynamics of the episome with respect to the nuclear membrane. It was decided not to merely insert the S/MAR sequence into the pDsRed1-LamC vector as the backbone was significantly different to that of the original pEPI-1 and this would add more variables to any investigations performed with the pEPI-1 derivatives.

Scrutiny of the sequences of pEPI-DsRed1 and pDsRed1-LamC revealed no suitable restriction sites to allow the insertion of the *LaminC* gene between the DsRed1 and S/MAR components. However, the *LaminC* fragment offered complementary restriction sites to those present in the multiple cloning site (MCS) of the family of vectors used to construct the original pEPI-1 plasmid. A red fluorescent version of pEGFP-C1 (the pEPI-1 backbone) containing an enhanced DsRed had become available and was selected for use. pDsRed1-LamC and pDsRed2-C1 were digested with *Bgl*II and *Ksp*I and the required fragments excised and purified from an agarose gel. A ligation reaction was performed and the products transformed into DH5α *E. coli*, which were then cultured on kanamycin plates. Colonies were picked, amplified and the plasmid purified prior to control digestions with *Bgl*II to confirm the correct construct had been produced (Figure 5.2 (a)). This resulting plasmid, pDsRed2-LamC, was trialled by transient transfection into HeLa cells to verify the new construct was functional. CLSM analysis of fixed cells 48 hours posttransfection showed red fluorescence localised to the nuclear boundary (Figure 5.2 (b)). This was taken as evidence that the construct was working and could be used for the production of a pEPI-1 derivative.

The next step was to insert the S/MAR into pDsRed2-LamC. The S/MAR sequence was inserted into pEPI-1 using the *Bgl*II and *Eco*RI sites in the MCS; this was unsuitable as the 3' *LaminC* sequence extends further into the MCS than *Eco*RI. Consequently, a modified version of the S/MAR was required to enable cloning into the remaining restriction sites in the MCS of pDsRed2-LamC. PCR primers (2.1.2.9) were designed that flanked the S/MAR sequence and also contained a 5' *Ksp*I and a 3' *Bam*HI restriction site. The PCR product (Figure 5.3 (a)) was ligated into pDsRed2-LamC using the *Ksp*I and *Bam*HI

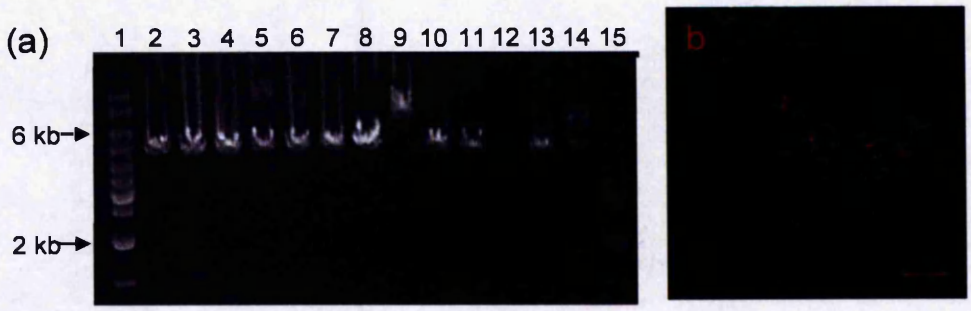


Figure 5.2. Construction and trial transfection of pDsRed-LamC. (a) Restriction analysis of the ligation reaction pDsRed2-C1 and *LaminC*. Plasmid minipreps were digested with *Bgl*II and run on a 0.8% agarose gel to identify plasmids containing the *LaminC* insert. DNA was visualised by ethidium bromide staining. Lanes 1 and 15, 1 kb ladder. Lane 2 is the original vector for comparison. Lane 12 contains no recovered plasmid. Lanes 3-8, 10, 11 and 13 represent re-ligated vector. Lane 9 represents a successful ligation reaction as seen by the increased size of the fragment compared with the vector in Lane 2. Lane 14 contains non-specific products. (b) CLSM image of HeLa cells transiently transfected with pDsRed2-LamC. Transfected HeLa cells were cultured for 48 hours prior to fixation and image capture. Red = DsRed2-LaminC.

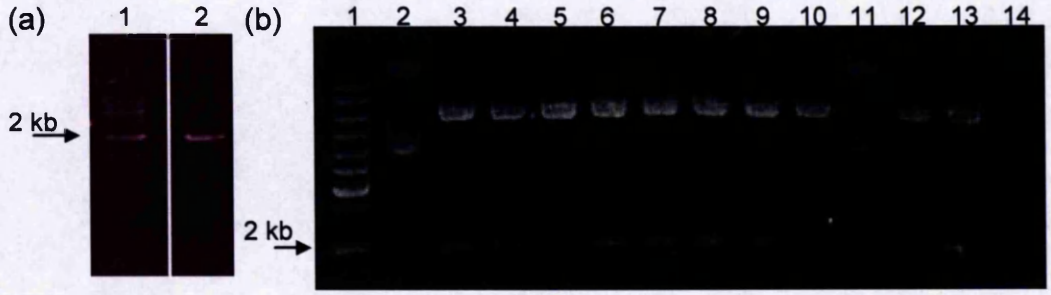
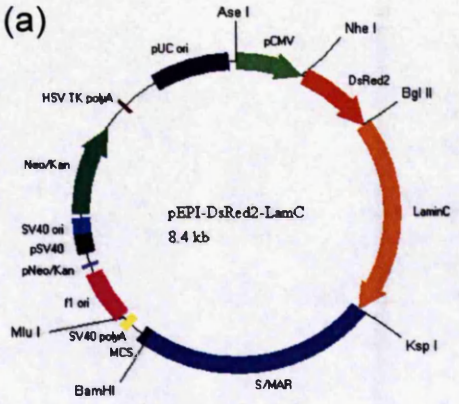


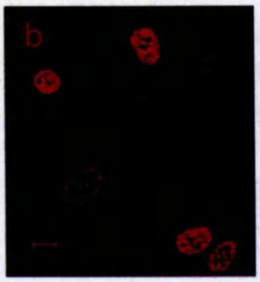
Figure 5.3. Restriction analysis of the ligation reaction between pDsRed2-LamC and the S/MAR. (a) A PCR reaction was performed according to 2.1.2.9 (Materials and Methods) at 62°C to amplify the S/MAR and insert terminal restriction sites. An aliquot of the product was run on a 0.8% agarose gel and visualised using ethidium bromide. Lane 1, 1 kb ladder. Lane 2 shows a 2 kb PCR product; the S/MAR. PCR products were restricted with *Ksp*I and *Bam*HI then ligated into correspondingly digested pDsRed2-LamC. (b) Plasmid minipreps of the ligation products purified from *E. coli* were digested with *Ksp*I and *Bam*HI and run on a 0.8% agarose gel to identify plasmids containing the 2 kb S/MAR insert. DNA was visualised by ethidium bromide staining. Lane 1, 1 kb ladder. Lanes 2 and 11 are non-specific products. Lane 5 represents re-ligated vector. Lanes 3, 4, 6-10, 12 and 13 represent successful ligation reactions reproducing the vector and insert fragments following restriction. Lane 14 is the original *Ksp*I and *Bam*HI digested pDsRed2-LamC vector for comparison

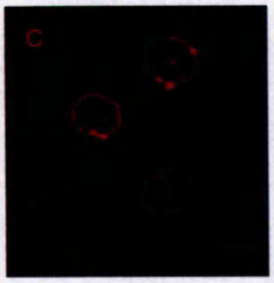
sites present in both. Figure 5.3 (b) shows control digests of plasmids recovered from DH5α *E. coli* and confirms that the S/MAR has been inserted into pDsRed2-LamC to create pEPI-DsRed2-LamC.

The resulting plasmid was trialled in HeLa cells to verify it worked in a similar manner to the parent vector, pEPI-1 (Figure 5.4). The vector expressed DsRed-LaminC in both transient and stable culture. The mean efficiency of stable clone

(a)







(d)

Transfection number	1	2	3	4	5	6
EGFP positive cells	3700	3200	3500	4100	3900	3300
Number of colonies	106	81	89	101	92	87
Efficiency of stable clone formation	2.89	2.53	2.54	2.46	2.36	2.63

Mean efficiency of stable clone formation = 2.57%

Figure 5.4. Analysis of HeLa cells transfected with pEPI-DsRed2-LamC. (a) Vector map of pEPI-1-DsRed2-LamC. (b) HeLa cells were transfected with pEPI-DsRed2-LamC and cultured for 48 hours prior to fixation and image capture. (c) Transfected HeLa cells were cultured in media supplemented with 500 µg ml^{-1} G418 for 3 weeks prior to fixation and image capture. Red = DsRed2-LaminC (d) Efficiency of stable clone formation calculated as in section 3.2.2.

formation was calculated as described in 3.2.2 and was found to be 2.57%, similar to that of pEPI-1 which was 2.61%. This suggests that the vector was performing in a comparable way to pEPI-1 during the establishment phase. Additionally, there was no significant loss of expression observed in over 5 weeks of drug-free culture following the initial selection period suggesting the vector is able to mediate continual expression of the reporter gene. Vector rescue confirmed the episomal status of the new construct and restriction analysis with *Ksp*I and *Bam*HI and comparison with the original plasmid indicated that the vector remained genetically unaltered (data not shown). The conclusion was that this version of pEPI-1 was functional and could be used in future studies.

5.2.3 Truncation of *lac*O-128 for insertion into pEPI-DsRed2-LamC and characterisation of the result in transfected HeLa cells

The next step in the construction of a plasmid that would allow the detection of individual episomes in live stably transfected cells was to insert an array of *lac* operator sites into pEPI-DsRed2-LamC. As previously discussed, the entire array of repeats from pUC21-128lac (Chubb *et al.* 2002) would be unnecessarily large so a truncated version was required. The method of construction of pUC21-128lac meant that it was made up of 16 copies of an 8-mer repeat of the *lac* operator flanked by *Eco*RI sites. A series of partial *Eco*RI digestions were performed to determine the optimal reaction conditions for the generation of a fragment in the 3-5 kb range (2.1.2.2). A partial digest using the appropriate conditions was performed and the reaction was terminated after 5 minutes by using a QIAQuick spinprep kit to remove the enzyme, after which the digest products were added to a ligation reaction with linearised pEGFP-C1. It was decided to use the plasmid pEGFP-C1 as a shuttle vector rather than the larger pEPI-DsRed2-LamC plasmid as this larger vector may reduce the efficiency of the cloning step. pEGFP-C1 was prepared for the ligation reaction by digestion in the MCS with *Eco*RI, alkaline phosphatase treatment to prevent religation and finally by ethanol precipitation to purify and concentrate the sample. Stbl2 *E. coli* were transformed with the ligation reaction and incubated on 30 μg ml^{-1} kanamycin plates. Colonies were amplified and the plasmids recovered; control digests were performed to establish if any transformants contained plasmid with an insert of an appropriate size. Two controls were performed, a double digest with *Bam*HI and *Bgl*II to ascertain the length of the *lac* operator insert and an *Eco*RI digest to confirm the integrity of the *lac* operator repeats. Figure 5.5 (a) shows a plasmid with an insert of approximately 2.4 kb (pEGFP-LacO). Transient double transfections were performed with this shuttle vector and p3'SS-GFP-lac-NLS to assess whether the *lac*O array was bound by sufficient LacI to produce a signal when viewed with the CLSM (Figure 5.5 (a)). This length of repeats appeared to allow detection of the shuttle vector above background EGFP when transiently transfected into HeLa cells.

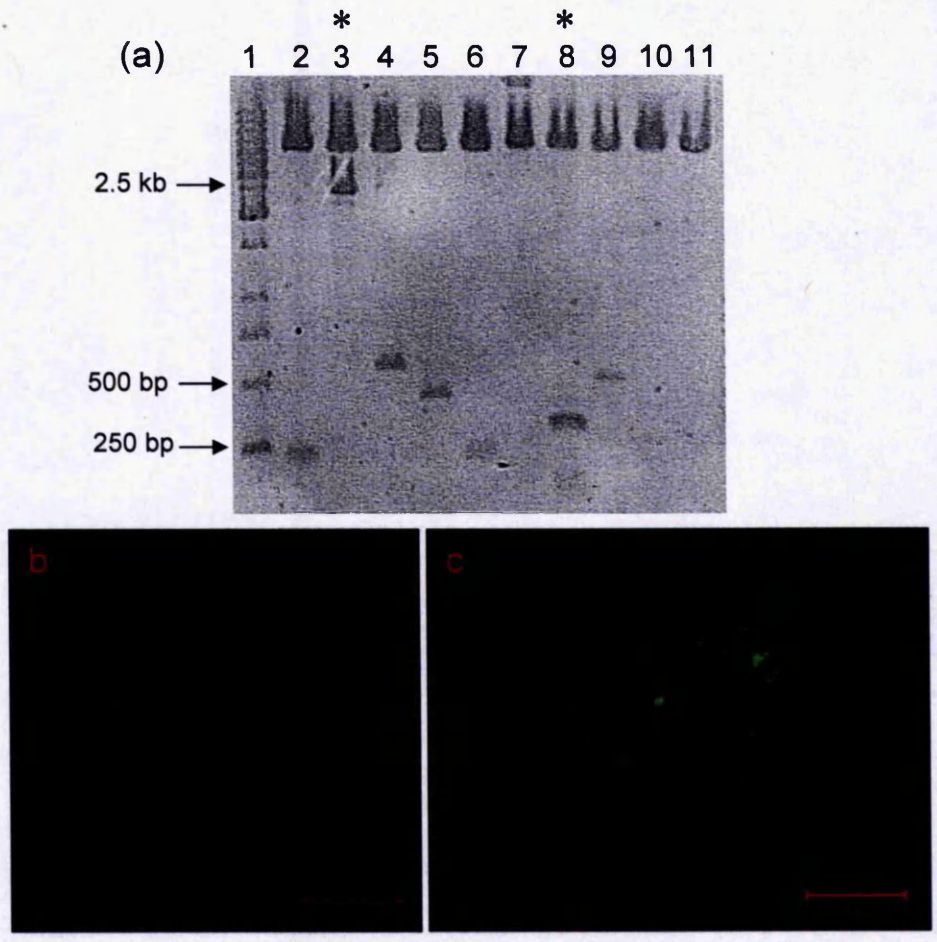


Figure 5.5. Construction of pEGFP-LacO and trial double transfections with p3'SS-GFP-lac-NLS. Partial digests (see Materials and Methods 2.1.2.2 for details) of the 128-mer *lac*O array were ligated into linearised pEGFP-C1 and transformed into *E. coli*. Purified DNA was visualised by ethidium bromide staining. Lane 1, 1 kb ladder. **Five** plasmid minipreps were digested with either *Bam*HI and *Bgl*II (Lanes 2-6) or *Eco*RI (Lanes 7-11) and run on a 0.8% agarose gel to assess the length and integrity of the *lacO* array insert. Lane 3 contains the pEGFP-C1 vector with a ~2.4 kb insert. The corresponding *Eco*RI digest in Lane 8 produces the pEGFP-C1 vector backbone and the 300 bp *lacO* basic repeat suggesting the 2.4 kb *lac*O array has not recombined. (b) Double transfections using pEGFP-LacO and p3'SS-GFP-lac-NLS were performed in HeLa cells. Images were captured 72 hours posttransfection. (a) control p3'SS-GFP-lac-NLS only transfection shows no localisation of signal. (c) double transfection showing a clear localisation of signal indicating that LacI-EGFP has bound to the *lacO* array on pEGFP-LacO vectors. Green = LacI-EGFP and EGFP (reporter gene from pEGFP-LacO).

The potential applications of this system were such that interest in developing it was high in our lab. A colleague, Kang Zheng, was also attempting to produce a shortened version of the *lac*O array for his studies and he produced a fragment of 3.5 kb (approximately 60 *lac* operator repeats, half the original length) by *Eco*RI partial digestion of pUC21-128lac and using pBlueScript as the shuttle vector. It was decided to proceed with this length of repeats rather than the 2.4 kb to maximise the signal to noise ratio in cells. Zheng removed the array from

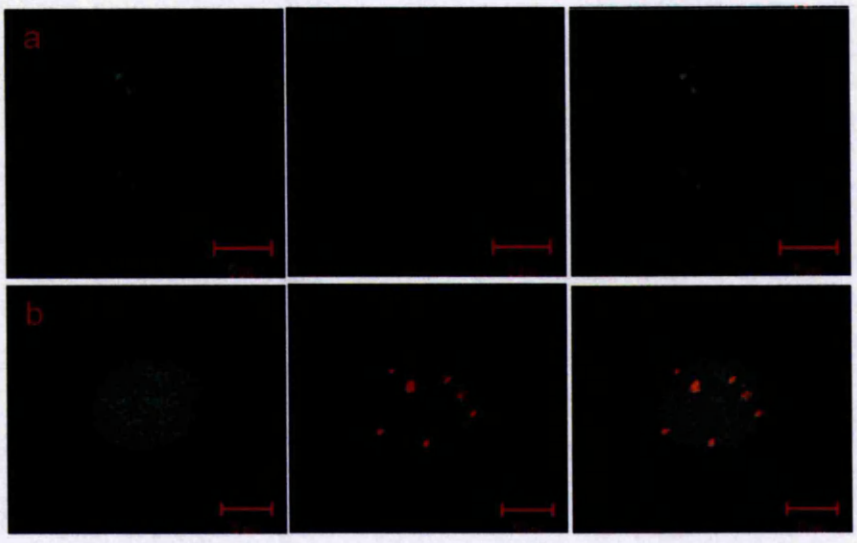


Figure 5.6. Localisation of LacI-EGFP signal in transient transfections to plasmids containing *lacO* arrays but not to plasmids lacking this sequence. (a) HeLa cells were transiently transfected with pEPI-DsRed2-H2B-Lac60 and p3'SS-GFP-lac-NLS and were fixed and examined using the CLSM 48 hours later. A clear localisation of signal is seen and this is assumed to be due to LacI-EGFP molecules binding to *lac*O arrays in pEPI-DsRed2-H2B-Lac60. Green = LacI-EGFP, red = DsRed2-H2B. (b) HeLa cells transiently transfected with pEPI-DsRed2-LamC (lacks the operator array) and p3'SS-GFP-lac-NLS demonstrate no localisation of the signal into discrete foci. Green = LacI-EGFP, red = DsRed2-LaminC.

pBlueScript by digestion with *Mlu*I and inserted it into the *Mlu*I site of pEPI-DsRed2-LamC to produce pEPI-DsRed2-LamC-Lac60. In order to visualise the position of the plasmid with respect to chromatin, Zheng also produced a plasmid containing the *H2B* gene in place of the *LaminC* gene; pEPI-DsRed2-H2B-Lac60.

Trial double transfections using these new pEPI-1 derivatives combined with the p3'SS-GFP-lac-NLS were performed to assess the whether the signal generated would be sufficiently above the background noise. Figure 5.6 indicates that a clear green fluorescence corresponding to the colocalisation of LacI-EGFP molecules to the pEPI-1 derivative with the *lac*O array in HeLa cells can be clearly seen above any non-specific background. The p3'SS-GFP-lac-NLS and pEPI-DsRed2-LamC control double transfection showed no such localisation of signal indicating that the LacI-EGFP protein was in fact specifically labelling the *lac*O repeats in the pEPI-1 derivative used. These observations were encouraging and it was decided to proceed with establishing stable clones with the new pEPI-1 derivatives to allow analysis of the long-term behaviour of the episome.

5.2.4 Characterisation of clones stably transfected with pEPI-DsRed2-H2B-Lac60 or pEPI-DsRed2-LamC-Lac60

The new pEPI-1 derivatives were assessed to determine whether they behaved in the same manner as the original vector. It was noted that the transfection efficiencies of these new larger vectors was reduced, typically around 20%. This was not unexpected as larger plasmids are harder to transfect into cells. In addition, it was observed that the efficiency of stable clone formation was lower than pEPI-1 or pEPI-DsRed2-LamC at around 2% and this was similar for both pEPI-DsRed2-H2B-Lac60 and pEPI-DsRed2-LamC-Lac60. Analysis of stable clones (named EPI-DHL- or EPI-DLL- respectively) suggested that there was mainly a uniform level of expression of the reporter genes. This was much harder to assess than with EGFP as a reporter as the DsRed-LaminC seemed to form aggregates at the nuclear membrane rather than forming a discrete ring around the nucleus as was expected. It was difficult to determine whether there were any different phenotypes with respect to gene expression similar to those seen in Chapter 4. The localisation of signal to the nuclear membrane rather than a diffuse signal across the whole of the cell meant that any changes in reporter protein levels wouldn't be readily detected.

Analysis of the stable clones at various intervals following the removal of G418 selective pressure indicated that the 'stable' state was not entirely so. There was a loss of reporter gene expression in ~20% of cells after 8 weeks, and a near complete loss of expression after 16 weeks, of drug-free culture suggesting that the presence of the *lac*O array was destabilising the actions of the S/MAR. This gradual loss of expression was found to be due to inactivation of the episome rather than episome loss as reintroduction of G418 did not result in additional cell death. Additionally, successful vector rescue indicated that the episome was still present, just silent. These technical issues should still have allowed the use of the *lac*O-LacI system to monitor the behaviour of the vectors in the short to mid-term as even though the episomal behaviour wasn't stable in the long-term, the investigation of the process of establishment and short-term maintenance would still be valuable. Unfortunately, transfection of these stable clones with p3'SS-GFP-lac-NLS yielded no discernable localisation of signal when compared with the control p3'SS-GFP-lac-NLS only transfection (data not

shown). It was unclear what the problem was since the vectors had appeared to produce a clear signal when used in transient double transfections. The formation of chromatin on the *lac*O array may have been masking some LacI binding sites meaning that these reduced numbers of bound LacI-EGFP couldn't be detected above the background noise of unbound LacI-EGFP. This possibility was investigated in couple of ways; cells were fixed and analysed by CLSM for up to 5 days posttransfection with (varying amounts of) p3'SS-GFP-lac-NLS in an attempt to effectively reduce and remove the unbound component. It was hoped that as the p3'SS-GFP-lac-NLS vector was lost from the culture the levels of LacI-EGFP would also diminish making it easier to detect the *lac*O bound fraction. This did not produce any improvement so it was decided to attempt to visualise the signal in live cells in case the fixing procedure was affecting it. Again, the p3'SS-GFP-lac-NLS vector was transfected into stable cells at various concentrations and cells were monitored for up to 7 days but no unambiguous signals could be reliably distinguished above the background. In a final attempt to determine whether any detectable levels of LacI-EGFP were *lac*O bound in stable HeLa cells, a region of the nucleus of live cells was photobleached until all of the unbound LacI-EGFP was undetectable. It was hoped that this would remove all background and allow identification of the *lac*O bound fraction which would have escaped the photobleaching. This also failed to produce any results and so it was concluded that, for whatever reason, there was insufficient LacI-EGFP bound to these *lac*60 pEPI-1 derivatives in stable long-term clones to allow detection in live or fixed cells.

5.3 Discussion

5.3.1 The use of DsRed as a reporter instead of EGFP in pEPI-1 is unsuitable as detection levels are too low in transiently transfected HeLa cells

In order to minimise the EGFP background in stable pEPI-1$^+$ HeLa cells transiently transfected with p3'SS-GFP-lac-NLS for visualisation of individual episomes, the EGFP reporter in pEPI-1 was replaced with DsRed. It was decided to exchange EGFP for DsRed in pEPI-1 rather than in p3'SS-GFP-lac-NLS because previously described difficulties in producing a functional LacI-EGFP protein made it possible that modification of p3'SS-GFP-lac-NLS would result in a non-functional construct (Belmont & Straight, 1998). In addition, the usefulness of future generations of pEPI-1 in gene therapy applications would require the synthesis of therapeutic proteins. If the system is so sensitive that it cannot tolerate such a small modification in reporter gene then it would not bode well for the development of this vector.

A DsRed version of pEPI-1 was generated and trialled in HeLa cells for its effectiveness. Unfortunately, even though there is a clear red fluorescent signal when it is produced as a fusion with LaminC and localised to a cell substructure, there was only a low level detected when it was freely diffusible (Figure 5.1 (c)). This posed a significant problem as transient transfections introduce thousands more copies of the vector than would be stably maintained as episomes following drug selection if this vector were to behave the same as the parent pEPI-1. If the level of DsRed produced during a transient transfection was barely detectable then it could be assumed that there would be no red fluorescence identifiable in stable populations. Double transfections of this vector with a LaminB1-EGFP expressing vector were used to verify that there were no problems with the transfection; the transfection efficiency was over 30% each time as estimated LaminB1-EGFP expression (data not shown). It was unclear why this construct did not produce a brighter signal when freely diffusible, but it should be remembered that early versions of GFP were not terribly effective when not in a fusion with a localised protein (Niswender *et al.* 1995).

5.3.2 Construction of pEPI-DsRed2-LamC and characterisation of behaviour in transfected HeLa cells

The release of an enhanced version of DsRed opened the possibility of redoing the construction of the vector in 5.3.1 with this new, brighter mutant. However, it was decided to engineer pEPI-1 to express a DsRed2-LaminC fusion protein instead of a freely diffusible reporter protein. There was concern that the DsRed2 produced would leak through into the green channel during microscopy and add to the green background present from unbound LacI-EGFP protein. Expression of DsRed2-LaminC would also allow easy identification of the nuclear boundary without the use of chromatin stains which could prove useful during live cell imaging. The DsRed2-LamC plasmid was constructed and transfection into HeLa cells followed by CLSM analysis confirmed the vector was functional (Figure 5.2). The S/MAR was inserted into this vector and the new construct was also trialled in HeLa cells. In addition to the expression of the DsRed2-LaminC reporter during the transient phase, G418 selection for 3 weeks generated stable clones with pEPI-DsRed2-LamC maintained as an episome, as verified by vector rescue (data not shown), which continued to express the reporter even after removal of selection (Figure 5.4 (c)).

If this construct was to be used as a basis for further modifications, such as the insertion of the *lac*O array, it was essential that it was confirmed that the vector behaved in the same way as pEPI-1 otherwise any observations made with it might not be applicable to pEPI-1 and therefore would be of little use. The efficiency of stable clone formation in HeLa cells was investigated and was found to be comparable to that of pEPI-1 – 2.57% for pEPI-DsRed2-LamC and 2.61% for pEPI-1. The stable populations generated continued to express DsRed2-LamC when cultured in the absence of selection pressure (Figure 5.4 (c)). In addition, the vector was recovered by Hirt extraction demonstrating it remained episomal. These observations all suggest that pEPI-DsRed2-LamC effectively functions the same way as pEPI-1 and can be substituted for it in future studies. A version of pEPI-1 expressing a truncated rat NGF-receptor gene as a fusion with EGFP has also been shown to be functional as an episome (Jenke *et al.* 2004a) suggesting that insertion of genes upstream of the S/MAR does not abolish the episomal properties of the vector.

5.3.3 Truncation of *lac*O-128 for insertion into pEPI-DsRed2-LamC and characterisation of the result in transfected HeLa cells

The final step in the construction of a vector that would allow its direct visualisation in living cells was to insert a length of *lac*O repeats. The 128-mer would almost triple the size of the vector so a truncated version was required. Partial digestion of the array with *Eco*RI produced fragments of varying repeat length, one of which, corresponding to ~60 repeats, was ultimately cloned into pEPI-DsRed2-LamC. Another version was also constructed where H2B was substituted for LaminC. These vectors were transiently transfected into HeLa cells along with the p3'SS-GFP-lac-NLS to assess the intensity of the signal generated. CLSM analysis of fixed cells revealed there were bright green fluorescent signals present in the doubly transfected cells but not in the p3'SS-GFP-lac-NLS singly transfected cells indicating that the LacI-EGFP protein was in fact localising to either pEPI-DsRed2-H2B-Lac60 or pEPI-DsRed2-LamC-Lac60, as appropriate (Figure 5.6). Based on these observations it was decided to proceed with establishing stable long-term clones of these pEPI-1 derivatives.

5.3.4 Characterisation of clones stably transfected with pEPI-DsRed2-H2B-Lac60 or pEPI-DsRed2-LamC-Lac60

If these constructs were to be used to investigate the mechanisms of establishment and maintenance of pEPI-1 then it would be essential to demonstrate the vectors were behaving in the same way. The efficiency of stable clone formation (described in section 3.2.2) was calculated to be approximately 2% for both constructs. This was noticeably lower than that observed for pEPI-1 or pEPI-DsRed2-LamC. It is unclear why this efficiency is lower, whether the increase in vector size is somehow destabilising the delicate balance that allows the small percentage of pEPI-1 molecules to become stably episomal or whether it is due to the presence of 3.5 kb of repetitive bacterial DNA. This size versus bacterial DNA issue could be investigated by replacing the 3.5 kb *lac*O sequence with a 3.5 kb gene of mammalian origin and

observing whether there is still a reduction in the efficiency of stable clone formation. If the size of the vector is the problem then this is an issue that would have to be addressed before pEPI-1 could be useful as a gene therapy vector since genes and their associated regulatory elements are often much larger than 3.5 kb.

Stable clones were generated by G418 selection of pEPI-DsRed2-H2B-Lac60 or pEPI-DsRed2-LamC-Lac60 transfected HeLa cells for 3 weeks. Clones were monitored for reporter gene expression at regular intervals for up to 16 weeks culture in non-selective media. It was found that there was a loss of reporter gene expression in ~20% of cells in all clones after 8 weeks. This loss of expression continued until there was virtually a complete loss of reporter gene expression detected by CLSM at 16 weeks post-selection (~1% of cells were still positive for reporter gene expression). Reintroduction of G418 did not result in any significant cell death and vector rescue of Hirt extracts indicated that the episome was still present in the clones, just inactive for reporter gene expression. This loss of expression is troubling since the advantage pEPI-1 has over other prototype gene therapy vectors is its ability to mediate long-term expression in culture without the use of selection. The downregulation of transcription could be as a result the properties of the *lac*O insert. The extra bacterial DNA might be a target for the host cell defences which would treat it as an infection and seek to promote the formation of epigenetic modifications associated with silent chromatin. The S/MAR appears to protect the pEPI-1 domain from silencing but this protective capacity may be overridden by too much bacterial DNA. Alternatively, the presence of such a repetitive sequence could be the cause of the gene inactivation. Repetitive DNA tracts in mammalian genomes are frequently found as heterochromatin and it is possible that the cell recognises the repetitive nature of the sequence and epigenetic modifications are made to heterochromatinise the domain. The replacement of the *lac*O sequence with a similar length of either a mammalian gene or a repetitive sequence, such as α-satellite DNA, in this construct should allow differentiation between these possibilities. It is unlikely that the size of the vector *per se* is the cause of the instability since YACs of 15.8 kb have been shown to be stably maintained episomally (Cossons *et al.* 1997).

Despite the instability of these pEPI-1 derivatives in long-term non-selective culture, it was decided that they could still be extremely valuable tools for investigating the dynamics episomal state in short-term studies. Additionally, the loss of reporter gene expression could be correlated with a change in nuclear localisation. For instance, the LacI-EGFP labelling of the episome might reveal that it moves from a nuclear interior position to a more peripheral one when gene expression is downregulated as changes in transcriptional status and nuclear localisation appear to be intimately linked.

The loss of gene expression in long-term culture was not the biggest problem with these clones; transfection with p3'SS-GFP-lac-NLS failed to produce any clear localisation of signal. Various different attempts were made to observe LacI-EGFP bound to episome but it appears there are insufficient numbers of LacI-EGFP bound to produce any detectable signal using the CLSM in fixed or live cells. It is possible that the formation of heterochromatin on the episome was masking some of the binding sites, but no signal was seen even immediately after the selection period when the reporter gene expression was still good, indicating that the vector had 'active' open chromatin at this point.

A report published since these investigations were carried out may go some way to explaining why this approach failed (Wang *et al.* 2005). The use of single molecule studies of LacI-EGFP bound to the 256-mer *lac*O array found that the mean occupancy is 2.5%. Since each site binds a dimer, it was therefore estimated that 26 LacI-EGFP molecules were bound to 256 repeats of the *lac*O sequence. They suggested a model to explain these long-range interactions whereby the strain energy induced by the binding of the LacI influences the ability of subsequent molecules to bind in a non-linear way (strain energy varies as the square of the twisting angle or bending radius). An increase in strain energy could therefore strongly inhibit the occupancy of neighbouring sites. Based on these figures, it could be estimated that the mean occupancy of 2.5% of a 60-mer of *lac* operator repeats would result in the localisation of less than 5 EGFP molecules, less than the threshold of detection (Niswender *et al.* 1995) The obvious question would be if the 60-mer occupancy of the 60-mer is below the detection level then how was it visualised in transient double transfections? One possible explanation for this would be that the excess of plasmids introduced were forming aggregates and effectively amplifying the signal. As the

copy number dropped, during the establishment phase, there were insufficient vectors left to colocalise and produce a detectable signal. It also follows that each green dot observed by CLSM following the double transient transfection did not represent a single vector as was previously believed but instead was the result of multiple vectors collected together.

There remains the possibility that individual episomes in stable clones could still be visualised in live cell studies if the intensity of the label could be improved. The use of EviTags (Evident Technologies) conjugated to LacI could allow detection of the *lac*O array. The tags are reported to be brighter than conventional fluorophores and withstand longer periods of microscope analysis. Recombinant LacI could be conjugated to these labels and *in vitro* then microinjected or electroporated into stable clones where they can still be detected more than seven days later.

Chapter Six: Analysing Vector Copy Number and Transgene Expression

6.1 Introduction

The failure to detect any signal from the *lac*O bound LacI-EGFP label in stably established HeLa cells prevented the use of this system to visualise individual episomes in live cells and to relate this information to the levels of gene expression observed. However, the use of antibody staining against EGFP component of LacI-EGFP in fixed cells would still provide invaluable information about the nature of the episomal state. The benefit of antibody staining would be the amplification of signal afforded by the binding of multiple fluorophores to the *lac*O array (via interaction of the primary antibody with EGFP and the secondary antibody with the primary). Immunostaining is also potentially much more powerful than FISH as it better preserves the nuclear architecture, a factor that must be included if a complete picture about how and why pEPI-1 works is to be constructed.

The analysis of EGFP levels and mean copy number in clones stably established with pEPI-1 provided a useful insight into the influence copy number and transcriptional control had on the amount of reporter gene present in these clones. The observation that there was a range of EGFP expression levels in different cells from the same clone raised the question of the nature of the variation. A variegating clone that was FACSorted into two subpopulations with respect to fluorescence intensity ('high' and 'normal') did not stably maintain the expression levels in the subpopulations. In fact, when analysed 2 weeks later there was no appreciable difference between the sorted populations; both subpopulations had the same ratio of normal and high EGFP expressing cells as each other and the unsorted population. This indicated that the variations in expression levels were dynamic and that any cell within the clone had the ability to up or down regulate expression. It was unclear exactly how dynamic switch was in these cells – did the expression levels change gradually over many cell cycles as the chromatin epigenetics gradually evolved? Or was there a rapid

upregulation and how prolonged was it? The use of live cell imaging to monitor the changes in the levels of EGFP present in cells would possibly help to answer some of these questions. Therefore, even though the live cell imaging studies cannot make use of the *lac*O-LacI system to directly visualise the relationship between gene expression and episome number, there is still a lot of information to be gathered about the nature of the variation in reporter gene expression. This characterisation of the dynamic nature of gene expression would be useful for the future development of pEPI-1 as a gene therapy vector since the appropriate amount of gene product would need to be produced at the right time, and the overexpression of this gene product, even in only 20% of cells, might be catastrophic.

The aims of this chapter were to (a) investigate the use of antibody staining against LacI-EGFP bound to the *lac*O array on pEPI-DsRed2-H2B-Lac60 or pEPI-DsRed2-LamC-Lac60 and (b) monitor the changes in EGFP levels with live cell imaging in order to better understand the nature of the variation.

6.2 Results

6.2.1 Using immunostaining to determine the copy number in established HeLa clones

Clones were established as previously described (Chapter 3) following transfection of HeLa cells with either pEPI-DsRed2-H2B-Lac60 or pEPI-DsRed2-LamC-Lac60. To prevent any ambiguity in results, all clones were investigated only up to 6 weeks after the initial selection phase; at this point there still appeared to be transgene expression in virtually all cells in a clone and so it was assumed that the behaviour of the episome was similar to that of pEPI-1 up to the time point. Clones were transfected with p3'SS-GFP-lac-NLS and were cultured for 24 hours, after which time they were seeded onto glass coverslips. Cells were synchronised by incubating with aphidicolin 36 hours after transfection for a period of 12 hours, after this time cells were washed for 30 minutes in at least 3 changes of media to remove the maximum amount of drug and release the cells into S phase. Cells typically underwent mitosis ~12 hours after release – this could be monitored using a standard bench-top light microscope. The cells were fixed using methanol after mitosis had been observed meaning that the majority of cells studied by immunostaining would be in early G1 thus eliminating any influence of cell cycle on vector copy numbers recorded (cells in G2 would have undergone a round of replication without division and so will appear to have double the copy number). The optimum incubation times and antibody concentrations were determined (section 2.4 for a fuller protocol), briefly, all steps were carried out on ice in a darkened humidified chamber; cells were methanol fixed followed by blocking in PBS^+ (2% (w/v) BSA in PBS) for 1 hour. Cells were washed once and then incubated with the primary antibody, a rabbit polyclonal anti-GFP, at a dilution of 1:200 in PBS^+ for 1 hour. The primary antibody was removed by washing with PBS. The second blocking step involved incubation with 2% goat serum in PBS^+ for 30 minutes followed by PBS washing. The secondary antibody, Alexa Fluor 488 goat anti-rabbit, was diluted to 5 $\mu g\ ml^{-1}$ in PBS^+ and incubated with the cells for 30 minutes. The secondary antibody was removed by washing with PBS, after

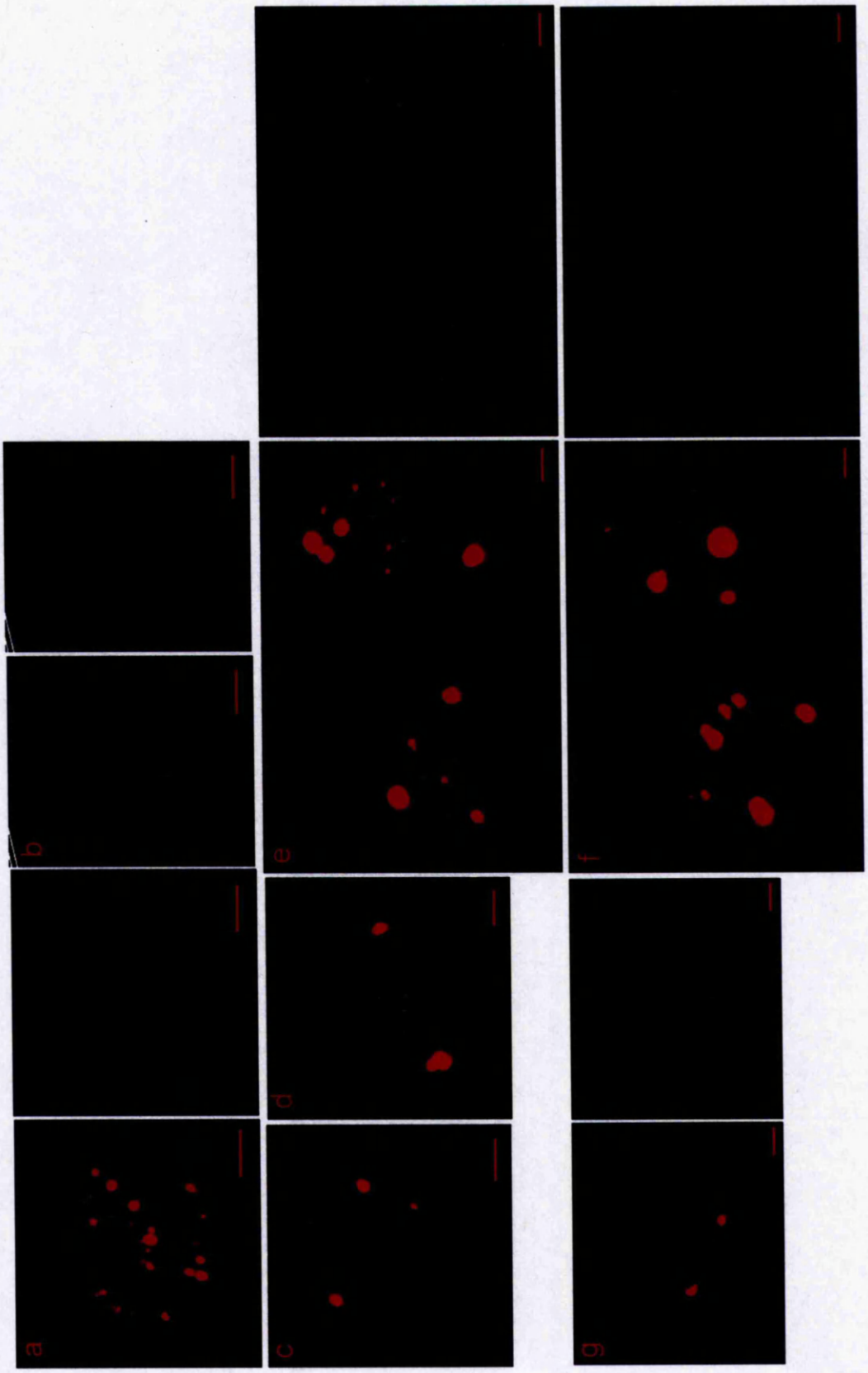

a
b
c
d
e
f
g

Figure 6.1. Representative CLSM images of immunostained episomes in stably established pEPI-DsRed2-LamC-Lac60 HeLa clones. Clones were transfected with p3'SS-GFP-lac-NLS and synchronised using aphidicolin 24 hours later. Cells were cultured for 48 hours after release then fixed and immunostained to visualise episomes using the CLSM via labelling of LacI-GFP bound to the *lacO* sites on the episomes (see Chapter 2 for details). (a) Antibody control; clone EPI-DLL-8 mock transfected shows no signal when immunostained. (b) Antibody control; clone EPI-DHL-1 was mock transfected and shows no staining following immunostaining. (c) Unprocessed image of established pEPI-DsRed2-LamC transfected with p3'SS-GFP-lac-NLS shows no localisation of signal following immunostaining. (d-g) established clones were transfected with p3'SS-GFP-lac-NLS and show clear signals above background following immunostaining. (d) Unprocessed image of clone EPI-DLL-8 (e) Clone EPI-DLL-2. (f) Clone EPI-DLL-3. (g) Clone EPI-DLL-8. Red=DsRed2-LaminC, green = Alexa Fluor 488. Scale bar represents 5 µm. Images were produced using the Imaris software and are 'snapshots' of 3 dimensional 'Surpass' images.

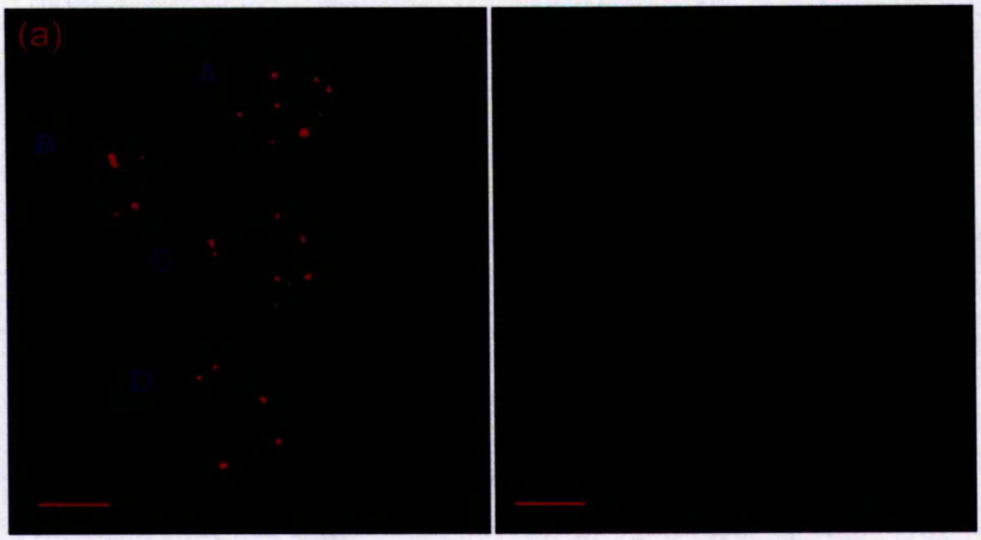
this the cells were mounted on glass slides and examined by CLSM whereby the number of signals per cell was recorded. Figure 6.1 shows representative images (processed using the Imaris software, images are a 'snapshot' of the 'Surpass' three dimensional projection) of immunostained episomes in HeLa clones stably established with pEPI-DsRed2-LamC-Lac60. There were no signals seen when cells were immunostained without transfection of p3'SS-GFP-lac-NLS (Figure 6.1 (a) and (b)) indicating that the antibody is specifically labelling the LacI-EGFP protein. Transfection of p3'SS-GFP-lac-NLS into a stable population of pEPI-DsRed2-LamC cells (no *lac*60 insert) shows no localisation of signal following immunostaining indicating that the signal obtained is genuinely produced by the localisation of LacI-EGFP specifically bound to *lac*O arrays in the 'EPI' vectors in positive cells (Figure 6.1 (c) and (d)). Slides (e-g) are representative images from clones EPI-DLL-2, EPI-DLL-3 and EPI-DLL-8; these populations were stably established with pEPI-derivative episomes carrying the *lac*O array which were bound by transiently transfected LacI-EGFP. Antibody staining against EGFP allowed identification of individual episomes in fixed samples of these clones. The number of signals detected from 50 cells prepared in this way from 3 different pEPI-DsRed2-LamC-Lac60 and 3 different pEPI-DsRed2-H2B-Lac60 clones was collated in Table 6.1 and shows copy number in these cells ranges from 1 to 12 with standard deviations up to ±2.76 for clone EPI-DLL-3. The mean copy number for each clone is reasonably similar, estimated to be approximately 4 vector copies per cell. This is slightly lower than the estimate of approximately 6 copies per cell for clones established with the original pEPI-1 vector obtained in section 4.2.2.

Nucleus	Copy .Number					
	pEPI-DLL-2	pEPI-DLL-3	pEPI-DLL-8	pEPI-DHL-1	pEPI-DHL-2	pEPI-DHL-5
1	7	1	4	5	3	5
2	4	6	3	4	3	3
3	3	7	3	6	4	3
4	6	1	2	6	7	2
5	10	4	4	7	5	5
6	1	6	4	3	3	2
7	7	4	3	3	4	2
8	2	7	4	4	4	4
9	2	10	9	3	4	7
10	1	8	9	4	3	4
11	3	7	4	2	4	7
12	2	2	4	9	9	3
13	2	6	7	5	3	4
14	2	3	3	5	11	4
15	4	4	2	2	3	3
16	4	1	1	5	2	3
17	2	1	2	4	2	3
18	1	10	2	8	4	3
19	2	9	3	4	2	6
20	7	2	7	6	6	3
21	3	10	2	2	7	8
22	4	5	2	2	3	5
23	3	7	4	6	8	4
24	7	4	6	3	3	4
25	2	3	2	5	3	6
26	6	3	2	5	4	2
27	3	4	3	3	4	1
28	2	2	7	4	3	5
29	2	6	1	3	4	3
30	2	3	4	3	2	6
31	4	4	7	4	7	3
32	2	1	4	4	9	2
33	2	2	3	1	3	4
34	1	2	3	11	3	3
35	4	2	1	4	1	4
36	2	3	2	3	3	5
37	6	1	2	2	4	4
38	2	4	6	2	4	7
39	8	7	6	3	5	9
40	2	12	3	1	3	3
41	3	3	8	5	2	2
42	6	2	12	3	3	5
43	12	2	2	4	10	3
44	2	5	4	3	3	2
45	1	3	4	2	3	4
46	3	2	3	9	4	2
47	9	3	2	3	3	3
48	3	3	3	3	4	3
49	4	6	2	4	3	5
50	2	4	3	4	4	3
Total	184	217	193	206	208	196
Mean	3.68	4.34	3.86	4.12	4.16	3.92
St. Dev.	±2.53	±2.76	±2.34	±2.05	±2.14	±1.71

Table 6.1. pEPI-DsRed2-LamC-Lac60 or pEPI-DsRed2-H2B-Lac60 copy number in established HeLa clones. Cells were treated as in Figure 6.1, Briefly, HeLa clones established with pEPI-derivative episomes carrying the *lac*O array were transiently transfected with p3'SS-GFP-lac-NLS. LacI-EGFP molecules localised to individual episomes were visualised by antibody staining and allowed determination of episome number in individual cells. Copy number for each clone was calculated as the mean copy number recorded for 50 cells from a clone.

6.2.2 Post-mitotic distribution of episomes

There was a big difference in the highest and lowest number of episomes detected in clonal cells established with the pEPI-1 vector (measured by FISH), pEPI-DsRed2-LamC-Lac60 and pEPI-DsRed2-H2B-Lac60 (measured by immunostaining of LacI-EGFP bound to the *lac*O array). To better understand these imbalances in copy number, it was decided to perform analyses on daughter cells to ascertain the distribution of vectors in each nucleus and hence how frequently a segregation imbalance occurred. Cells of clone EPI-DLL-8, a typical established clone, were transfected with p3'SS-GFP-lac-NLS and cultured for 24 hours prior to seeding onto coverslips at low density to allow easy identification of daughter cells. Aphidicolin was used 12 hours after seeding to synchronise the cells; they were released 12 hours later by thorough washing with 3 media changes over a 30 minute period. The cells were fixed and immunostained when they had undergone either one or two mitotic events and CLSM analysis was used to assess the distribution of vector molecules in the daughter pairs. Figure 6.2 (b) depicts the post-mitotic distribution observed for 25 daughter pairs of clone EPI-DLL-8 and shows that there is equal segregation of molecules into daughter cells in only 32% of the mitoses studied. There was identical segregation or a difference of only 1 in 72% of the cases and there were no imbalances greater than 3 (and this only occurred in 12% of cases) indicating that in general the segregation of vector molecules into daughter cells is reasonably efficient. To further investigate the segregation of episomes at mitosis, it was decided to measure the distribution of molecules after two mitoses to see if there was a pattern to the mis-segregation. Figure 6.2 (a) is a representative image of the result of two mitoses and the distribution observed was 5, 2, 2, 5. The arrangement of cells in this field suggests that A and B are daughters of one cell from the first mitosis and C and D are daughters from the second cell created after the first round of cell division. This means that even though both divisions produced unequal segregation of episomes, there is an apparent symmetry to the imbalance. Table 6.2 shows the analysis of 10 cells that have undergone two rounds of mitosis and the observed copy number in each daughter cell. Nuclei labelled A and B have



(b)

Early G1 Daughter Pair	Copy Number		Irregular
	Nucleus A	Nucleus B	molecules
1	2	2	0
2	3	6	3
3	7	4	3
4	1	2	1
5	4	4	0
6	7	7	0
7	4	5	1
8	3	2	1
9	3	2	1
10	1	1	0
11	2	2	0
12	2	3	1
13	4	6	2
14	6	4	2
15	3	4	1
16	8	9	1
17	12	9	3
18	2	4	2
19	4	3	1
20	4	4	0
21	3	3	0
22	2	2	0
23	4	2	2
24	2	3	1
25	3	4	1
Total	96	97	27
Mean	3.84	3.88	1.08

Figure 6.2. Analysis of post-mitotic distribution of episomes in clone EPI-DLL-8. Cells were transfected with p3'SS-GFP-lac-NLS and cultured for ~16 hours prior to seeding onto coverslips at a low density. Cells were allowed to adhere after the trypsin treatment for 6-8 hours then were synchronised with aphidicolin. Cells were cultured for 1 or 2 mitoses after which time they were fixed (during early G1, immediately after the mitotic event), immunostained and analysed using the CLSM. (a) Representative of image of the result of two mitoses. Note the unequal segregation of vector molecules in the daughter cells. Red = DsRed2-LaminC, green = Alexa Fluor 488. Scale bar represents 10 µm. Images were produced using the Imaris software and are 'snapshots' of 3 dimensional 'Surpass' images. (b) Table showing the episome number observed in 25 early G1 daughter pairs following a single round of cell division.

Early G1	Copy Number				Irregular	Molecules	Between:
Daughter Pair	Nucleus A	Nucleus B	Nucleus C	Nucleus D	A and B	C and D	AB and CD
1	5	2	2	5	3	3	0
2	2	4	2	4	2	2	0
3	3	3	3	4	0	1	1
4	4	3	4	4	1	0	1
5	6	6	7	5	0	2	2
6	4	5	6	4	1	1	0
7	5	7	3	3	2	0	2
8	4	2	4	2	2	2	0
9	4	4	4	4	0	0	0
10	2	2	4	3	0	1	1
Total	39	38	39	38	11	12	7
Mean	3.9	3.8	3.9	3.8	1.1	1.2	0.7

Table 6.2 Post mitotic distribution of vector molecules at the four cell stage. Cells were treated as in Figure 6.2 and the number of signals obtained after immunostaining cells that had undergone 2 rounds of mitosis was counted and compared between daughter cells. Nuclei A and B were designated as the mitotic product of one daughter cell and nuclei C and D the product of the other daughter cell produced from a single mother cell.

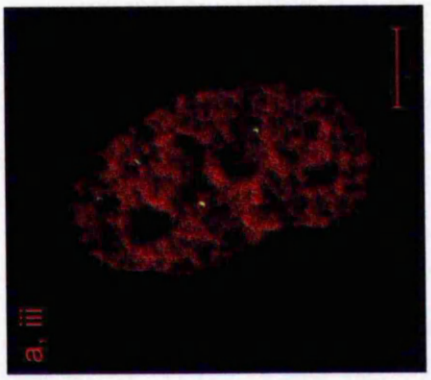
been deemed to be the mitotic product of one cell, and those labelled C and D the product of the other cell produced after the first mitosis (Figure 6.2 (a)). There was the same number of mis-segregated molecules between cells A and B as there were in cells C and D in 50% of the daughter sets studied. There appears to be a greater variation in copy number between two daughters than between the total episomes detected in these daughters and the total counted in the second set of daughters produced after the two rounds of mitosis. Put another way, the total number of episomes varies between A and B or C and D more than it varies between AB and CD. Taking the cells in Figure 6.2 as an example of this, there is an imbalance of 3 between daughter cells A and B, there is also an imbalance of 3 between daughter cells C and D but the total number of vectors in A and B is the same as that in C and D, 7 in each daughter pair.

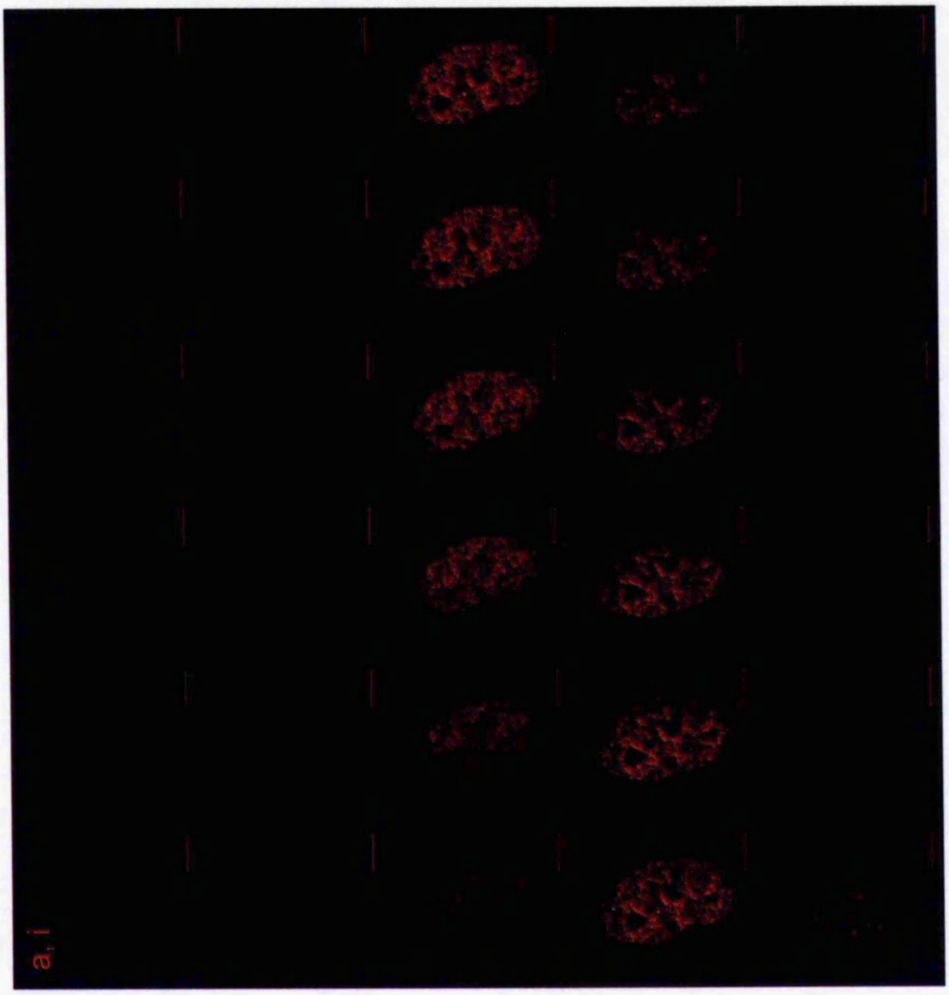
6.2.3 Episomes appear to be generally localised to areas of low chromatin density

Established pEPI-DsRed2-H2B-Lac60 clones expressing the reporter DsRed-H2B were employed to provide information about the association of the episomes with different regions of the nucleus. Cells from clone EPI-DHL-1 or EPI-DHL-5 were transfected with p3'SS-GFP-lac-NLS and cultured for 24 hours

prior to seeding onto coverslips. Aphidicolin was used 12 hours after seeding to synchronise the cells; they were released 12 hours later by thorough washing with 3 media changes over a 30 minute period. Cells were monitored for the onset of mitosis and were fixed shortly afterwards, during early G1. Cells were immunostained and CLSM was used for image capture. Figure 6.3 shows a gallery of a Z-stack taken of a typical cell from each clone. The enlarged images from the gallery show that many of the episomes are situated in areas of lower red fluorescence (DsRed2-H2B) suggesting that they are localised either to chromatin poor regions or to regions of more active, open chromatin. In addition to this, it was observed that in clone EPI-DHL-5 there were more episomes nearer the nuclear periphery than in clone EPI-DHL-1; interestingly, the expression levels of DsRed2-H2B appear to be lower in this clone than in EPI-DHL-1. This agrees with many observations that show the nuclear localisation of a gene can influence its expression, with less active genes typically being found in a more nuclear peripheral position.


a, ii


a, iii


a, i

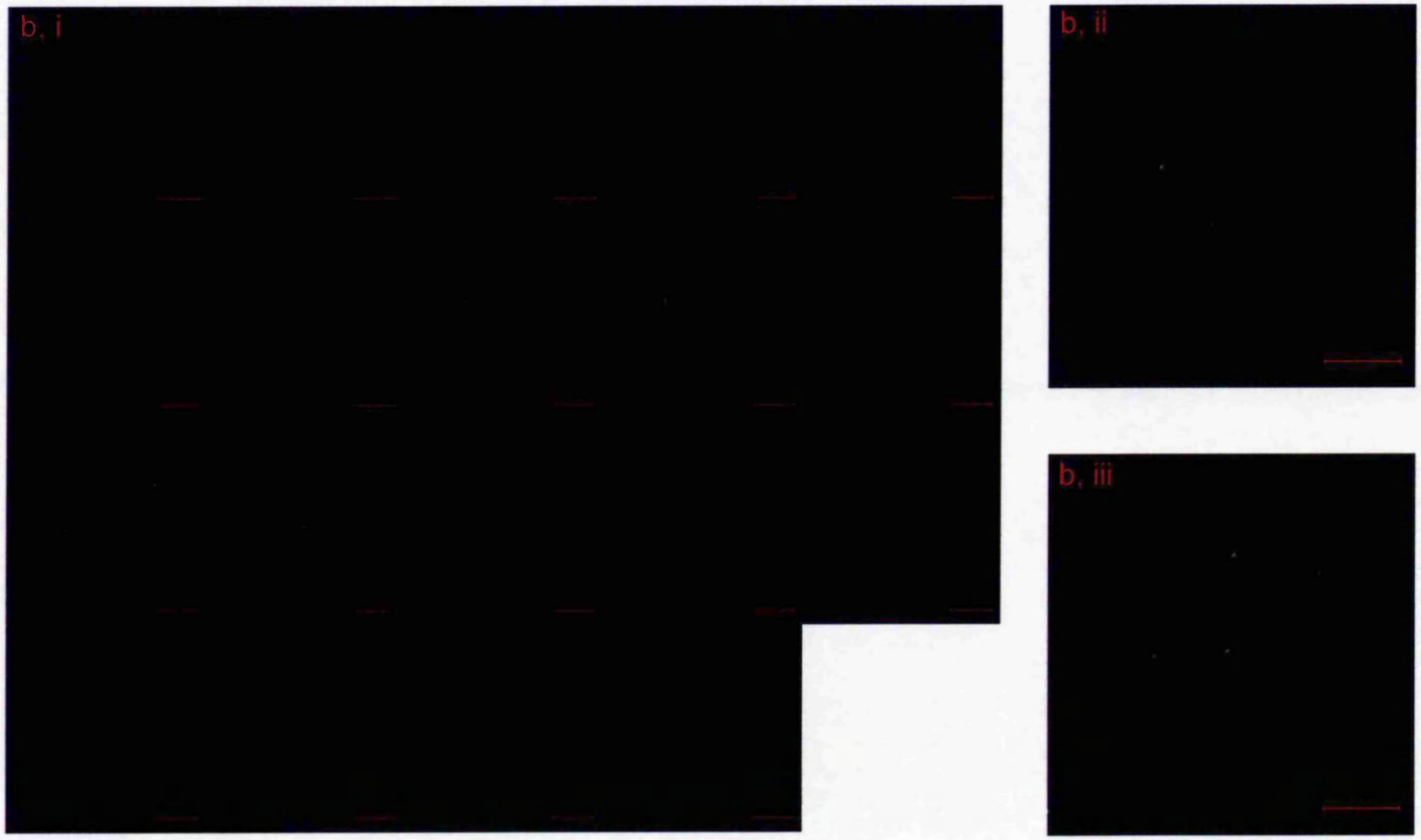


Figure 6.3. Visualisation of pEPI-DsRed2-H2B-Lac60 episomes in established HeLa cells by immunostaining. Clones were transfected with p3'SS-GFP-lac-NLS and synchronised using aphidicolin 24 hours later (see M+M for details). Cells were cultured for 48 hours after release then fixed and immunostained to visualise episomes using the CLSM via labelling of LacI-GFP bound to the *lacO* sites on the episomes. (a) Clone EPI-DHL-1 and (b) Clone EPI-DHL-5. (i) gallery of Z-stack, (ii) and (iii) enlarged images from the gallery. The episomes appear to be localised to areas of the nucleus with less dense DsRed2-H2B fluorescence suggesting that they are colocalised with more open euchromatin as opposed to the more densely packed heterochromatin. Red = DsRed2-H2B, green = Alexa Fluor 488.

6.2.4 Investigating EGFP intensity in live cells

The use of immunostaining to visualise individual episomes in stable clones provided information about the distribution of episomes in daughter cells. However, it did not provide detailed information about the level of gene expression in these cells making it difficult to assess the impact of copy number on reporter protein synthesis. It was decided to utilise live cell imaging to monitor the variations in EGFP expression in stable established pEPI-1^{+} HeLa clones to ascertain the nature of the variation. Cells were seeded into poly-lysine coated, gridded, glass-bottomed culture dishes at least 24 hours prior to microscopy. Once the cells had settled, they were treated with aphidicolin for 12 hours to synchronise them so that comparisons between cells would be valid; they would be at a similar stage in the cell cycle and so any fluctuations in gene expression due to cell cycle phase would be equivalent. Cells were cultured in colourless media supplemented with 25 mM HEPES to help buffer the solution. Cells were cultured in a standard incubator and were only moved to the microscope briefly at each time point for image capture. Image capture was performed on a heated stage in a heated, humidified chamber with 5% CO_2 to minimise the stress to the cells. To further reduce the damage caused by microscope analysis the mercury lamp was never used to view the EGFP in the cells; therefore, cells were randomly selected at the beginning of the experiment and not chosen on the basis of EGFP levels.

Figures 6.4, 6.5 and 6.6 show representative time courses of cells from clones B7 (mean copy number of 4.80), B4 (mean copy number of 5.88) and B8 (mean copy number of 5.84) respectively. Intensity analysis of fixed samples of these clones had suggested that there were three distinct phenotypes, a low, a normal and a variegating phenotype (Chapter 4). FACSorting clone B8 into subpopulations with respect to EGFP expression resulted in the reappearance of the variegating phenotype by the time the fractions were examined by CLSM 2 weeks later suggesting that all cells in the population have the potential to be highly variable for EGFP expression. This analysis did not provide any information about the temporal aspect of the variation so the intensity of EGFP fluorescence was monitored in cells from clones B7, B4 and B8. There was very little variation in the intensity of fluorescence recorded for all three clones over

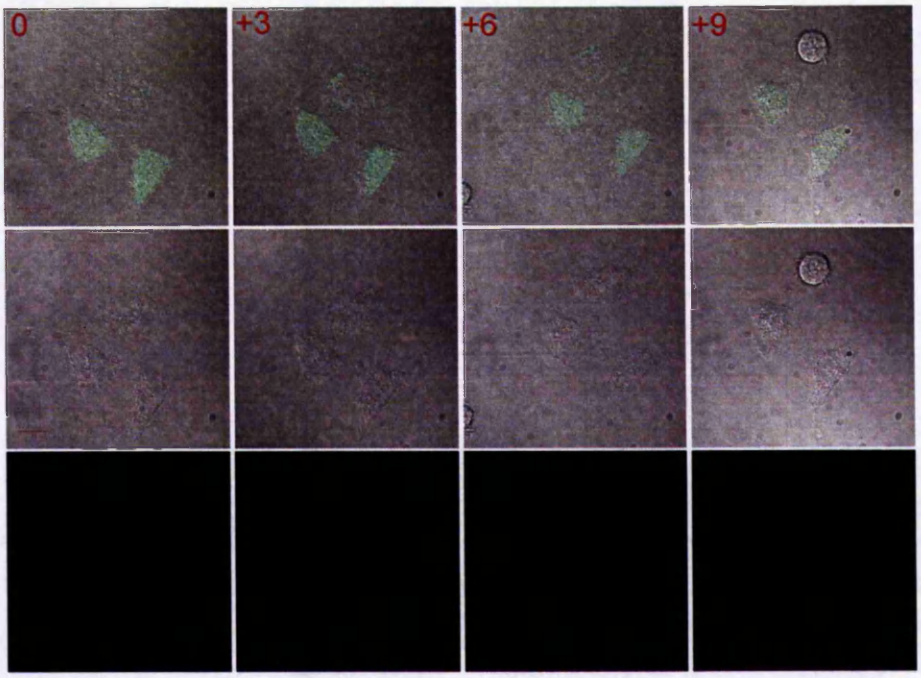


Figure 6.4. EGFP fluorescence remains constant during live cell imaging of clone B7. Cells from clone B7 (a 'low' EGFP expressing clone) were seeded into live cell dishes, synchronised with 2.5 μg ml^{-1} for 12 hours then released and allowed to divide before imaging commenced. Images were captured at 3 hour intervals. Green = EGFP.

one period of study. Cells in clone B4 had similar intensities of fluorescence to each other, as did cells in clone B7, even though the intensity level was different between B4 and B7. However, there was a massive difference in intensity between the cells in the variegating clone, B8 even though it did not appear to alter much over the period of study.

This stability in EGFP expression was not always seen in cells in clone B8, significantly, no major fluctuations were ever observed with clone B4 or clone B7 cells indicating that this was a property of the variegating phenotype. Observation of B8 cells revealed pronounced changes in fluorescence over the duration of the study. Cell 'A' had no detectable green fluorescence at the beginning of the time course. The cell divided between image capture at 7 and at 8 hours to produce two daughter cells also with no detectable green fluorescence. This condition is steady for the next 10 hours as no fluorescence due to EGFP was ever observed. Cells B1 and B2 are almost certainly daughters and are the product of a recent cell division. These two cells are possibly the most interesting cells in this experiment as there is total asymmetry

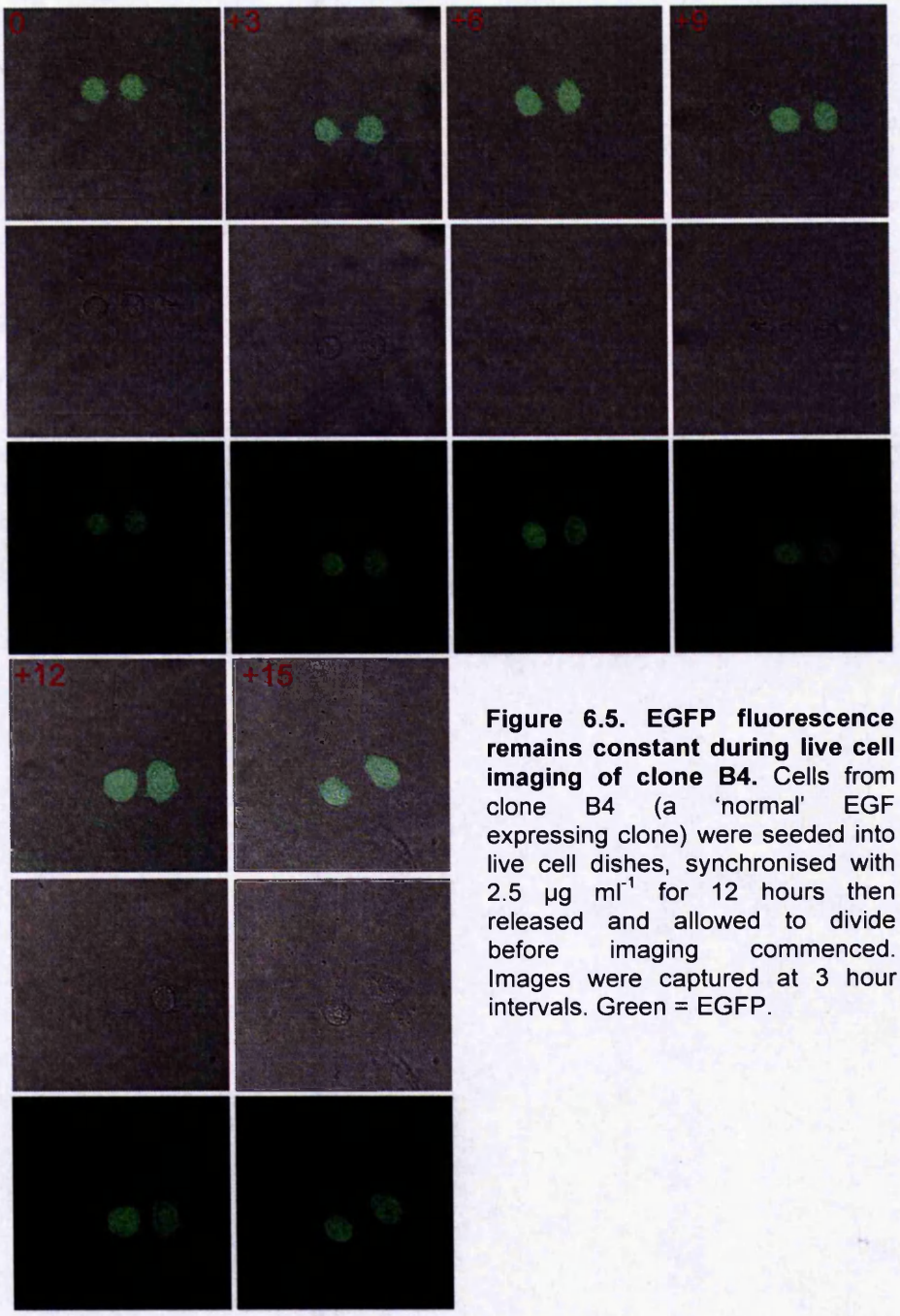


Figure 6.5. EGFP fluorescence remains constant during live cell imaging of clone B4. Cells from clone B4 (a 'normal' EGF expressing clone) were seeded into live cell dishes, synchronised with 2.5 μg ml^{-1} for 12 hours then released and allowed to divide before imaging commenced. Images were captured at 3 hour intervals. Green = EGFP.

in fluorescence at 0 hours – B1 had no detectable EGFP whereas B2 had very high levels of fluorescence. The high levels of EGFP in B2 remained constant throughout the period of observation, showing no fluctuation in 17 hours. Up to the +13 hour time point there was no fluorescence observed in cell B1, but subsequent images showed a steady rise in fluorescence over the remainder of the time course. The final level of fluorescence was still well below that of B2 and of other green cells in the field, though one can only speculate whether it

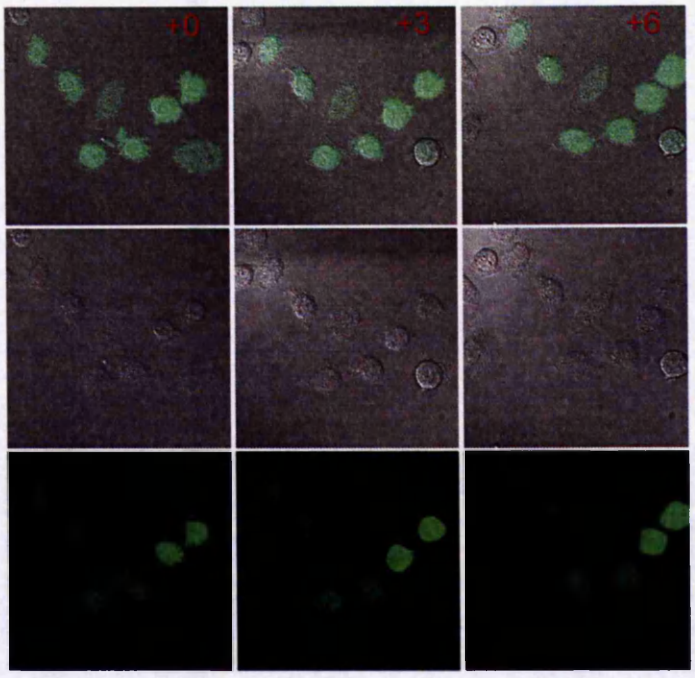
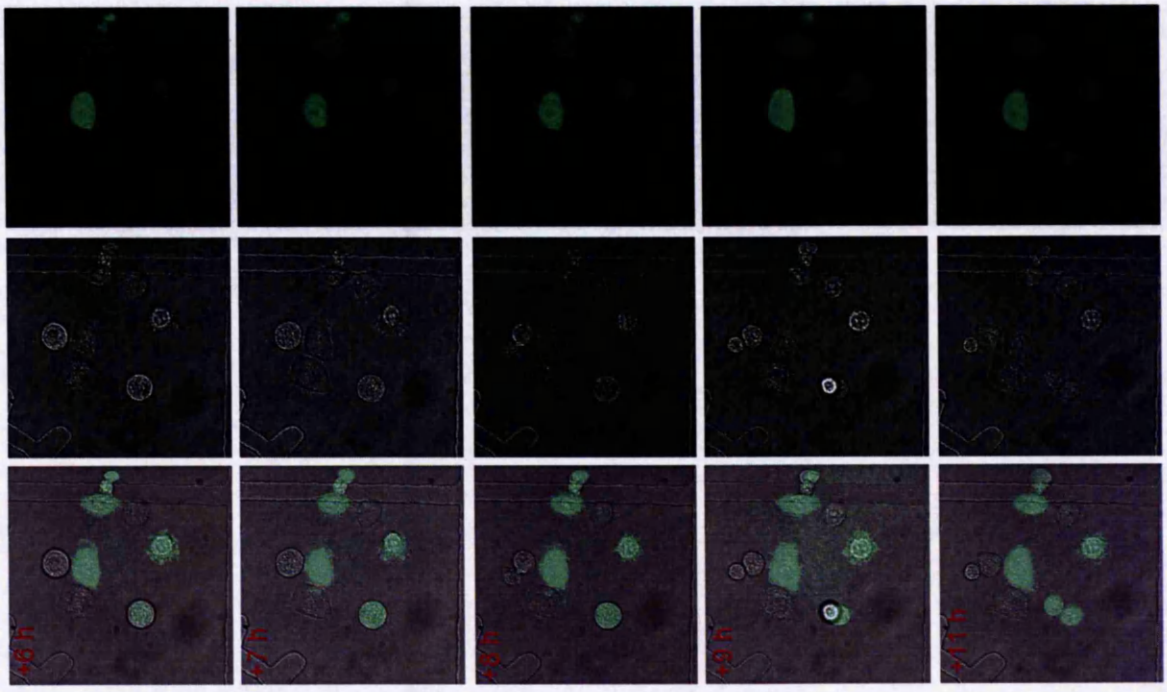


Figure 6.6. EGFP fluorescence remains constant during live cell imaging of clone B8. Cells from clone B8 (a 'variegating' clone) were seeded into live cell dishes, synchronised with 2.5 µg ml^{-1} for 12 hours then released and allowed to divide before imaging commenced. Images were captured at 3 hour intervals. Green = EGFP.

would have continued to rise until it reached the level of B2 in the period after the study concluded. Cell C had a 'normal' level of EGFP and it stayed steady throughout the period of study. Cell D had 'normal' amounts of fluorescence at the beginning of the study; it divided at about +9 hours and the resulting two daughters contained comparable levels of EGFP immediately post mitosis. This 'normal' level of fluorescence persisted in both daughter cells for the remainder of the observation period. Cell E had no detectable EGFP at any stage in the time course. As a control, to rule out the possibility that the imaging process was in some way causing these variations in expression in clone B8, cells from clone B4 were imaged concurrently. Figure 6.8 (a) shows that even though there was a modest difference in the fluorescence between the cells, the intensity remained constant for the duration of the study. Another field from the same culture dish showed a similar result (Figure 6.8 (b)); in this field the cell has consistent levels of fluorescence for the period of observation and undergoes mitosis at +14 hours giving rise to two daughters with similar EGFP levels.


+6 h
+7 h
+8 h
+9 h
+11 h

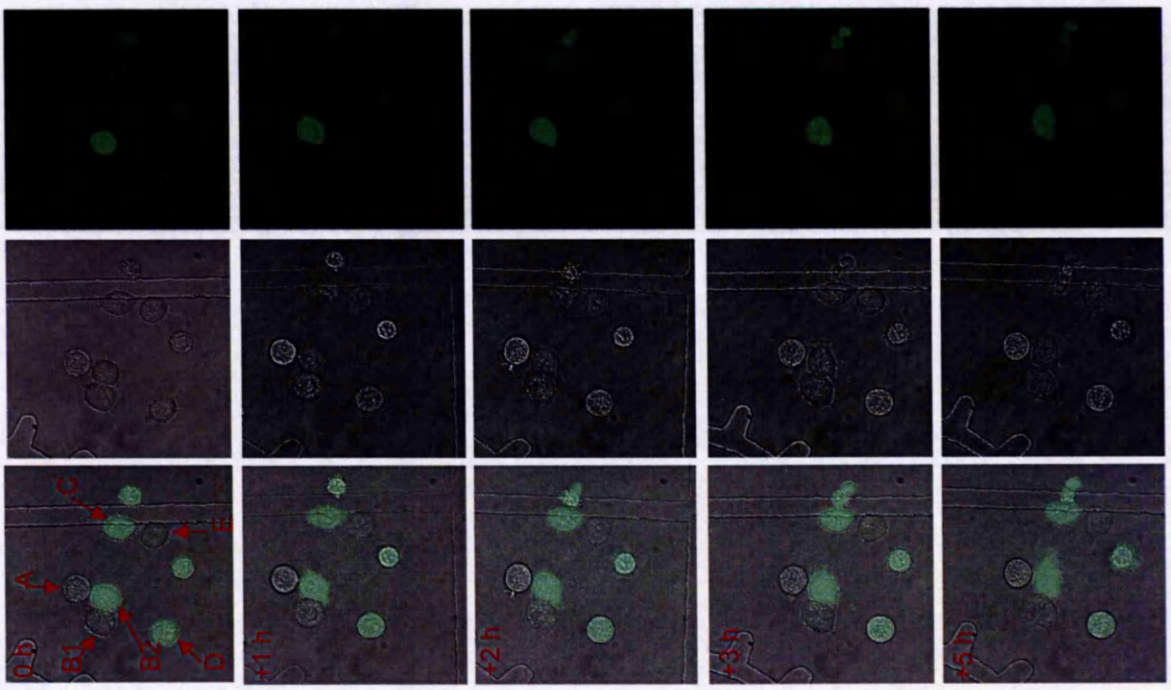

C
E
A
B1
B2
D
0 h
+1 h
+2 h
+3 h
+5 h

Figure 6.7. Analysis of real time EGFP levels fluctuating in live HeLa clone B8. Cells were seeded into live cell dishes, synchronised with 2.5 µg ml^{-1} for 12 hours then released. Image capture began as cells entered mitosis.

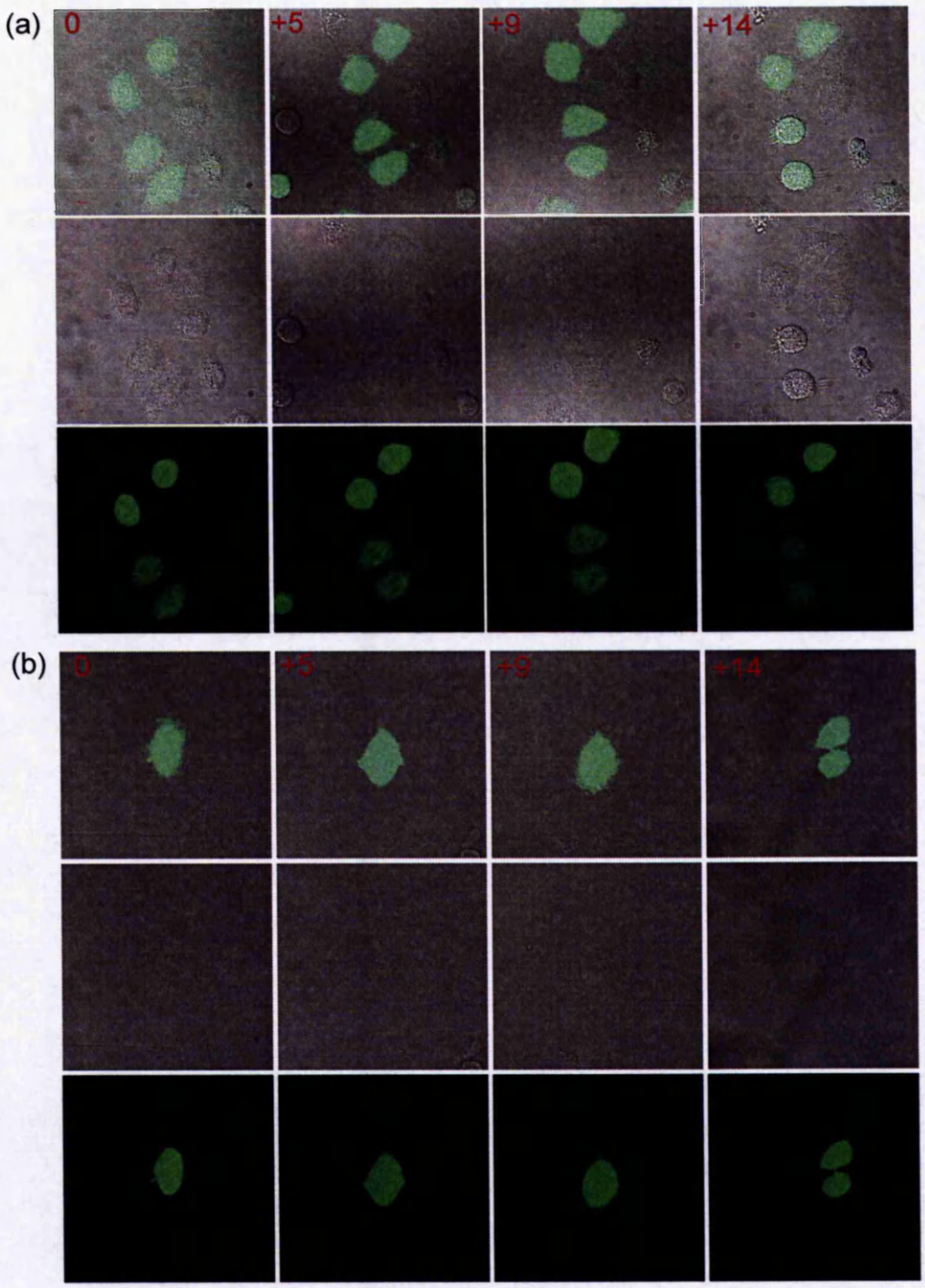


Figure 6.8. Timecourse of clone B4 EGFP expression. Cells were seeded into live cell dishes, synchronised with 2.5 μg ml^{-1} for 12 hours then released and allowed to divide before imaging commenced. (a) and (b) represent two different fields from the same live cell dish containing clone B4 cells. Green = EGFP. Images were captured alongside those in Figure 6.7, only an edited selection are shown here as representative of the data set.

6.3 Discussion

6.3.1 Determination of copy number using Immunostaining

It was ascertained that the signal produced by LacI-EGFP proteins binding to a 60-mer *lac*O insert in a pEPI-1 derivative was not sufficiently bright to be detected in established clones (section 5.2.4). By employing immunofluorescence against the EGFP component of the fusion protein, and the inherent amplification process involved, it was possible to directly label individual pEPI-1 derivative episomes in long-term clones. Fifty cells each clone were examined by CLSM and Z-stacks were generated, these data were analysed using the Imaris software in the 'Surpass' mode to label foci of a diameter greater than 0.5 μm. Table 6.1 shows the copy number for three clonal populations with pEPI-DsRed2-LamC-Lac60 and three with pEPI-DsRed2-H2B-Lac60 episomes. The mean copy number for each clone is similar to each other but slightly lower than pEPI-1 clones (approximately 4 per cell compared with approximately 6), and the copy number per cell in each clone exhibits the same variation as pEPI-1 episomes in established clones. Analyses were performed on early G1 cells to prevent any effect cell cycle would have on the copy numbers observed therefore, it is believed that copy numbers in the region of 10 per cell are real and not the product of a copy number of 5 that has been replicated and is in G2. It is unclear exactly why there is a lower mean copy number for stable clones with these pEPI-1 derivatives. It is possible that there is some influence of episome size, or as previously mentioned, the presence of a large insert of repetitive and/or bacterial DNA may be destabilising the mechanisms responsible for maintenance, expression or segregation meaning that fewer episomes become stably established. Chen *et al.* (2004) tested numerous vectors some with and some without bacterial sequences for the expression of a transgene when delivered to mice. They concluded that the minicircle forms of vectors, that is, those lacking the prokaryotic sequences, consistently expressed higher levels of the transgene. They propose that in many cases, the DNA silencing of transgenes is mediated by a covalent linkage of the expression cassette with bacterial DNA elements. This is supported by

the observation that these clones are unstable in the long-term and they begin to lose transgene expression around 8 weeks after the removal of selection.

6.3.2 Post-mitotic Distribution of Episomes

To investigate the nature of the variation in copy number and whether a segregation defect was involved in the general instability of these pEPI-1 derivatives, analyses were performed on daughter nuclei to assess the number of episomes present immediately after mitosis. The initial study was performed after one round of mitosis, on early G1 cells, and provided some interesting results (Figure 6.2 (b)). There appeared to be a relatively inefficient segregation process so that there was an equal distribution of molecules in only 32% of daughter pairs. There is a copy number difference of one or zero in 72% of cases, at first glance this gives the impression that the segregation process is reasonably efficient in nearly three quarters of mitoses, but there is a perturbing aspect to these observations. For there to be a segregation imbalance of one episome it would mean there had to be an odd number of episomes in the G2 nucleus. If all vector molecules replicated once and once only per cell cycle as is reportedly the case with the parent vector, pEPI-1 (Schaarschmidt *et al.* 2004) then there should never be an odd number of molecules in the cell at mitosis to allow an imbalance of one in the daughter nuclei. For example, a cell with a copy number of 5 would double this to 10 during S phase and so an unequal segregation into daughter cells should take the form of +2 or more and the total of molecules in both daughters should sum to 10. If this cell produced two daughters with an imbalance of one then it would mean that the copy number could not have been 2n at mitosis. The suggestion that there are an odd number of vector molecules at mitosis has serious implications for the future of this vector in gene therapy applications as this indicates there was a replication defect as well as a segregation one. How this defect manifests itself is unclear, unequal segregation and this apparent replication defect should lead to the gradual loss of the episomes from long-term culture. This is most definitely not the case in many published studies, for example, Jenke *et al.* (2004a) demonstrated the maintenance of expression (and by extension,

persistence of the episome) in all cells of a clone established with a pEPI-1 derivative in HaCat cells for more than 16 weeks. If there were such a serious deficiency in segregation and/or replication it would have been evident in a subpopulation of this clone by 16 weeks. This implies that any events that result in the loss of vector molecules must be balanced by an event that restores it, otherwise each time an episome is unreplicated the mean copy number in a population would drop, something which is not seen. The reasons why there is no drop off in copy number over time needs to be elucidated as a matter of urgency as future generations of this vector will be unsuitable for gene therapy if there are outstanding issues about faithful replication and segregation.

To investigate the uneven segregation of vector molecules at mitosis further analyses were performed on cells after two rounds of mitosis. It was found that there is a balancing of the mis-segregation when two sets of daughter pairs are compared. In the example presented in Figure 6.2 (a), the copy number in each nucleus is 5, 2, 2, 5. It has been assumed, due to the position of each cell relative to the other, that cells A and B are the product of one mitosis and C and D the product of the other. This would mean that even though there is an imbalance of 3 in the second set of daughter cells, the total number of episomes in cells A and B is 7 and the total in C and D is 7 indicating that there is some predictability to the uneven segregation. The key piece of information missing here is the copy number of the parent cells, 1, 2 and M (Figure 6.9). Was there an imbalance in copy number between cells '1' and '2' and how did a total copy number of 7 (in AB and CD) come into being? The use of live cell imaging with the LacI-GFP bound to *lac*O in pEPI-1 derivatives could have helped to answer these questions as cells could be tracked through multiple cell cycles and the efficacy of replication and segregation would be able to be fully assessed.

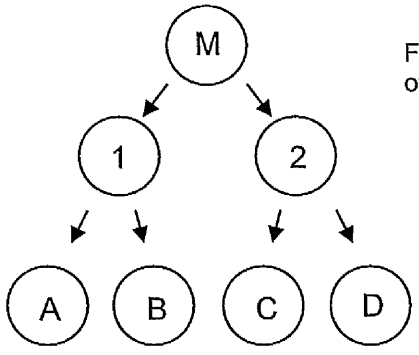


Figure 6.9. Hypothetical segregation profile of the cells imaged in Figure 6.2 (a).

6.3.3 Episomes appear to be generally localised to areas of low chromatin density

The reporter gene of pEPI-DsRed2-H2B-Lac60, that is, DsRed2-H2B was exploited to provide information regarding the nuclear localisation of episomes in established clones. The immunostaining of LacI-EGFP molecules bound to the *lac*O arrays of the episomes allowed a comparison of their location with the general distribution of H2B, taken as being representative of the chromatin as a whole. The DsRed2-H2B appears to form less aggregates than DsRed2-LaminC and a clearly defined chromatin structure can be seen. However, there remains the possibility that the fusion is somehow disrupting the normal nuclear architecture and so reduces how meaningful these observations might be. It would be relatively easy to determine whether the replication pattern has been perturbed in these cells by staining with 5-bromo-2-deoxy-uridine and checking whether there is a normal distribution of early and late replicating foci. If this were normal then it is likely the cell is functioning normally. However, clones expressing DsRed2-H2B appear to grow normally and they maintain expression of the transgene for more than 8 weeks in the absence of selection, if the fusion protein was so damaging then it would be unlikely the cells would remain healthy for that length of time.

Figure 6.3 shows representative images of Z-stacks from two clones, EPI-DHL-1 and EPI-DHL-5. The enlarged images from the gallery suggest that the episomes are associated with areas of less dense red fluorescence. This lower density of DsRed2-H2B (and by extension, lower density of chromatin) could be for two reasons, firstly, it may represent the IC or secondly it may represent decondensed chromatin. Either way this suggests that the majority of the episomes are in contact with 'active' regions of the nucleus, as transcription occurs in the IC, with active gene domains looping out of the chromosome territories and a decondensed open chromatin structure within a chromosome territory would also be enriched for many activating factors and provide an environment refractory to silencing.

An interesting comparison can be made between the two clones, EPI-DHL-5 expresses lower levels of DsRed2-H2B than EPI-DHL-1 and it also appears to have more episomes in a more nuclear peripheral position. A more robust

analysis of the link between the transgene expression and the location of the episomes with respect to the nuclear periphery would need to be performed before any firm conclusions can be drawn from these observations.

6.3.4 Investigating EGFP intensity in live cells

Stable pEPI-1$^+$ clones were tracked at regular intervals for more than 24 hours to assess the variation in EGFP expression. In general, there is very little variation in expression levels in cells from the 'low' or 'normal' phenotypes (Figures 6.4 and 6.5). The variegating phenotype demonstrated pronounced differences in EGFP expressed in some cells in relatively short time frames. Figure 6.7 demonstrates many different aspects of the expression of the reporter gene in this clone. There are cells that (divide and) produce no detectable levels of EGFP over the 18 hours of observation, there are other cells that divide and appear to equally segregate the ability to express EGFP, and there is one daughter pair that has a complete asymmetry in expression. Even more remarkable is the steady increase in EGFP after about 14 hours of observation. It is unlikely that this upregulation in transcription is due to a time lag imposed by mitosis, either because of an unequal segregation of mRNA or a transcriptional delay in one daughter cell following mitosis as the cells are halfway through the next cycle. The increase in expression does not seem to occur early enough to be a product of an uneven mitosis. This begins to characterise this curious sub-set of established clones by giving a temporal aspect to observations gained from analysis of fixed cells in Chapter Four.

It could be suggested that the variations in green fluorescence, and by extension, level of EGFP expressed, are due to random fluctuations in gene expression that are seen normally in the endogenous gene expression levels. However, there are a number of pieces of evidence that suggest it is not the case and that the variation in EGFP levels is a real feature of the variegating phenotype of established pEPI-1 HeLa clones. In one study, HeLa cells were transfected with pEGFP-C1 (the backbone of pEPI-1) and stable (integrated) clones were obtained. The mean fluorescence was measured at various points in the cell cycle and was shown to vary by a factor of less than 2 (Brunner *et al.*

2000). This demonstrates that, in some contexts at least, there is little variation in the activity of the CMV promoter across the different phases of the cell cycle making it unlikely that the variation seen here is due to cell cycle factors. In addition to this, the cells were synchronised so that even if there were cell cycle variations, it would be expected that a higher proportion of cells in the study would exhibit fluctuations, and in a similar timeframe to each other.

Another suggested cause of this variable gene expression pattern could be purely 'noise' (Raser & O'Shea, 2005). 'Noise' can occur due to the inherently random nature of chemical reactions within a cell when only a few molecules or targets of a specific type exist. This phenomenon could account for the diversity exhibited by genetically identical cells or organisms, even when they have identical histories of environmental exposure. The impact of this randomness can be seen in the difference of expression of two 'identical' genes in yeast (Raser & O'Shea, 2004). CFP and YFP reporter proteins were expressed from two identical promoters integrated at the same locus on homologous chromosomes in *S. cerevisiae* used in the diploid form. By comparing the difference in expression profiles for the two reporter genes in the same cell and in the levels between cells, it was shown that 20% of the variation in gene expression could be directly attributed to 'intrinsic noise', that is to random, stochastic processes within the cell. It is possible that the same could be true for the different pEPI-1 episomes, however, the lack of any significant variation in gene expression levels in the low or normal clones suggests that something 'different' is happening in the variegating clones. Further to this, a recent study has shown that it is unlikely such small fluctuations in gene expression levels will be manifested as an observable difference in green fluorescence (Raj *et al.* 2006). The number of mRNA molecules produced by an integrated reporter gene were counted by FISH. They found there were massive variations in the numbers of mRNA molecules between cells and that the variations occurred because the mRNAs were synthesised in short but intense bursts of transcription. These bursts of transcription correlate with the transition of the gene from an inactive state to an active one and end when they transition back to the inactive state. It was further shown that these bursts were intrinsically random and were not due to external factors, such as the levels of transcriptional activators. However, the bursts in mRNA levels were also shown

to be buffered by the slow degradation rate of the reporter gene (EGFP). This makes it unlikely that a random, natural fluctuation is responsible for the variation in fluorescence in the clones studied, as on this scale any differences would be masked by steady EGFP levels.

6.3.5 Conclusions

The variegating clones may be an extreme version of stochastic events, where a number of the episomes are in a more compact, less active region of the nucleus and the number of interacting activating factors will be higher than for the episomes in the more active, open regions. To compound this problem for the variegating episomes, there is likely to be fewer available general activating factors in the less active regions making the chances of a successful interaction fewer. When it does occur, however, the chromatin structure will be opened and the histone modifications altered making the transition to the 'on' state more likely whilst this open configuration is maintained. Cell division or replication may interfere with the open, active domain formed on the variegating pEPI-1 episomes and so return it to the basal state of more compact and less active.
The observation that these derivatives of pEPI-1 are unstable in the long-term is troubling because mitotic stability is a critical component of any future gene therapy vector. It is essential that future studies address this issue and establish whether it is a product of the increased size of the vector or the bacterial repeat elements that cause the loss of expression, and in some cases, the loss of the vector itself.

Chapter Seven: Discussion

7.1 Conclusions

This study aimed to determine the mechanism of establishment of pEPI-1 as a stably maintained episome in HeLa cells and to investigate the factors involved with this behaviour. This was attempted in a number of ways; the establishment phase was investigated with the intention of improving the efficiency of this step; the stable phase was examined in both fixed and live cells to characterise the expression profile of the reporter gene and provide information regarding vector copy number; finally, individual episomes were labelled in an attempt to investigate their nuclear localisation. There was a degree of success with respect to determining the factors involved in the establishment phase in the sense that a number were discounted as having no effect leading the to conclusion, in conjunction with work published by the Lipps group (for example, Stehle *et al.* 2003; Jenke *et al.* 2004b), that it is the evolution of a set of chromatin modifications due to transcription and replication of the vector that is most likely to be the key factor in the generation of stable pEPI-1^{+} HeLa cells. It is therefore essential that any future studies concerning the establishment phase of this vector address this issue and more fully define the complex interplay between transcription and replication and how they work to create an 'episomal' set of chromatin modifications, the timescale of this evolution, and precisely what the modifications are.

An important observation from this study is the apparent lack of requirement for drug selection during the establishment phase. This makes the future use of this vector system more attractive as it would be undesirable to have to have to administer drug selection for 3-4 weeks or to have the bacterial components present in the construct.

The copy number is not the same in clonal HeLa cells, some cells having four times the number as others; this is in agreement with a previous study concerning the vector in CHO cells (Baiker *et al.* 2000) and this raises many questions. The clonal populations maintain the episome for more than 100 generations with a mitotic stability of 99%, yet analysis of early G1 daughter

cells reliably showed a disparity in copy number. It is unclear at this stage how, even though there does not appear to be a faithful 50-50 segregation of vectors at mitosis, the population maintains the episomes in long-term culture without selection pressure. A more robust understanding of how copy number varies between clonal cells, and its effect on the expression of the reporter gene must be carried out if this system is to be used for gene therapy as the production of therapeutic protein must be under tight control.

This study has shown that the level of EGFP expression varies between clonal cells and investigation of the 'variegating' phenotype suggests that gene expression is not solely dependent on copy number; the cells in this phenotype had a comparable copy number to those expressing 'normal' levels of EGFP yet some cells had more than double the EGFP intensity. Further investigation is needed to determine precisely the difference between these cells and a good place to begin would be to compare the chromatin modifications present on episomes from cells expressing low, normal and high levels of reporter gene; Papapetrou *et al.* (2006) restored EGFP expression to MEL cells by treating them with sodium butyrate, a HDAC inhibitor, showing that expression of the reporter gene on the vector can be influenced by altering the chromatin modifications present.

7.2 Future Directions

This vector system potentially has great value in the field of gene therapy; the removal of viral and bacterial components would avoid many of the issues surrounding current vectors such as transformation or provoking an immune response. However, there are many outstanding issues that need to be resolved before pEPI-1 has any therapeutic value such as the extremely low efficiency of stable clone formation and the variation in gene expression and copy number in clonal cells. Even if pEPI-derivatives are eventually unsuitable for gene therapy, it could still be of great value as a minimal system for studying the interplay between chromatin modifications and nuclear localisation.

In terms of developing pEPI-derivatives as future gene therapy vectors, it is essential that it be clear which cells support stable long-term expression; a

comprehensive study of reporter gene expression in as many different human cell types as possible needs to be undertaken so that the limits of the system become apparent. The ability of this vector to establish episomal status in different primary cells should be investigated as well as using transformed and immortalised cells. Can the CMV promoter be substituted for cell-type specific promoters to allow targeted expression, or an inducible promoter to confer regulated expression? And can cells be transfected *in vivo*?

Future work should focus on improving the efficiency of stable clone formation. If this phase is dependent on the evolution of a specific set of chromatin modifications then can this process be enhanced? The addition of a HDAC inhibitor restored EGFP expression in MEL cells (Papapetrou *et al.* 2006); could the addition of compounds that affect the post-translational modification of chromatin at specific time points in the establishment phase promote the switch from plasmid to episome?

Another area which requires greater investigation is the nature of the expression cassette; what is the maximum insert size that can be supported by one S/MAR? Many genes and their associated regulatory sequences are much larger than the EGFP reporter gene currently in place and so it would be expected that the vector would need to be tens or hundreds of kilobases in size. The vector appears to support two separate expression cassettes at present, EGFP and Kan/Neo^r, what is the stoichiometry of expression? The use of two reporter genes under the control of identical promoters would help to determine the control of gene expression across the construct. Do different promoters confer different levels of reporter gene expression? Substituting pCMV for other commonly used promoters would help to answer this. Taken one step further, if there are differences in expression due to different promoters, will two different promoters in the same construct express reporter genes at different levels? This could have great therapeutic value as numerous subunits of a protein or complex could be encoded on the same vector to provide defined ratios of each.

Finally, another area that would be interesting to investigate is whether there a pattern to the nuclear localisation of the episomes. The S/MAR used in this study was from the 5'-region of the human β-interferon gene. The S/MAR is normally found on chromosome 9p21 and it is tempting to speculate that the

localisation of the episomes may be determined by this. Using FISH to visualise different regions of chromosomes in conjunction with pEPI-1 FISH would help to determine whether the vector has a preferred location. It would be interesting to see if the nuclear localisation alters if the β-interferon S/MAR is substituted for another from a different chromosome region.

This form of episomal replicon is a relatively new one and so there are many more questions to be asked than are answered at present. However, careful characterisation of the system should provide a wealth of information about the interplay between nuclear localisation and chromatin structure, and we hope it will also serve as the starting point for a therapeutically viable gene therapy vector, the value of which is evident.

Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y., and Kohwi-Shigematsu, T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes & Development 14*, 521-535.

Baiker, A., Maercker, C., Piechaczek, C., Schmidt, S.B.A., Bode, J., Benham, C., and Lipps, H.J. (2000). Mitotic stability of an episomal vector containing a human scaffold/matrix-attached region is provided by association with nuclear matrix. *Nature Cell Biology 2*, 182-184.

Ballestas, M.E., Chatis, P.A., and Kaye, K.M. (1999). Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen. *Science 284*, 641-644.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature 410*, 120-124.

Baxter, J., Merkenschlager, M., and Fisher, A.G. (2002). Nuclear organisation and gene expression. *Current Opinion in Cell Biology 14*, 372-376.

Bell, A.C., West, A.G., and Felsenfeld, G. (2001). Gene regulation - Insulators and boundaries: Versatile regulatory elements in the eukaryotic genome. *Science 291*, 447-450.

Belmont, A.S. (2001). Visualizing chromosome dynamics with GFP. *Trends in Cell Biology 11*, 250-257.

Belmont, A.S., Dietzel, S., Nye, A.C., Strukov, Y.G., and Tumbar, T. (1999). Large-scale chromatin structure and function. *Current Opinion in Cell Biology 11*, 307-311.

Belmont, A.S., and Straight, A.F. (1998). *In vivo* visualization of chromosomes using *lac* operator-repressor binding. *Trends in Cell Biology 8*, 121-124.

Berezney, R. (2002). Regulating the mammalian genome: the role of nuclear architecture. *Advances in Enzyme Regulation 42*, 39-52.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. *Nature 447*, 407-412.

Bode, J., Benham, C., Knopp, A., and Mielke, C. (2000). Transcriptional augmentation: Modulation of gene expression by Scaffold/Matrix-Attached regions (S/MAR elements). *Critical Reviews in Eukaryotic Gene Expression 10*, 73-90.

Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C., and Kohwishigematsu, T. (1992). Biological Significance of Unwinding Capability of Nuclear Matrix Associating DNAs. *Science 255*, 195-197.

Bode, J., and Maass, K. (1988). Chromatin Domain Surrounding the Human Interferon-β Gene as Defined by Scaffold-Attached Regions. *Biochemistry 27*, 4706-4711.

Bonnerot, C., Grimber, G., Briand, P., and Nicolas, J.F. (1990). Patterns of Expression of Position-Dependent Integrated Transgenes in Mouse Embryo. *Proceedings of the National Academy of Sciences of the United States of America 87*, 6331-6335.

Boussadia, O., Amiot, F., Cases, S., Triqueneaux, G., Jacquemin-Sablon, H., and Dautry, F. (1997). Transcription of *unr* (upstream of N-*ras*) down-modulates N-*ras* expression *in vivo*. *FEBS Letters 420*, 20-24.

Brantl, S., and Wagner, E.G.H. (1997). Dual function of the copR gene product of plasmid pIP501. *Journal of Bacteriology 179*, 7016-7024.

Bridger, J.M., Boyle, S., Kill, I.R., and Bickmore, W.A. (2000). Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. *Current Biology 10*, 149-152.

Brunner, S., Sauer, T., Carotta, S., Cotten, M., Saltik, M., and Wagner, E. (2000). Cell cycle dependence of gene transfer by lipoplex polyplex and recombinant adenovirus. *Gene Therapy 7*, 401-407.

Bulger, M., Sawado, T., Schubeler, D., and Groudine, M. (2002). ChIPs of the β-globin locus: unravelling gene regulation within an active domain. *Current Opinion in Genetics & Development 12*, 170-177.

Bulger, M., Schubeler, D., Bender, M.A., Hamilton, J., Farrell, C.M., Hardison, R.C., and Groudine, M. (2003). A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse β- globin locus. *Molecular and Cellular Biology 23*, 5234-5244.

Cai, S.T., Han, H.J., and Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nature Genetics 34*, 42-51.

Calos, M.P. (1996). The potential of extrachromosomal replicating vectors for gene therapy. *Trends in Genetics 12*, 463-466.

Cavazzana-Calvo, M., Hacein-Bey, S., Basile, C.D., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., Bousso, P., Le Deist, F., and Fischer, A. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science 288*, 669-672.

Chao, M.V., Martinson, H.G., and Gralla, J.D. (1980). *Lac* Operator Nucleosomes .2. *Lac* Nucleosomes Can Change Conformation to Strengthen Binding by *Lac* Repressor. *Biochemistry 19*, 3260-3269.

Chen, J., and Matthews, K.S. (1992). Deletion of Lactose Repressor Carboxyl-Terminal Domain Affects Tetramer Formation. *Journal of Biological Chemistry 267*, 13843-13850.

Chen, Z.Y., He, C.Y., Meuse, L., and Kay, M.A. (2004). Silencing of episomal transgene expression by plasmid bacterial DNA elements *in vivo*. *Gene Therapy 11*, 856-864.

Cheung, P., Allis, C.D., and Sassone-Corsi, P. (2000). Signalling to chromatin through histone modifications. *Cell 103*, 263-271.

Chubb, J.R., Boyle, S., Perry, P., and Bickmore, W.A. (2002). Chromatin motion is constrained by association with nuclear compartments in human cells. *Current Biology 12*, 439-445.

Chung, J.H., Bell, A.C., and Felsenfeld, G. (1997). Characterization of the chicken β-globin insulator. *Proceedings of the National Academy of Sciences of the United States of America 94*, 575-580.

Cockell, M., and Gasser, S.M. (1999). Nuclear compartments and gene regulation. *Current Opinion in Genetics & Development 9*, 199-205.

Collas, P. (1998). Modulation of plasmid DNA methylation and expression in zebrafish embryos. *Nucleic Acids Research 26*, 4454-4461.

Cooper, M.J., Lippa, M., Payne, J.M., Hatzivassiliou, G., Reifenberg, E., Fayazi, B., Perales, J.C., Morrison, L.J., Templeton, D., Piekarz, R.L., and Tan, J. (1997). Safety-modified episomal vectors for human gene therapy. *Proceedings of the National Academy of Sciences of the United States of America 94*, 6450-6455.

Cossons, N., Nielsen, T.O., Dini, C., Tomilin, N., Young, D.B., Riabowol, K.T., Rattner, J.B., Johnston, R.N., ZannisHadjopoulos, M., and Price, G.B. (1997). Circular YAC vectors containing a small mammalian origin sequence can associate with the nuclear matrix. *Journal of Cellular Biochemistry 67*, 439-450.

Craig, J.M., Boyle, S., Perry, P., and Bickmore, W.A. (1997). Scaffold attachments within the human genome. *Journal of Cell Science 110*, 2673-2682.

Cremer, T., Kreth, G., Koester, H., Fink, R.H.A., Heintzmann, R., Cremer, M., Solovei, I., Zink, D., and Cremer, C. (2000). Chromosome territories, interchromatin domain compartment, and nuclear matrix: An integrated view of the functional nuclear architecture. *Critical Reviews in Eukaryotic Gene Expression 10*, 179-212.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *Journal of Cell Biology 145*, 1119-1131.

Dean, D.A., Strong, D.D., and Zimmer, W.E. (2005). Nuclear entry of nonviral vectors. *Gene Therapy 12*, 881-890.

Dodd, I.B., and Egan, J.B. (2002). Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. *Molecular Microbiology 45*, 697-710.

Escriou, V., Carriere, M., Bussone, F., Wils, P., and Scherman, D. (2001). Critical assessment of the nuclear import of plasmid during cationic lipid-mediated gene transfer. *Journal of Gene Medicine 3*, 179-187.

Eszterhas, S.K., Bouhassira, E.E., Martin, D.I.K., and Fiering, S. (2002). Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Molecular and Cellular Biology 22*, 469-479.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990). A Deletion of the Human β-Globin Locus Activation Region Causes a Major Alteration in Chromatin Structure and Replication across the Entire β-Globin Locus. *Genes & Development 4*, 1637-1649.

Fukuda, Y., and Nishikawa, S. (2003). Matrix attachment regions enhance transcription of a downstream transgene and the accessibility of its promoter region to micrococcal nuclease. *Plant Molecular Biology 51*, 665-675.

Gasser, S.M. (2001). Positions of potential: Nuclear organization and gene expression. *Cell 104*, 639-642.

Gerasimova, T.I., and Corces, V.G. (1998). Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell 92*, 511-521.

Geyer, P.K. (1997). The role of insulator elements in defining domains of gene expression. *Current Opinion in Genetics & Development 7*, 242-248.

Ghosh, S.S., Gopinath, P., and Ramesh, A. (2006). Adenoviral vectors - A promising tool for gene therapy. *Applied Biochemistry and Biotechnology 133*, 9-29.

Grimes, B.R. and Monaco, Z.L. (2005). Artificial and engineered chromosomes: developments and prospects for gene therapy. *Chromosoma 114*, 230-241.

Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J., Thrasher, A.J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., Fischer, A., Cavazzana-Calvo, M., Davies, E.G., Kuis, W., Lundlaan, W.H.K., and Leiva, L. (2002). Sustained correction of X-linked severe combined immunodeficiency by *ex vivo* gene therapy. *New England Journal of Medicine 346*, 1185-1193.

Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M., and Fischer, A. (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *New England Journal of Medicine 348*, 255-256.

Hart, C.M., and Laemmli, U.K. (1998). Facilitation of chromatin dynamics by SARs. *Current Opinion in Genetics & Development 8*, 519-525.

Heng, H.H.Q., Goetze, S., Ye, C.J., Liu, G., Stevens, J.B., Bremer, S.W., Wykes, S.M., Bode, J., and Krawetz, S.A. (2004). Chromatin loops are selectively anchored using scaffold/matrix- attachment regions. *Journal of Cell Science 117*, 999-1008.

Heun, P., Laroche, T., Shimada, K., Furrer, P., and Gasser, S.M. (2001). Chromosome dynamics in the yeast interphase nucleus. *Science 294*, 2181-2186.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. *Journal of Molecular Biology 26*, 365-369.

Houtsmuller, A.B., and Vermeulen, W. (2001). Macromolecular dynamics in living cell nuclei revealed by fluorescence redistribution after photobleaching. *Histochemistry and Cell Biology 115*, 13-21.

Hu, X., Eszaterhas, S., Pallazzi, N., Bouhassira, E.E., Fields, J., Tanabe, O., Gerber, S.A., Bulger, M., Engel, J.D., Groudine, M., and Fiering, S. (2007) Transcriptional interference among the murine β-like globin genes. *Blood 109* 2210-2216.

Ikeno, M., Inagaki, H., Nagata, K., Morita, M., Ichinose, H., and Okazaki, T. (2002). Generation of human artificial chromosomes expressing naturally controlled guanosine triphosphate cyclohydrolase I gene. *Genes to Cells 7,* 1021-1032.

Jackson, D.A., Yuan, J., and Cook, P.R. (1988). A Gentle Method for Preparing Cyto-Skeletons and Nucleo-Skeletons and Associated Chromatin. *Journal of Cell Science 90*, 365-378.

Jagura-Burdzy, G., and Thomas, C.M. (1997). Dissection of the switch between genes for replication and transfer of promiscuous plasmid RK2: Basis of the dominance of trfAp over trbAp and specificity for KorA in controlling the switch. *Journal of Molecular Biology 265*, 507-518.

Jenke, A.C.W., Scinteie, M.F., Stehle, I.M., and Lipps, H.J. (2004a). Expression of a transgene encoded on a non-viral episomal vector is not subject to epigenetic silencing by cytosine methylation. *Molecular Biology Reports 31*, 85-90.

Jenke, A.C.W., Stehle, I.M., Herrmann, F., Eisenberger, T., Baiker, A., Bode, J., Fackelmayer, F.O., and Lipps, H.J. (2004b). Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. *Proceedings of the National Academy of Sciences of the United States of America 101*, 11322-11327.

Jenke, B.H.C., Fetzer, C.P., Stehle, I.M., Jonsson, F., Fackelmayer, F.O., Conradt, H., Bode, J., and Lipps, H.J. (2002). An episomally replicating vector binds to the nuclear matrix protein SAF-A in vivo. *EMBO Reports 3*, 349-354.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science 293*, 1074-1080.

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

Kellum, R., and Elgin, S.C.R. (1998). Chromatin boundaries: Punctuating the genome. *Current Biology 8*, R521-R524.

Kimura, M., and Yamaguchi, I. (1998). Convergent transcription units and their promoters at both ends of Pot2, an inverted repeat transposon from the rice blast fungus. *Journal of Biochemistry 124*, 268-273.

Kipp, M., Gohring, F., Ostendorp, T., van Drunen, C.M., van Driel, R., Przybylski, M., and Fackelmayer, F.O. (2000). SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Molecular and Cellular Biology 20*, 7480-7489.

Kurz, A., Lampel, S., Nickolenko, J.E., Bradl, J., Benner, A., Zirbel, R.M., Cremer, T., and Lichter, P. (1996). Active and inactive genes localize preferentially in the periphery of chromosome territories. *Journal of Cell Biology 135*, 1195-1205.

Lehman, C.W., and Botchan, M.R. (1998). Segregation of viral plasmids depends on tethering to chromosomes and is regulated by phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America 95*, 4338-4343.

Lei, D.C., Kunzelmann, K., Koslowsky, T., Yezzi, M.J., Escobar, L.C., Xu, Z., Ellison, A.R., Rommens, J.M., Tsui, L-C., Tykocinski, M., and Gruenert, D.C. (1996). Episomal expression of wild-type CFTR corrects cAMP-dependent chloride transport in respiratory epithelial cells. *Gene Therapy 3*, 427-436.

Leight, E.R., and Sugden, B. (2001a). The *cis*-acting family of repeats can inhibit as well as stimulate establishment of an oriP replicon. *Journal of Virology 75*, 10709-10720.

Leight, E.R., and Sugden, B. (2001b). Establishment of an oriP replicon is dependent upon an infrequent, epigenetic event. *Molecular and Cellular Biology 21*, 4149-4161.

Li, G., Sudlow, G., and Belmont, A.S. (1998). Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: Precise choreography of condensation/decondensation and nuclear positioning. *Journal of Cell Biology 140*, 975-989.

Liebich, I., Bode, J., Reuter, I., and Wingender, E. (2002). Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Research 30*, 3433-3442.

Lipps, H.J., Jenke, A.C.W., Nehlsen, K., Scinteie, M.F., Stehle, I.M., and Bode, J. (2003). Chromosome-based vectors for gene therapy. *Gene 304*, 23-33.

Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D., and Felsenfeld, G. (2001). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science 293*, 2453-2455.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature 389*, 251-260..

Mackey, D., and Sugden, B. (1999). The linking regions of EBNA1 are essential for its support of replication and transcription. *Molecular and Cellular Biology 19*, 3349-3359.

Manzini, S., Vargiolu, A., Stehle, I.M., Bacci, M.L., Cerrito, M.G., Giovannoni, R., Zannoni, A., Bianco, M.R., Forni, M., Donini, P., Papa, M., Lipps, H.J., and Lavitrano, M. (2006). Genetically modified pigs produced with a nonviral episomal vector. *Proceedings of the National Academy of Sciences of the United States of America 103*, 17672-17677.

Marshall, W.F., Straight, A., Marko, J.F., Swedlow, J., Dernburg, A., Belmont, A.S., Murray, A.W., Agard, D.A., and Sedat, J.W. (1997). Interphase chromosomes undergo constrained diffusional motion in living cells. *Current Biology 7*, 930-939.

Martin, D.I.K., Fiering, S., and Groudine, M. (1996). Regulation of β-globin gene expression: Straightening out the locus. *Current Opinion in Genetics & Development 6*, 488-495.

Mecsas, J. and Sugden, B. (1987). Replication of plasmids derived from bovine papilloma virus type 1 and Epstein-Barr virus in cells in culture. *Annual Review of Cell Biology 3,* 87-108.

McKnight, R.A., Shamay, A., Sankaran, L., Wall, R.J., and Henninghausen, L. (1992). Matrix-Attachment Regions Can Impart Position-Independent Regulation of a Tissue-Specific Gene in Transgenic Mice. *Proceedings of the National Academy of Sciences of the United States of America 89*, 6943-6947.

McMorrow, T., van den Wijngaard, A., Wollenschlaeger, A., van de Coreut, M., Monkhorst, M., Trimborn, T., Fraser, P., van Lohuizen, M., Jenuwein, T., Djabali, M., Philipsen, S., Grosveld, F., and Milot, E. (2000). Activation of the β-globin locus by transcription factors and chromatin modifiers. *EMBO Journal 19*, 4986-4996.

Miller, J.H., and Reznikoff, W.A. (1980). The Operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 177-220.

Monaco, Z.L., and Moralli, D. (2006). Progress in artificial chromosome technology. *Biochemical Society Transactions 34*, 324-327.

Nakagomi, K., Kohwi, Y., Dickinson, L.A., and Kohwishigematsu, T. (1994). A Novel DNA-Binding Motif in the Nuclear Matrix Attachment DNA- Binding Protein SATB1. *Molecular and Cellular Biology 14*, 1852-1860.

Nehlsen, K., Broll, S., and Bode, J. (2006). Replicating minicircles: Generation of nonviral episomes for the efficient modification of dividing cells. *Gene Therapy and Molecular Biology 10B*, 233-243.

Nickerson, J.A. (2001). Experimental observations of a nuclear matrix. *Journal of Cell Science 114*, 463-474.

Niswender, K.D., Blackman, S.M., Rohde, L., Magnuson, M.A., and Piston, D.W. (1995). Quantitative imaging of green fluorescent protein in cultured cells-comparison of microscope techniques, use in fusion proteins and detection limits. *Journal of Microscopy 180*, 109-116.

Ostermeier, G.C., Liu, Z.D., Martins, R.P., Bharadwaj, R.R., Ellis, J., Draghici, S., and Krawetz, S.A. (2003). Nuclear matrix association of the human β-globin locus utilizing a novel approach to quantitative real-time PCR. *Nucleic Acids Research 31*, 3257-3266.

Papapetrou, E.P., Ziros, P.G., Micheva, I.D., Zoumbos, N.C., and Athanassiadou, A. (2006). Gene transfer into human hematopoietic progenitor cells with an episomal vector carrying an S/MAR element. *Gene Therapy 13*, 40-51.

Park, P.C., and De Boni, U. (1999). Dynamics of structure-function relationships in interphase nuclei. *Life Sciences 64*, 1703-1718.

Pemov, A., Bavykin, S., and Hamlin, J.L. (1998). Attachment to the nuclear matrix mediates specific alterations in chromatin structure. *Proceedings of the National Academy of Sciences of the United States of America 95*, 14757-14762.

Phi-Van, L., and Stratling, W.H. (1988). The Matrix Attachment Regions of the Chicken Lysozyme Gene Co- Map with the Boundaries of the Chromatin Domain. *EMBO Journal 7*, 655-664.

Piechaczek, C., Fetzer, C., Baiker, A., Bode, J., and Lipps, H.J. (1999). A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. *Nucleic Acids Research 27*, 426-428.

Pikaart, M.I., Recillas-Targa, F., and Felsenfeld, G. (1998). Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes & Development 12*, 2852-2862.

Prosch, S., Stein, J., Staak, K., Liebenthal, C., Volk, H.D., and Kruger, D.H. (1996). Inactivation of the very strong HCMV immediate early promoter by DNA CpG methylation in vitro. Biological Chemistry Hoppe-Seyler *377*, 195-201.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biology 4*, 1707-1719.

Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. *Science 304*, 1811-1814.

Raser, J.M., and O'Shea, E.K. (2005). Noise in gene expression: Origins, consequences, and control. *Science 309*, 2010-2013.

Robinett, C.C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., and Belmont, A.S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *Journal of Cell Biology 135*, 1685-1700.

Romig, H., Fackelmayer, F.O., Renz, A., Ramsperger, U., and 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 Journal 11*, 3431-3440.

Saha, S., Haggardljungquist, E., and Nordstrom, K. (1987). The Cox Protein of Bacteriophage-P2 Inhibits the Formation of the Repressor Protein and Autoregulates the Early Operon. *EMBO Journal 6*, 3191-3199.

Schaarschmidt, D., Baltin, J., Stehle, I.M., Lipps, H.J., and Knippers, R. (2004). An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO Journal 23*, 191-201.

Schubeler, D., Francastel, C., Cimbora, D.M., Reik, A., Martin, D.I.K., and Groudine, M. (2000). Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human β-globin locus. *Genes & Development 14*, 940-950.

Schubeler, D., Groudine, M., and Bender, M.A. (2001). The murine β-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proceedings of the National Academy of Sciences of the United States of America 98*, 11432-11437.

Schwaiger, M., and Schubeler, D. (2006). A question of timing: emerging links between transcription and replication. *Current Opinion in Genetics & Development 16*, 177-183.

Sclimenti, C.R. and Calos, M.P (1998). Epstein-Barr virus vectors for gene expression and transfer. *Current Opinion in Biotechnology 9*, 476-479.

Sessa, G., and Ruberti, I. (1990). Assembly of correctly spaced chromatin in a nuclear extract from Xenopus-Laevis oocytes. *Nucleic Acids Research 18*, 5449-5455.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science 311*, 844-847.

Simons, R.W., Hoopes, B.C., McClure, W.R., and Kleckner, N. (1983). 3 Promoters near the Termini of Is10 - Pin, Pout, and Piii. *Cell 34*, 673-682.

Stehle, I.M., Scinteie, M.F., Baiker, A., Jenke, A.C.W., and Lipps, H.J. (2003). Exploiting a minimal system to study the epigenetic control of DNA replication: the interplay between transcription and replication. *Chromosome Research 11*, 413-421.

Stoll, S.M., Sclimenti, C.R., Baba, E.J., Meuse, L., Kay, M.A., and Calos, M.P. (2001). Epstein-Barr virus/human vector provides high level, long-term expression of α_1-antitrypsin in mice. *Molecular Therapy 4*, 122-129

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature 403*, 41-45.

Tseng, W.C., Haselton, F.R., and Giorgio, T.D. (1997). Transfection by cationic liposomes using simultaneous single cell measurements of plasmid delivery and transgene expression. *Journal of Biological Chemistry 272*, 25641-25647.

Tsukamoto, T., Hashiguchi, N., Janicki, S.M., Tumbar, T., Belmont A.S., and Spector, D.L. (2000). Visualization of gene activity in living cells. *Nature Cell Biology 2*, 871-878.

Tumbar, T., and Belmont, A.S. (2001). Interphase movements of a DNA chromosome region modulated by VP16 transcriptional activator. *Nature Cell Biology 3*, 134-139.

Tumbar, T., Sudlow, G., Belmont, A.S. (1999). Large-scale chromatin unfolding and remodelling induced by VP16 acidic activation domain. *Journal of Cell Biology 145*, 1341-1354.

Udvardy, A. (1999). Dividing the empire: boundary chromatin elements delimit the territory of enhancers. *EMBO Journal 18*, 1-8.

van Rijn, P.A., Goosen, N., Turk, S., and Vandeputte, P. (1989). Regulation of Phage Mu Repressor Transcription by Ihf Depends on the Level of the Early Transcription. *Nucleic Acids Research 17*, 10203-10212.

Vazquez, J., Belmont, A.S., and Sedat, J.W. (2001). Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Current Biology 11*, 1227-1239.

Verschure, P.J., van der Kraan, I., Manders, E.M.M., and van Driel, R. (1999). Spatial relationship between transcription sites and chromosome territories. *Journal of Cell Biology 147*, 13-24.

Volpi, E.V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ragoussis, J., Trowsdale, J., and Sheer, D. (2000). Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *Journal of Cell Science 113*, 1565-1576.

Wade-Martins, R., White, R.E., Kimura, H., Cook, P.R., and James, M.R. (2000). Stable correction of a genetic deficiency in human cells by an episome carrying a 115 kb genomic transgene. *Nature Biotechnology 18*, 1311-1314.

Walther, W., and Stein, U. (2000). Viral vectors for gene transfer - A review of their use in the treatment of human diseases. *Drugs 60*, 249-271.

Wang, Y.M., Tegenfeldt, J.O., Reisner, W., Riehn, R., Guan, X.J., Guo, L., Golding, I., Cox, E.C., Sturm, J., and Austin, R.H. (2005). Single-molecule studies of repressor-DNA interactions show long-range interactions. *Proceedings of the National Academy of Sciences of the United States of America 102*, 9796-9801.

Xu, J., and Koudelka, G.B. (2000). Mutually exclusive utilization of P-R and P-RM promoters in bacteriophage 434 O-R. *Journal of Bacteriology 182*, 3165-3174.

Yasui, D., Miyano, M., Cai, S.T., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature 419*, 641-645.

Yates, J.L., and Guan, N. (1991). Epstein-Barr Virus-Derived Plasmids Replicate Only Once Per Cell-Cycle and Are Not Amplified after Entry into Cells. *Journal of Virology 65*, 483-488.