Development of a Tunable Mammalian Protein Expression System and an Investigation of Promoter Interference in Three Promoters Often Utilized in the Production of Biopharmaceuticals

By Nathan R West (MBiomedSci, MSc)

Thesis submitted for the degree of Doctor of Philosophy (PhD)

Department of Chemical and Biological Engineering

University of Sheffield

Final submission: August 2014

Thesis Abstract

Cell line engineering strategies for improved biopharmaceutical production in mammalian cells often involve the expression of one/multiple genes to try and improve the cellular processes involved in recombinant protein production. Most strategies are relatively simple, involving the use of a strong constitutive promoter for expression of one or more proteins to help increase production. Results often vary and can be cell line and product specific and mean a generic strategy is unlikely to be found. There is a need for more sophisticated expression systems which can express multiple genes but in a controlled fashion and tuned to meet the needs of a specific product. This thesis can be split into two distinct parts but both concern the expression of multiple genes in mammalian cells and recombinant protein production.

A tunable mammalian expression system for multi-gene engineering composed of elements of the mammalian unfolded protein response has been developed. ATF6 (activating transcription factor 6) and its binding element (ERSE – ER stress response element) were used to control the expression of the reporter proteins SEAP (secreted alkaline phosphatase) and GFP (green fluorescent protein). By expressing different amounts of ATF6 and by inserting different numbers of ERSEs upstream of a SV40 (Simian virus 40) promoter, driving SEAP/GFP gene transcription, the level of reporter protein expression could be manipulated in a controlled fashion. The system was capable of controlled/tunable expression of both reporter proteins when expressed alone and when they were co-expressed. This a novel use for ATF6 and ERSE and the first step towards the development of a tunable mammalian expression system for multi-gene engineering. This system could also be easily modified to include or use different transcription factors and binding sites as well as having the potential to use completely synthetic components. This work also showed that the presence of 'promoter interference' (the negative influence of one promoter on another) could be used to our advantage to increase the range of expression.

The SV40 early, human CMV (cytomegalovirus) major immediate-early and human EF1α (elongation factor 1 alpha) are constitutive promoters frequently used in recombinant protein production. The former being used mainly for expression of selection genes and the latter two for strong expression of recombinant proteins. The differences in the strengths of the promoters was demonstrated in CHO (Chinese hamster ovary) cells (CMV > EF1α > SV40) and also their abilities to negatively affect the expression from a co-expressed promoter. The negative influence of one promoter on another is termed 'promoter interference'. The CMV promoter was shown to have the greatest negative effect on expression from another promoter, decreasing both SEAP mRNA and protein expression, while the SV40 had the least. SEAP expression from the SV40 was reduced the most by the presence of a competing promoter. The level of interference inflicted by a competing promoter (CMV > EF1α >

SV40) seemed to be relative to its strength. This is the first time these three important promoters have been compared in a way which not only demonstrates their relative strengths but also their ability to interfere with another promoter when present in the same transient expression system. This also has implications for their use in multi-gene engineering strategies if there is a need for controlled/tunable expression of multiple genes. The work with ATF6 and ERSE showed how 'promoter interference' could be used to our advantage and not necessarily be just a negative occurrence.

One hypothesis for why promoter interference occurs is there is competition between promoters for shared transcription factors (TFs). The promoters were analysed for potential transcription factor binding sites (TFBSs) using the programs MatInspector and ModelInspector. The analysis showed that the SV40 promoter had the least number and variety of potential TFBSs. Both the CMV and EF1α had greater numbers and variety of potential TFBSs. All three promoters had common potential TFBSs but the SV40 promoter shared a greater proportion of its sites with the other two promoters. The number of potential TFBSs and the proportion which were shared reflected both the strength and the ability of a promoter to interfere with another. All three contained TFBSs for the SP1 (specificity protein 1) family of TFs and over-expression of SP1 counter-acted the effects of promoter interference showing that it can affect the expression of all three promoters. However, promoter interference will involve more than just a single TF and also more than just competition for transcriptional activators. This is the first time these three promoters have been compared in terms of the potential TFBSs they contain. The TFBS analysis highlighted the complexity in the control of these promoters and with the effects of promoter interference means that they will be ill suited for the controlled expression of multiple genes without modification.

The work in this thesis was directed towards the controlled expression of multiple genes in mammalian cells for recombinant protein production. We have presented one novel way of controlling the expression of one/two genes with the rest of the thesis looking at the phenomenon of 'promoter interference' between three commonly used promoters. This thesis tries to highlight the importance of multi-gene expression systems as well showing that these three promoters may not be suitable without further modification and also the importance of considering promoter interactions when more than one is present in the same system. The switch to completely synthetic multi-gene control systems is something we envisage happening in the future as the complexities and capabilities of these systems grow.

Acknowledgements

I would first like to thank my supervisor, Prof. David James for giving me the chance to study for a PhD in his lab. I also want to thank him for all his support and advice throughout the duration of my PhD. I am also grateful to the BBSRC for funding me.

A special mention has to go to my lab group including all past and present members who made the experience of a PhD so much greater. I wish them all the best of luck with whatever they decide to do in the future and let them know I really treasured their friendship.

I would also like to say thank you to all the other PhD students, post-docs, supervisors, technicians and other members of staff in the department who made it enjoyable place to study.

A special thanks goes to my friends who don't work at the department (Rob & Shaun) or live in Sheffield (Gav & Bown). Sorry for all the special occasions that I may have missed due to being stuck in the lab and thank you for understanding and not being angered by my absence.

Another thank you goes to the people I started, firstly my Masters and then my PhD with who have become great friends. Leon, thanks for all your help and advice even if I did ask you some stupid questions. Stephen, the climbing sessions were always fun. Joseph, because you put me in your acknowledgements. Steve, for the sessions in the pub and laughs that really helped me unwind and ease the stress.

Lastly, I would like to especially thank my parents, Alan and Denise, my grandparents, Joan and Bill and my girlfriend Danying who without your love and support I would never have made it. It was worth it all in the end.

Contents

Abbreviation List

AP1 – Activator protein 1

ATF – Activating transcription factor

ATF6 – Activating transcription factor 6

BiP – Binding immunoglobulin protein

CD – Chemically defined

CDS – Coding DNA sequence

CHO – Chinese hamster ovary

CHOP – C/EBP homologous protein

CMV – Cytomegalovirus

CREB – cAMP response-element binding protein

CypB – Cyclophilin B

DHFR – Dihydrofolate reductase

DOE – Design of experiment

DTE – Difficult to express

EF1α – Elongation factor 1 alpha

EGR – Early growth response

EPO – Erythropoietin

ER – Endoplasmic reticulum

ERAD – ER associated degradation

ERSE – Endoplasmic reticulum stress response element

ETS – E-twenty-six

FACS – Fluorescent-activated cell sorting

GFP – Green fluorescent protein

GS – Glutamate synthetase

IGF – Insulin like growth factor

IRE1 – Inositol-like kinase 1

KLF – Kruppel-like factor

MAb – Monoclonal antibody

MAR – Matrix attachment region

MSX – Methionine sulphoximine

MTX – Methotrexate

NFAT – Nuclear factor of activated T-cells

NF-κB – Nuclear factor-kappa-light-chain-enhancer of activated B cell

OCT1 – Octamer binding protein 1

PDI – Protein disulphide-isomerase

PERK – Protein kinase-like ER kinase

RNAP II – RNA polymerase II

SEAP – Secreted alkaline phosphatase

S/MAR – Scaffold and matrix attachment region

SNARE – Soluble N-ethylmaleimide-sensitive factor attachment receptor

SP1 – Specificity protein 1

SV40 – Simian virus 40

TF – Transcription factor

TFBS – Transcription factor binding site

TOP – Terminal oligopyrimidine tract

tPA – Tissue plasminogen activator

UCOE – Ubiquitous chromatin opening element

UPR – Unfolded protein response

UTR – Untranslated region

XBP1 – X-box binding protein

ZF – Zinc finger

Chapter 1 – Introduction and Research Aims/Objectives

1.1 – Introduction

Global sales of biopharmaceuticals were a reported $99 billion in 2009 and is an industry which continues to grow (Walsh, 2010). It has come a long way since the production of the first therapeutic protein produced in Chinese hamster ovary (CHO) cells in the mid 1980s (Kaufman *et al.*, 1985). The production yields of recombinant proteins have increased >20 fold over the last ~30 years (De Jesus & Wurm, 2011). CHO cells are the main cell line used for the production of recombinant therapeutic proteins and especially monoclonal antibody (MAb) products (O'Callaghan & James, 2008). It has now become routine to be able to produce therapeutic recombinant proteins in g L^{-1} amounts in large scale bioprocesses (De Jesus & Wurm, 2011). This is thanks to improvements in all the major processes which are involved in recombinant protein production including improved stable cell line creation/selection (Fan *et al.*, 2013; Lai *et al.*, 2013; Fan *et al.*, 2012), improved bioprocesses (like optimized media and feeding strategies) (Wurm, 2004), to genetic manipulation of cell lines for improved growth, survival and recombinant protein production (Li *et al.*, 2010).

With the recent publication of the CHO genome we have entered a new era for the production of biopharmaceuticals (Brinkrolf *et al.*, 2013; Lewis *et al.*, 2013; Li *et al.*, 2010). The publication of the CHO genome has lead to the development and use of omics technologies for the analysis of CHO cell lines producing recombinant proteins which were much tougher to carry out without it. There is also an increasing number of novel therapeutics which are being developed. However, some of these products and more conventional MAb products suffer from unexpectedly low titres which hinder their prospects of becoming a clinically important biopharmaceutical. In the future there is going to be a need for sophisticated rational multi-gene engineering strategies which are optimized and based on the requirements of a specific product rather than a generic approach. For this to occur, new multi-gene expression systems are needed which can not only express multiple genes simultaneously (Kriz *et al.*,

2010) but which can also discretely control (tune) the expression levels of multiple genes in CHO cells.

Although details of some systems for mammalian cells have been published (Endo *et al.*, 2013; Stapleton *et al.*, 2012) there are currently no published papers describing a system which can be used for the controlled/tunable expression of multiple genes in CHO cells.

1.2 – Research Aims and Objectives

1.2.1 – Aims

To develop a mammalian tunable protein expression system were the expression of one or multiple genes can be controlled and expressed at a desired level.

To investigate the effects of 'promoter interference' between three promoters, the SV40 (Simian virus 40) early promoter, the human CMV (cytomegalovirus) major immediate-early promoter and the human EF1α (elongation factor 1 alpha) promoter, all commonly utilized in biopharmaceutical production with a view to assessing their suitability in multi-gene engineering strategies.

1.2.2 – Objectives

1. Construct a system for the controlled/tunable expression of one/two proteins

2. Compare the strengths and promoter interference effects of three promoters (SV40, CMV and EF1α)

3. Analyse the SV40, CMV and EF1α promoters for their potential transcription factor binding sites. For differences and similarities between the promoters which could potentially impact both their expression and their ability to interfere with another promoter

Chapter 2 – Literature Review

2.1 – The Development of Biopharmaceutical Production in Mammalian Cells

2.1.1 – Introduction

CHO cells have been the predominant mammalian cell line used for the research and production of biopharmaceuticals over the last 25-30 years (O'Callaghan & James, 2008; Birch & Racher, 2006; Walsh, 2006). They are used due to their ability to produce fully glycosylated recombinant proteins as well as having a long and favourable track record with drug regulators (O'Callaghan & James, 2008; Dinnis & James, 2005). Other mammalian cell lines such as murine lymphoid (NS0 and SP2/0), human embryonic kidney (HEK)-293 and BHK21 (baby hamster kidney) cells have also been used (O'Callaghan & James, 2008; Birch & Racher, 2006). CHO cells have been used to produce a variety of recombinant therapeutic proteins including blood factors (e.g. – coagulation factor VIII), thrombolytics (e.g. – tissue plasminogen activator (tPA)), growth factors (erythropoietin (EPO)) and hormones (e.g. – thyrotrophin-α) (Walsh, 2010). However, recombinant monoclonal antibodies (MAb) are probably the biggest and most important class of therapeutic proteins and their sales in 2009 accounted for over a third ($38 billion) of the $99 billion global biopharmaceutical market (Walsh, 2010).

CHO cells originated from an immortalised cell taken from a primary culture of ovarian cells from a Chinese hamster (*Cricetulus griseus*) in 1957 (Hacker *et al.*, 2009). Although not initially used for the production of biopharmaceuticals, it was in the 1980s when they were first used in the production of recombinant therapeutic proteins. The first recombinant protein to be FDA-approved was tPA in 1986/7 (Kim *et al.*, 2012; De Jesus & Wurm, 2011; Kaufman *et al.*, 1985). A typical production run for a batch culture lasted around a week in the 1980s with low maximum cell densities (1-2 x 10^6 cells mL^{-1}) and low specific (10-20 pg $cell^{-1}$ day^{-1}) and volumetric (50-100 mg L^{-1}) productivities (Wurm, 2004). In the last 25 years through improvements in the bioprocessing of recombinant proteins we now have fed-batch cultures whose production runs can last three weeks, achieve maximum cell densities of 15 x 10^6 cells

mL^{-1} and titres of 5 g L^{-1} (De Jesus & Wurm, 2011). Some biopharmaceutical companies have claimed, mainly through conference presentations, to have bioprocesses which have achieved over 10 g L^{-1} for the production of MAb's, but high titres such as these are still limited by downstream purification processes (Kelley, 2009). These increases have been made through a combination of different methods but have mainly been through optimization of cell media and improvements in the bioprocesses of large scale production so that cells remain healthier at higher cell densities for longer in culture and thus produce more recombinant protein.

2.1.2 – Media optimization and the development of large scale bioprocesses

Commercial chemically defined (CD) media, which is usually of excellent consistent quality, is readily available for academic research and is suitable for some production processes. However, many biopharmaceutical companies use custom made media which are usually optimised for the production of a specific product (Wurm, 2004). Often different media formulations will be used for different stages of a production process (Wurm, 2004). The exact composition of such media and even the commercial media remains undisclosed by the companies producing it and the biopharmaceutical companies using it to protect their intellectual property. One known change which has occurred as the biopharmaceutical industry grew is the removal of serum and other animal derived components and the switch to CD media. Serum was used to provide the factors which were important for growth such as carbohydrates, lipids, growth factors and other peptides but was removed due its heterogeneous character, its high cost and the risk of virus transmission (Grillberger *et al.*, 2009; Wurm, 2004). In fact serum and other animal derived products are avoided in commercial production processes with pressure from drug regulators being another major reason for their removal.

Additives which improve media performance in the absence of serum and have been reported in the literature include insulin and insulin-like growth factors (IGFs) which have been shown to have anti-apoptotic effects and help maintain high viable cell densities (Morris & Schmid, 2000; Sunstrom *et al.*, 2000; Baserga *et al.*, 1997). The use

of non-animal derived hydrolysates which contain mixtures of lipids, oligopeptides, amino acids, iron salts, vitamins and other small molecules in trace amounts have also been shown to increase both cell specific growth rate and cell specific productivity but not necessarily both at the same time (Kim & Lee, 2009; Sung *et al.*, 2004). However, the complex composition of hydrolysates means there is lot-to-lot variation and there can be significant variability in different batches of media if they are used. CD media containing specific amounts of vitamins, inorganic salts, trace elements, lipids, amino acids, insulin or IGFs and proprietary components have been developed over many years through methods for optimizing media composition including medium blending, single component titration, spent medium analysis and metabolite profiling (Li *et al.*, 2010). As mentioned earlier media may be customised for a specific product but this can be an extremely time-consuming task but this can be shortened through the use of a design of experiment (DOE) approach (Castro *et al.*, 1992).

The actual culture process itself has not changed too dramatically. The production of recombinant therapeutic proteins from mammalian cells is still carried out mainly in large-scale stirred-tank bioreactors and the culture can still be split into three phases – seeding, inoculation and production. However, the technology for monitoring, analysis and control of culture parameters and the feeding strategies for large scale cultures have been improved dramatically. An optimized feed medium and feeding strategy will likely be developed while developing the media for a process. This is due to the feed composition sometimes being similar to the bulk media but also because similar considerations are taken in to account. These include its effects on cell growth and productivity, lactate and ammonia accumulation as well as nutrient consumption including levels of glucose, glutamine and other amino acids in the media which are all inherently linked (Khattak *et al.*, 2010; Li *et al.*, 2010; Zhou *et al.*, 1997; Zhou *et al.*, 1995). Limiting lactate and ammonia accumulation is essential due to their known deleterious effects on cell growth, productivity and product quality (Khattak *et al.*, 2010; Altamirano *et al.*, 2006; Chee Furng Wong *et al.*, 2005; Altamirano *et al.*, 2001).

The monitoring, analysis and control of physical (agitation speed, gas flow rates, temperature etc), chemical (dissolved oxygen and CO_2, pH, osmolality etc) and biological (viable cell concentration, viability, NADH and LDH levels etc) parameters

have also improved bioprocess performance and all can affect the quality of the product produced including important post-translational modifications such as glycosylation (Grainger & James, 2013; Gawlitzek *et al.*, 2009; Walsh & Jefferis, 2006). Monitoring and analysis of process parameters can occur both online and offline, with online analysis allowing continuous monitoring and the ability to adjust parameters quickly and efficiently, while offline analysis is used to measure parameters which are difficult or too costly to measure online as well validating online measurements. The growth in our capabilities for monitoring bioprocess parameters has allowed us to gain a greater understanding of the parameters which most affect culture performance, such as pH (Wayte *et al.*, 1997) and lactate accumulation (Altamirano *et al.*, 2006; Chee Furng Wong *et al.*, 2005; Altamirano *et al.*, 2001), and allow us to make adjustments accordingly. We can already monitor numerous parameters producing huge amounts of data even for a single production run. The challenge in the future is to be able to analyse such large data sets and to extract meaningful results and this will involve developments in data mining techniques and multivariate analysis (Le *et al.*, 2012).

2.2.3 – Creation and selection of high producing stable cell clones

Although the increases in production of recombinant therapeutic proteins has mainly been through improvements in the large scale culture of mammalian cells, processes upstream of this are also vitally important and development of these processes has led to some good increases in productivity but have mainly seen dramatic decreases in the time and labour required and have lowered the cost of manufacturing. The major upstream process is that of creating a stable cell line which expresses the desired therapeutic protein at a sufficiently high enough expression level to make it viable for use in the large scale production of the product. Like many of the processes involved in recombinant protein production the creation of stable cell lines is time and labour intensive. Therefore methods which help create and select high expressing cell clones in a more efficient and time effective way have been developed to help reduce the time and labour required. This has been achieved through our greater understanding

of molecular biology and both improvements in expression systems and clone selection technology.

The first part of creating a high expressing stable cell line is the transfection and then incorporation of an expression vector, containing the genes for both the recombinant protein and a selection marker, into one of the cells chromosomes. This usually occurs through homologous recombination but the expression of both the selection gene and recombinant protein can be negatively affected by its random integration site within a chromosome (Kim *et al.*, 2004; Zahn-Zabal *et al.*, 2001). Different DNA elements can be inserted into expression vectors to negate the positional effects of the site of chromosomal integration. Scaffold and matrix attachment regions (S/MAR) can be inserted into expression vectors and act to insulate the recombinant genes after integration from negative positional effects by maintaining a transcriptionally active chromatin structure (Girod *et al.*, 2007; Mirkovitch *et al.*, 1984). The β-globin MAR, chicken lysozyme 5' MAR and the β-interferon SAR have all been successfully used to increase the numbers of high producing cell clones (Girod *et al.*, 2005; Kim *et al.*, 2005; Kim *et al.*, 2004). Another element that insulates recombinant genes from negative positional effects is the ubiquitous chromatin opening element (UCOE) which has been shown to maintain chromatin in an open transcriptionally active state and increase both the numbers of high producing clones and recombinant protein production (Ye *et al.*, 2010; de Poorter *et al.*, 2007; Benton *et al.*, 2002).

Insertion of recombinant genes through homologous recombination results in their random integration into chromosomes but through site-specific recombination they can be incorporated more selectively into highly transcriptionally active areas of chromosomes (Huang *et al.*, 2007). This has been accomplished using a number of differing strategies. The Flp/FRT system, from *Saccharomyces cerevisiae* and marketed as Flp-In™ by Life Technologies (Paisley, UK), and the Cre/LoxP system, from bacteriophage P1, have both been used to create stable cell lines expressing recombinant proteins in mammalian cells (Kameyama *et al.*, 2010; Huang *et al.*, 2007; Wiberg *et al.*, 2006; Kito *et al.*, 2002). Other systems, which use the λ and ΦC31 integrases, which target attP and attB sites respectively, have also been used for site-specific recombination of recombinant proteins (Campbell *et al.*, 2010; Russell *et al.*,

2006; Smith & Thorpe, 2002). These integrase systems have the advantage of irreversible integration of your recombinant protein when compared to the Flp/FRT and Cre/LoxP systems.

The main expression systems used in CHO cells are Lonza's GS (glutamate synthetase) expression system™ (Bebbington *et al.*, 1992) (Lonza Biologics, Slough, UK) and the DHFR (dihydrofolate reductase) expression system (Kaufman & Sharp, 1982). Both require cell lines were there is deficient or the absence of expression of a particular selection gene, GS for the Lonza system and *DHFR*$^{-}$ cells (Urlaub & Chasin, 1980) for the corresponding system, and also the use of a selection drug, MSX (methionine sulphoximine) which inhibits GS and MTX (methotrexate) which inhibits DHFR. The DNA expression vector containing the gene for the recombinant protein will also contain the gene for either GS or DHFR and allow cells which incorporate the vector to survive in the presence of the appropriate selection drug. Through multiple rounds of cell selection using increasing concentrations of a selection drug, gene amplification can occur (Schimke, 1984) making some cells even more resistant to the drug while also increasing the expression of the therapeutic protein in some of these cells. Recently researchers at Eli Lily (Indianapolis, US) improved the GS system by both creating full GS knockouts of the Lonza CHOK1SV cell line and weakening the promoter (SV40) used to express the GS gene thus increasing the selection stringency for high producers when used to make stable cell lines (Fan *et al.*, 2013; Fan *et al.*, 2012). A new version of the GS system, called GS Xceed™ (http://www.lonza.com/custom-manufacturing/development-technologies/gs-xceed-gene-expression-system.aspx) was released in 2012 by Lonza which uses the GS knockout version of the CHOK1SV cell line. Improvements for the DFHR system have included both codon de-optimization of the selection gene and insertion of protein destabilizing elements to weaken the selection protein (Westwood *et al.*, 2010; Ng *et al.*, 2007). Another method for improving a neomycin expression system included mutating the selection protein to have less affinity for binding the selection drug (Sautter & Enenkel, 2005). All these methods lead to increasing stringency in selection of cells producing high levels of recombinant proteins.

Even though improvements in the insertion of expression vectors and expression systems have improved the proportion of high producing cells in a pool stably expressing a recombinant protein, they are still relatively rare. Isolation of high producing clones was traditionally done by limited dilution cloning and, although simple, this is a long and time consuming process with many limitations. More advanced high-throughput methods can be used to find higher numbers of high producing clones and include the use of FACS (fluorescent-activated cell sorting) and automated methods such as the ClonePix system (Genetix, Sunnyvale, US) and CellExpress™ + LEAP™ (laser-enabled analysis and processing) (Richardson *et al.*, 2010).

FACS is a high-throughput technology that is capable of sorting thousands of cells in a short period of time analysing multiple fluorescent wavelengths. It sorts cells by measuring the fluorescence signal associated with individual cells and since most recombinant proteins are secreted the major challenge for this technique is to create a fluorescent signal that remains associated with the cell. Several techniques have been developed for the use of FACS to select cell clones producing high levels of secreted recombinant protein. Some involve the immobilization of the secreted recombinant protein so that it is still associated with the cell and then the addition of fluorescent antibodies which has been shown to improve clone selection, but preparation of cell samples in this fashion can be time consuming and technically very challenging (Atochina *et al.*, 2004; Manz *et al.*, 1995). Another method used the hypothesis that the amount of secreted recombinant protein associated with the cell surface correlated with the production capacity of the cell and this was tested by cooling cells to between 0-4 °C and applying fluorescent antibodies and then sorting by FACS (Brezinsky *et al.*, 2003). Other methods involve fluorescently tagging cells intracellularly. This has been done through the co-expression of a fluorescent protein as a separate secondary selection protein and fluorescently tagging the selection protein metallotheonine (Bailey *et al.*, 2002; Meng *et al.*, 2000). Fluorescent labelling of MTX has also been used to tag cells using the DFHR expression system and have an advantage in that it did not require the expression of a secondary selection protein (Yoshikawa *et al.*, 2001).

The ClonePix FL system uses semi-solid media containing fluorescein isothiocyanante (FITC) conjugated antibodies, which can isolate single colonies of cells, capture their secreted recombinant protein, produce a halo of fluorescence surrounding the colony which can then analysed and high producing colonies selected and removed for further characterization (Dharshanan *et al.*, 2011). The process is automated and is an extremely efficient system capable of screening 10,000 cells an hour while being sensitive enough to isolate high producing clones who make up only 0.003% of an entire population and has already been successfully used in the isolation of high producing clones for a variety of recombinant proteins (Lai *et al.*, 2013; Dharshanan *et al.*, 2011; Serpieri *et al.*, 2010).

The Cell Xpress + LEAP platform is a fully automated self contained system which uses microplates which can capture secreted recombinant proteins (Richardson *et al.*, 2010; Browne & Al-Rubeai, 2007; Hanania *et al.*, 2005). It detects fluorescence associated with either cells or their secreted recombinant protein to detect the highest producing clones within a well. Other cells are then destroyed using a laser to leave only high producing clones within a well which can then be allowed to grow further before removal and expansion in larger wells (Browne & Al-Rubeai, 2007; Koller *et al.*, 2004). This system has been shown to be capable of routinely selecting clones with specific productivities of > 50 pg $cell^{-1}$ day^{-1} (Hanania *et al.*, 2005). One drawback of the system is the laser may also damage high producing clones (Browne & Al-Rubeai, 2007).

2.2.4 – Genetic engineering of mammalian cell lines

Various genetic engineering strategies have also been used in attempts to improve recombinant protein production in mammalian cells. These have included the expression of recombinant genes and the suppression of endogenous genes for the extension of culture duration, maximising viable cell density, increasing growth rates and the expression of genes to increase the protein folding and vesicle trafficking/secretion capacities of mammalian cells.

Through the expression/suppression of genes involved in apoptosis (type I programmed cell death) mammalian cells producing recombinant proteins have been

engineered to survive longer in culture and in some studies produce increased amounts of recombinant protein. The process of apoptosis is controlled by positive and negative regulators and both have been targeted for cell line engineering. The anti-apoptotic protein Bcl-2 (B-cell lymphoma 2) has been over-expressed and shown to promote the survival and increase protein production in recombinant protein producing cells both in the presence and absence of the apoptosis inducing agents such as sodium butyrate (NaBu) (Lee & Lee, 2003; Kim & Lee, 2001; Fussenegger *et al.*, 2000; Tey *et al.*, 2000). More recently a CHO cell line expressing both Bcl-2 and the pro-autophagy protein Beclin-1 has been developed which extended both culture duration and maintained a higher viability then a cell line only expressing Bcl-2 (Lee *et al.*, 2013). Bcl-xL (B-cell lymphoma-extra large) and Mlc-1 (Megalencephalic leukoencephalopathy with subcortical cysts 1) are also anti-apoptotic proteins which have been over-expressed and were also shown to increase product titres and increase viability in CHO cells (Majors *et al.*, 2008b; Chiang & Sisk, 2005; Figueroa *et al.*, 2003). Bcl-x(L) has also been shown to increase production and viability in a transient expression system (Majors *et al.*, 2008b). Bak (Bcl-2 homologous antagonist/killer) and Bax (Bcl2-associated X protein) are pro-apoptotic factors and have been deleted in a CHO cell line resulting in increased cell densities and product titres (Cost *et al.*, 2010). The apoptotic signalling pathway is mediated by capases, such as capase-3, -7, -8 and -9, interference with either their expression or function has been shown to have beneficial effects on cell growth and in some cases productivity (Sung *et al.*, 2007; Yun *et al.*, 2007; Kim & Lee, 2002). In another study four CHO cell lines expressing IFN-γ (Interferon gamma) were engineered for either the over-expression of FAIM (Fas apoptosis inhibitory molecule) or FADD (Fas-associated deathdomain), or the knock-down of ALG-2 (Apoptosis-linked gene 2) or REQUIEM, all of which are involved in apoptosis (Wong *et al.*, 2006b). All four cell lines showed an increase in viability and reductions in apoptosis while ALG-2 and REQUIEM increased the maximum viable cell density. All four were also shown to increase the final product titre of IFN-γ. XIAP (X-linked modifier of apoptosis) is an inhibitor of capases (-3, -7 and -9) and its over-expression has been shown to significantly decrease apoptosis in both CHO and HEK-293 cells (Sauerwald *et al.*, 2003; Sauerwald *et al.*, 2002). A cell line combining the expression of XIAP and unfolded protein response (UPR) protein XBP1 (X-box binding

protein 1) (which can decrease growth and induce apoptosis) was shown to increase recombinant protein production and improve maximum cell density and viability compared to a cell line only expressing XBP1 (Becker *et al.*, 2010).

As well as inhibiting apoptosis, the inhibition of autophagy (type II programmed cell death) is another cellular target for engineering of mammalian cells producing recombinant proteins and has been observed in the late stages of culture when nutrients become exhausted (Han *et al.*, 2011; Hwang & Lee, 2008). Over-expression of Akt (protein kinase B), which plays a role in cell survival, was shown to delay the onset of both autophagy and apoptosis in CHO cells in batch culture (Hwang & Lee, 2009). Bcl-2 expression has also been shown to interact with the autophagy mediator Beclin-1 and as mentioned previously both proteins have been over-expressed in the same cell line for extension of culture duration and higher viabilities (Lee *et al.*, 2013; Kim *et al.*, 2009). A chemical inhibitor of autophagy (3-methyl adenine) has also been shown to significantly improve the yield of tPA in a CHO fed-batch culture (Jardon *et al.*, 2012).

Mammalian cells producing recombinant proteins have also been engineered for an increase in cell specific growth rate and/or maximum viable cell density. The cell cycle transcription factor E2F-1 has been over-expressed in CHO cells and led to both an extra day of proliferation and increased viable cell density when compared to a control cell line (Majors *et al.*, 2008a) while over-expression of c-Myc (v-myc avian myelocytomatosis viral oncogene homolog), which is involved in cell cycle progression, increased cell specific growth rate and maximum viable cell density (Kuystermans & Al-Rubeai, 2009). Over-expression of two proteins which were shown to be up-regulated in fast growing cells, Cdkl3 (Cyclin-Dependent Kinase-Like 3) and Cox15 (Cytochrome c oxidase assembly protein COX15 homolog), led to increases in proliferation rates and maximum viable cell densities respectively in mammalian cells (Jaluria *et al.*, 2007). However, significant increases in the production of recombinant proteins were not seen when these proteins were over-expressed. When mTOR (mammalian target of rapamycin), a global metabolic sensor, was over-expressed increases in proliferation, cell size, viability and cell specific productivity in a CHO cell line producing a therapeutic recombinant protein was observed (Dreesen & Fussenegger, 2011). The

multiple effects of mTOR over-expression reflect its role as a global metabolic sensor and the multiple cellular processes which it is involved in. These include ribosomal protein synthesis (translation initiation and elongation) (Ma & Blenis, 2009; Richter & Sonenberg, 2005), cell survival (inhibition of autophagy) (Kim *et al.*, 2011; Jung *et al.*, 2010; Sudarsanam & Johnson, 2010), cell proliferation (promotes proliferation) and cell growth (Fingar & Blenis, 2004; Fingar *et al.*, 2004; Fingar *et al.*, 2002) and positively regulating mitochondrial ATP synthesis capacity (Schieke *et al.*, 2006) among many others (Laplante & Sabatini, 2012).

As well as engineering cells for increased survival/growth/viability, cell engineering strategies have been attempted which try to increase the capacity of mammalian cells to produce and secrete recombinant protein. This has involved the modulation of genes involved in protein synthesis, secretion and the UPR.

Molecular chaperones are proteins involved in the assembly and disassembly, and the folding and unfolding of proteins and are found mainly in the endoplasmic reticulum (ER), but can also be found in the nucleus and in mitochondria. Their roles in protein production make them an obvious genetic engineering target and numerous studies have been carried out for their effects on recombinant protein production in mammalian cells. The ER molecular chaperone PDI (protein disulfide-isomerase), which helps form disulfide bonds in newly assembled proteins (Freedman *et al.*, 1994), has been over-expressed in a number of studies. It was shown to increase the cell specific productivity of MAb producing CHO cell lines (Mohan *et al.*, 2007; Borth *et al.*, 2005), but not to have any effect on cell lines producing the recombinant proteins such as a TNFR:Fc (tumor necrosis factor receptor:Fc) fusion protein and TPO (thrombopoietin) (Mohan *et al.*, 2007; Davis *et al.*, 2000). Another isoform of PDI, ERp57, was shown to increase the production of TPO (Hwang *et al.*, 2003). Simultaneous over-expression of two proteins involved in protein folding and quality control in the ER, calnexin and calreticulin, also resulted in increased TPO production (Chung *et al.*, 2004). The protein BiP (Binding immunoglobulin protein) is also found in the ER and is important in both protein folding and assembly (Dorner & Kaufman, 1994; Haas & Wabl, 1983) and has previously been a engineering target but over-expression was showed to have no significant effect on recombinant protein

production (Borth *et al.*, 2005; Dorner & Kaufman, 1994). However, more recently BiP has been shown to increase MAb production from a stable mammalian cell line alone and in combination with other chaperones such as PDI, ERO1Lβ (Endoplasmic reticulum oxidoreductin-1-like beta) (Nishimiya *et al.*, 2013). In another more recent study the molecular chaperones BiP, PDI and CypB (Cyclophilin B), an ER peptidyl-prolyl-isomerase (Meunier *et al.*, 2002), were over-expressed with a panel of transiently expressed MAb whose final production titre differed >4-fold (Pybus *et al.*, 2013). Both BiP and CypB increased the final titre while BiP and PDI over-expression caused increases in cell specific productivity. However, this only occurred for a small number of the MAbs expressed, which shows along with the other studies mentioned that although the over-expression of molecular chaperones can improve recombinant protein production they are likely to be protein specific, depending on the bottleneck of protein production. For improved expression, investigation of the rate limiting step for production is likely to be required and one study showed that the bottleneck in a high expressing human MAb CHO cell lines was ER-Golgi transport (Hasegawa *et al.*, 2011).

The movement of proteins between intracellular compartments and to the cell surface for extracellular secretion is carried out by a complex vesicle trafficking system and the proteins involved are also cell engineering targets. One family of proteins which are important are the Sec1/Munc18 (secretory blood group 1/mammalian uncoordinated-18 (SM)) proteins and the SM proteins Sly1 (also known as SCFD1 - sec1 family domain containing 1) and Munc18 have been over-expressed in a MAb producing CHO cell line (Peng & Fussenegger, 2009). Over-expression of each protein alone increased production significantly, dual expression increased production further, while expression of both these proteins along with the UPR factor XBP1 increased production further still. SNARE (soluble N-ethylmaleimide-sensitive factor attachment receptor) proteins are another set of proteins important in vesicle trafficking and over-expression of the exocytic SNAREs, SNAP-23 (synaptosomal-associated protein, 23kDa) and VAMP8 (vesicle-associated membrane protein 8), were both shown to increase recombinant protein production in mammalian cells along with a MAb in a CHO cell line (Peng *et al.*, 2011). Interestingly other exocytic SNAREs had no positive effect on

recombinant protein production in the same study (Peng *et al.*, 2011). CERT (ceramide transfer protein) is a protein involved in protein transport between the Golgi and the plasma membrane and its over-expression increased the cell specific productivities and final titre of a therapeutic protein in CHO cells (Florin *et al.*, 2009). In a more recent study a mutant form of CERT, which is resistant to phosphorylation, was over-expressed in a tPA producing CHO cell line and specific productivity was increased (Rahimpour *et al.*, 2013).

The UPR is an intracellular response to a build up of unfolded proteins in the ER and modulated through a number of different proteins which can affect the function and expression of multiple target proteins to help alleviate such build ups (Kohno, 2010; Mori, 2009) (a more detailed description is given in figure 2.1). It has long been thought that the expression of recombinant proteins, especially those that are highly expressed and also proteins which are termed 'difficult-to-express', may induce a UPR and has made it a natural target for cell engineering. Indeed it has been shown in transient expression experiments in mammalian cells that recombinant proteins can induce a UPR, while stable cell lines in the late stages of culture and have shown increases in the UPR induced XBP1 protein (Pybus *et al.*, 2013; Ku *et al.*, 2010). The negative effects of an UPR can be to down regulate global and specific mRNA translation, induce ERAD (ER associated degradation) potentially degrading your recombinant protein and also potentially induce apoptosis and cell death (Kohno, 2010; Mori, 2009). This can result in low specific productivities. However, there are aspects of a UPR that might be beneficial for recombinant protein production and these include inducing the expression of ER proteins which aid the folding and processing of proteins (molecular chaperones – e.g.:- PDI, BiP etc.), as well as expanding the ER (endoplasmic reticulum) and improve secretion (Hetz, 2012). Therefore there has been work carried out to engineer cells to help negate the negative aspects of the UPR and also take advantage of the beneficial parts.

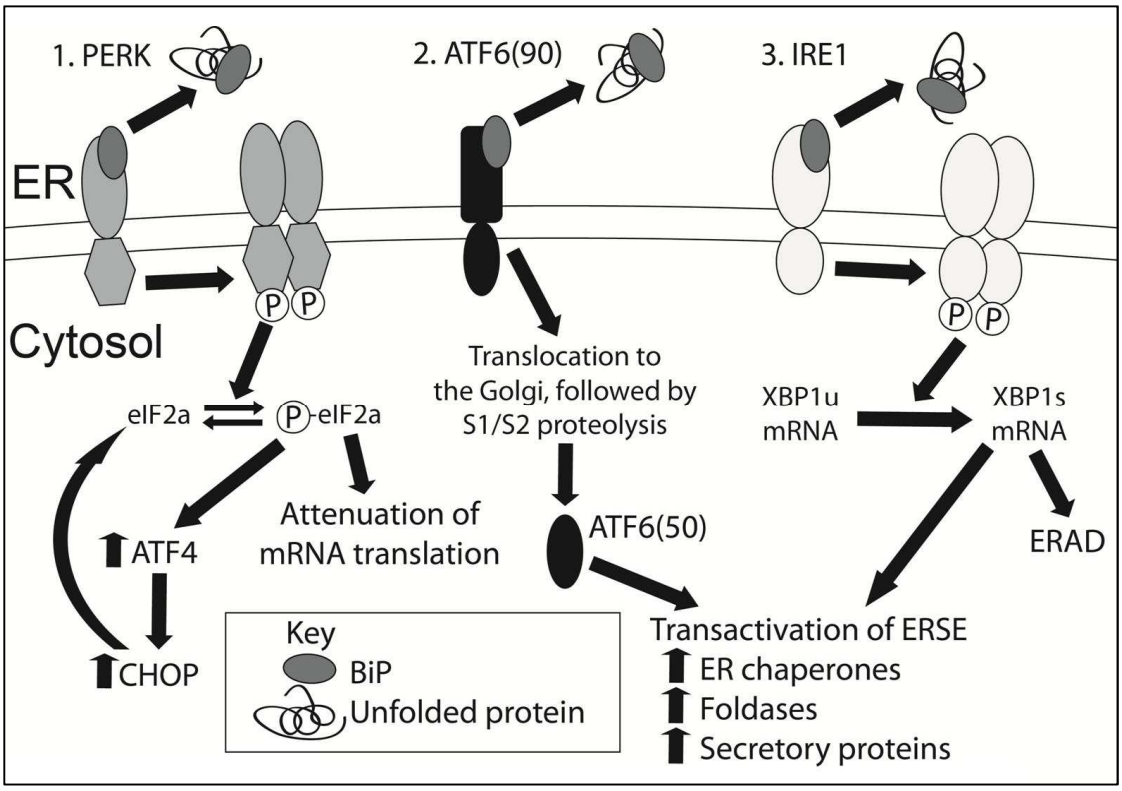

Figure 2.1 – Diagram of the unfolded protein response. The unfolded protein response (UPR) is an intracellular response to ER stress such as an accumulation of mis-folded/unfolded protein. When unfolded protein begins to accumulate in the ER the molecular chaperone BiP disassociates from the three ER transmembrane UPR receptors to bind unfolded protein so it is retained within the ER. **1. PERK** – (protein kinase-like endoplasmic reticulum kinase) disassociation of BiP from PERK leads to PERK dimerization and phosphorylation. Phosphorylated PERK can in turn phosphorylate eIF2a (eukaryotic initiation factor 2 alpha) which attenuates global mRNA translation preventing a further increase of unfolded protein inside the ER. Allowing the ER time to refold the unfolded protein or degrade it. Although translation is attenuated some proteins can be translated. One of these is ATF4 which has pro-survival functions. ATF4 can induce expression of CHOP which can lead to dephosphorylation of eEF1a thus allowing global mRNA translation to resume. This creates a negative feedback loop but this only occurs if the unfolded protein has been processed accordingly. **2. ATF6 –** (activating transcription factor 6) after BiP dissociation the ATF6 protein it translocates to the Golgi were it is cleaved by S1/S2 proteases from a 90 kDa form (ATF6(90)) to a 50 kDa form (ATF6(50)). ATF6(50) is then free to move to the nucleus were it acts as a potent transcriptional activator of UPR target genes which are important in protein folding and secretion. It targets UPR genes by binding ER-stress response elements (ERSE) which are upstream of these genes. **3. IRE1 –** (inositol-requiring kinase 1) like PERK, IRE1 forms a homodimer and is phosphorylated. Phosphorylation of IRE1 leads to the splicing of XBP1µ (X-box binding protein 1) mRNA which can then be translated into an activating transcription factor (XBP1s) which like ATF6(50) can induce the expression of target genes involved in protein folding and secretion. XBP1s is also involved in ERAD (ER associated degradation). If unfolded protein persists in the ER and this accumulation of protein is not removed a UPR can eventually induce apoptosis (Chakrabarti *et al.*, 2011).

XBP1 is the most studied UPR protein for increased recombinant protein production and is a transcription factor which induces the expression of target genes involved in protein entry into the ER, protein folding, ERAD, glycosylation, redox metabolism, lipid biogenesis, vesicular trafficking and autophagy (Acosta-Alvear *et al.*, 2007; Lee *et al.*, 2005; Shaffer *et al.*, 2004; Lee *et al.*, 2003). XBP1 over-expression was first shown to

increase production of a glycosylated reporter protein, as well as human VEGF (vascular endothelial growth factor) in CHO cells (Tigges & Fussenegger, 2006). The transient expression of EPO was also improved by XBP1 over-expression, but no increases in a MAb, IFN-γ or EPO production were seen from stable cell lines for these recombinant proteins (Ku *et al.*, 2010; Ku *et al.*, 2008). In another study, XBP1 over-expression was shown to increase the size of the ER and also the final titre of a MAb in stably expressing CHO cells (Becker *et al.*, 2008). It has also been shown to increase the expression of low expressing 'difficult-to-express' MAb's in a transient study (Pybus *et al.*, 2013). XBP1 was stably expressed along with the oxidoreductase ERO1Lα (Endoplasmic reticulum oxidoreductin-1-like alpha) for creation of a CHO cell line developed specifically for transient gene expression and showed increased MAb expression compared to the CHO-S cell line (Life Technologies, Paisley, UK) (Cain *et al.*, 2013).

The UPR target genes ATF4 (activating transcription factor 4), GADD34 (growth arrest and DNA damage-inducible protein 34) and CHOP (C/EBP homologous protein) have also been targets for cell line engineering. Over-expression of the transcription factor ATF4 and one of its target genes GADD34, which are involved in reversing the attenuation of translation caused by PERK (protein kinase-like endoplasmic reticulum kinase) phosphorylation of eIF2α (eukaryotic initiation factor 2 alpha), were both shown to increase the production of AT-III (Antithrombin III) in CHO cells (Haredy *et al.*, 2011; Ohya *et al.*, 2008). Another target of ATF4, CHOP, which plays a role in apoptosis, was shown to increase the expression of a MAb from a mammalian cell line both alone and also in combination with other UPR related proteins and ER chaperones such as ERdj3, ERdj5, BiP, PDI and ERO1Lβ (Nishimiya *et al.*, 2013). Combinations of CHOP and the other proteins increased MAb titre more than CHOP alone.

Along with XBP1, ATF6 is another important transcription factor which induces the expression of variety of target genes involved in protein folding and ER homeostasis when a UPR is induced (Bobrovnikova-Marjon & Diehl, 2007). Compared with XBP1 there is only a single study which showed ATF6 to successfully increase the production of recombinant proteins in mammalian cells (Pybus *et al.*, 2013).

One reason why proteins such as ATF6, XBP1 and mTOR are interesting targets for cell line engineering strategies is the fact they affect multiple targets. Many cell genetic engineering strategies involve the expression of a single protein which has a single or limited number of targets. Both of these approaches as already mentioned are often cell line and product specific. With the development of novel therapeutic proteins which can sometimes suffer from low product titres and increased understanding of the processes involved in recombinant protein production there will be both the product and the knowledge for identifying production bottlenecks and rationally designing more sophisticated cell engineering strategies for improving recombinant protein production in mammalian cells. This will be aided by the further development of –omics approaches and the development of mathematical models which can be used to both describe cellular processes important in recombinant protein production and also be used to identify potential engineering targets (Pybus *et al.*, 2013; Baik *et al.*, 2011; Doolan *et al.*, 2010; Wong *et al.*, 2006a; Wong *et al.*, 2006b). For some products for improved production there will likely be the need for the expression of multiple genes/proteins to improve multiple cellular processes. This will require novel and more sophisticated strategies for controlled multi-gene expression.

2.2.5 – Multi-gene engineering in mammalian cells

There are very few published multiple gene engineering strategies which can be used to control/tune the expression of multiple genes simultaneously in mammalian cells. This has been done more frequently in lower organisms such as prokaryotes and eukaryotic microbes (Dehli *et al.*, 2012; Wang *et al.*, 2009).

One interesting strategy which was carried out in CHO cells showed that the expression of three different proteins could be controlled by light inducible promoters which responded to different wavelengths of light (Muller *et al.*, 2013). The time of induction of each individual gene could be controlled by exposure to these different wavelengths of light. However, they did not demonstrate that they could control the actual levels of protein expression.

Another study showed that they could control the level of expression of two proteins simultaneously by engineering the *cis*-regulatory 5'-UTR (untranslated region) of each proteins corresponding mRNA (Endo *et al.*, 2013). Through this they were able to tune the translational efficiency of each proteins mRNA and simultaneously control the expression of two proteins. This was carried out in HeLa cells but due to their mechanism of control it should be easily transferable to other mammalian cells like CHO cells.

The development of further novel systems for the controlled/tunable of multiple genes will likely rely on the future developments in mammalian synthetic biology which is an exciting and fast growing research area (Grushkin, 2012).

Chapter 3 – Utilizing Elements of the UPR to Create a Tunable Protein Expression System for Mammalian Cells

3.1 – Abstract

In this chapter the development of a system for controlling the expression of one or more proteins in mammalian cells using elements of the mammalian unfolded protein response (UPR) is described. Over-expression of the active form of ATF6 (ATF6(50)) was used along with vectors for expression of SEAP/GFP containing either none or multiple copies (1x, 3x, 6x and 9x) of the ATF6 binding site, the ER stress response element (ERSE), to manipulate SEAP and GFP expression both separately and simultaneously in CHO cells.

Increasing amounts of an ATF6(50) vector increased SEAP expression when ERSEs were present via transactivation and decreased SEAP expression when ERSEs were not present by 'promoter interference'. We took advantage of this 'promoter interference' effect to increase the range of expression of our system. We also showed that increasing numbers of ERSEs, from 1x to 9xERSE, and co-expression of ATF6(50) caused increases in both SEAP and GFP expression.

This system was then used to control the expression of both SEAP and GFP simultaneously. Again this involved using different amounts of ATF6(50) along with differing numbers of ERSEs. Nine different combinations of simultaneous SEAP and GFP expression were achieved.

Further development is required to make this a more viable system for the controlled expression of one/multiple genes but these are the first steps towards a system which could not only be useful in the biopharmaceutical industry for cell line engineering strategies, for so called 'difficult to express' proteins as one example, but also in other areas of biotechnology and biological research.

3.2 – Introduction

Previously elements of the UPR, such as the UPR activators XBP1 and ATF6, have been used to try and increase the productivity of mammalian cells producing therapeutic recombinant proteins with varying degrees of success (Cain *et al.*, 2013; Pybus *et al.*, 2013; Becker *et al.*, 2008; Campos-da-Paz *et al.*, 2008; Ku *et al.*, 2008; Tigges & Fussenegger, 2006). Their DNA binding elements, ERSE and UPRE, have been inserted into DNA vectors for the detection and quantification of endogenous and induced UPR's (Du *et al.*, 2013). Endogenously these elements combine in the activation of UPR target genes to help alleviate an accumulation of unfolded protein but they have not been used together as part of a cell/synthetic biology engineering strategy (see figure 2.1 for a summary on the UPR).

The work in this chapter combined transient expression of a recombinant form of ATF6 with co-expression of a recombinant reporter protein (either SEAP (secreted alkaline phosphatase) a secreted glycoprotein or GFP (green fluorescent protein) an intracellular protein) which contains repeats of the ATF6 binding element (ERSE) upstream of its promoter (see figure 3.1).

ATF6 and ERSE were chosen due to them being known to endogenously act together for the activation of UPR target genes (Yoshida *et al.*, 2001b; Li *et al.*, 2000; Ubeda & Habener, 2000). The ERSE sequence is also relatively short (19 bp) (Yoshida *et al.*, 1998) meaning its number can be varied without creating extremely large vectors or vectors that vary to greatly in size. ATF6 was also shown to have a more of positive effect, in terms of increased expression, when co-expressed with ERSE containing vectors than XBP1 (figure 3.3).

It was hypothesized that vectors containing ERSEs would respond via transactivation by ATF6 co-expression and that by using differing numbers of ERSEs (1x, 3x, 6x and 9xERSE) as well as different amounts of ATF6 there would be different levels of transcriptional activation resulting in controlled expression of SEAP/GFP. This could be described as a *tunable protein expression system for mammalian cells*.

Although gene expression can be controlled, via gene copy number/DNA vector load, the majority of the time strong constitutive promoters are used and result in near 'all

or nothing' responses making it difficult to achieve discrete/fine control over expression when it is desired. *However, tunable expression allows for such control.*

3.2.1 – Brief Background – Tunable Promoters

Tunable protein expression is not a new idea and two of the oldest inducible expression systems, the Tet-On/Off (tetracycline-On/Off) and metallothionein (metal inducible promoters) systems, date back to the late 80's and early 90's (Gossen & Bujard, 1992; Dickerson *et al.*, 1989; Bunch *et al.*, 1988). There has always been a need in biological research for the controlled expression of recombinant proteins and for the further progression of biological research and especially synthetic biology this requirement is ever increasing. Protein expression can be controlled at the transcriptional level as well as post transcriptionally (i.e. – translation) (Farzadfard *et al.*, 2013; Dehli *et al.*, 2012; Stapleton *et al.*, 2012; Tigges *et al.*, 2009; Mijakovic *et al.*, 2005). This can involve the redesign/alteration of endogenous promoters as well as the creation of completely synthetic promoters (Yim *et al.*, 2013; Qin *et al.*, 2011; Alper *et al.*, 2005; Tornoe *et al.*, 2002). It can also involve co-expression or the presence of an effector molecule such as a transcription factor or chemical and also the insertion of post transcriptional regulatory sequences to control protein expression (Farzadfard *et al.*, 2013; Stapleton *et al.*, 2012; Tigges *et al.*, 2009; Bulter *et al.*, 2004).

Discrete control of protein expression has been achieved in the form of inducible/repressible promoters for time dependent expression (Blount *et al.*, 2012; Hurley *et al.*, 2012; Bowers *et al.*, 2004; Bateman *et al.*, 2001) and oscillator systems for mimicking biological rhythms (Purcell *et al.*, 2010; Tigges *et al.*, 2009; Stricker *et al.*, 2008). Also libraries containing large numbers of promoters that vary in strength in slight increments (Qin *et al.*, 2011; Tornoe *et al.*, 2002; Jensen & Hammer, 1998), but sometimes over several orders of magnitude, have been essential in areas such as metabolic control analysis (Solem *et al.*, 2008; Solem & Jensen, 2002).

One of the future challenges for this area of research is towards the discrete control of multiple genes in a system (Dehli *et al.*, 2012; Mijakovic *et al.*, 2005) and there already has been some successful research published (Muller *et al.*, 2013; Wang *et al.*, 2009).

The majority of the research has been done in prokaryotic and eukaryotic microbial systems. Research has been done in mammalian systems but there is less published literature. However this is changing and mammalian synthetic biology is currently one of the strongest and most innovative biological disciplines (Grushkin, 2012).

Synthetic biology, in the near and distant future, will have an ever increasing impact on not only biological research but also on practical applications in both the medical and biotechnology industries.

3.2.2 – Description of the tunable protein expression system

The system (fig 3.1) is currently a transiently expressed two/three vector system, one vector containing the gene for recombinant ATF6 and the other/s containing the gene for the reporter protein, either SEAP or GFP.

The ATF6 vector utilizes the CMV promoter for strong constitutive expression of ATF6. An ATF6 vector utilizing an SV40 promoter was also created but was not used due to it not having any effect on reporter vectors containing ERSEs. This was probably due to the weaker expression which is usually seen from a SV40 promoter compared with the CMV in some cell lines and which is demonstrated later in this chapter.

The SEAP and GFP reporter vectors contained an SV40 promoter and ERSE sequences were inserted upstream of this. The SV40 promoter was chosen for the expression of the reporter proteins due to its weak expression capabilities. This made it more amenable to manipulation of its transcriptional strength. ERSEs were also inserted upstream of the CMV promoter in a SEAP expressing vector but when co-expressed with the ATF6 vector no effect on SEAP expression was observed. Again this probably occurred due to the strength of the CMV promoter.

The ERSE sequence is 19 base pairs (bp) long (5' – CCAATCGGAGGCCTCCACG – 3') (Yoshida *et al.*, 1998). The active form of ATF6 binds the CCACG part of the sequence while the constitutively expressed NF-Y transcription factor binds CCAAT with both being required for transcriptional activation (Yoshida *et al.*, 2001b; Li *et al.*, 2000; Ubeda & Habener, 2000). Both SEAP and GFP vectors containing 1x, 3x, 6x and 9x

copies of the ERSE sequence were created using a method for the unidirectional (same 5' to 3' orientation) insertion of repeated DNA sequences (Jobbagy *et al.*, 2002). Multiple ERSE sequences were separated by 13 bp. We expect reporter vectors containing more ERSE sequences and therefore more ATF6 binding sites to give higher levels of reporter protein expression when ATF6 is co-expressed.

The ATF6 vector and the reporter vector/s are co-transfected together and transiently expressed. Figure 3.1 illustrates the hypothesized mechanism of our system.

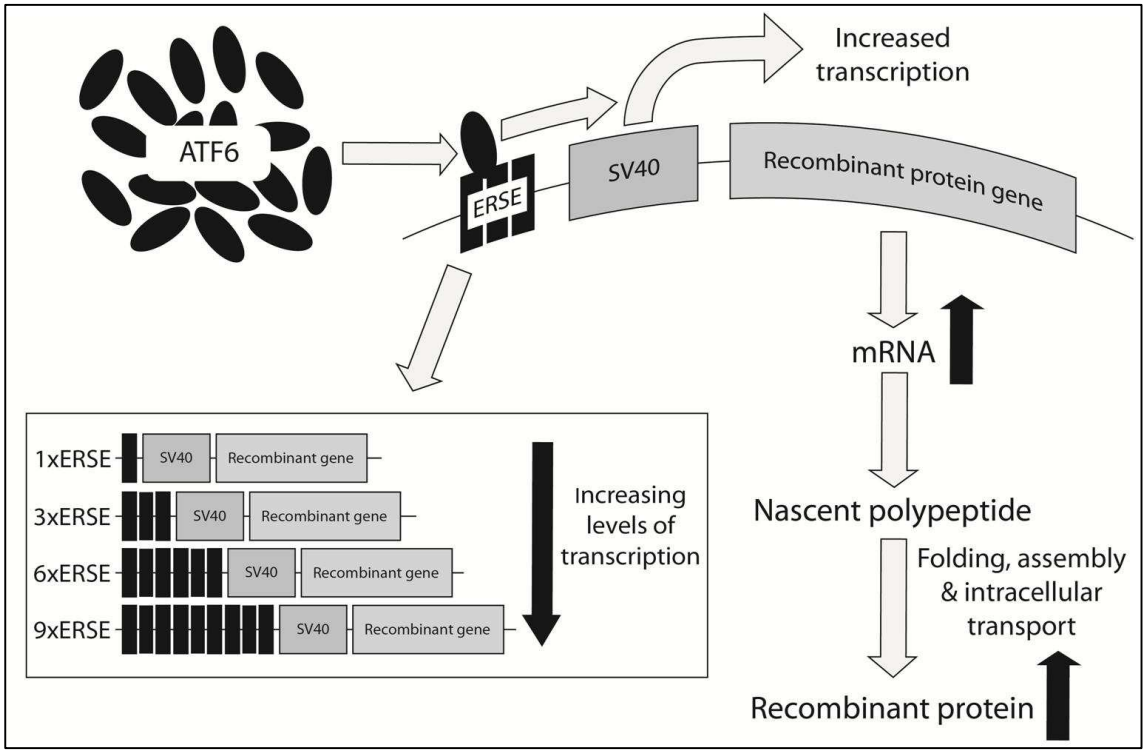

Figure 3.1 – Diagram of the tunable protein expression system. The diagram shows how the system will function. Transfection of a DNA vector containing the ATF6 gene will lead to the expression of recombinant ATF6 protein. The active form of ATF6 will then be able to bind ERSE elements upstream of the SV40 promoter in the co-transfected reporter vector and cause an increase in transcription from this vector. Altering the amount of ATF6 and number of ERSE will lead to different levels of mRNA transcription and overall recombinant protein expression.

3.3 – Aims and Objectives

Using the system created, we aim to discretely control protein expression in mammalian cells by manipulating the transcriptional activity of a commonly used promoter using components of the UPR (ERSE and ATF6). We also aim to show that it can be used to control the expression of two recombinant proteins simultaneously. All the work carried out has been done using a transient expression platform. A more detailed description of our system can be found in the previous section of this chapter.

The aims/objectives of this chapter are as follows:

1. Show that expression of recombinant ATF6 is capable of transactivating vectors containing ERSE upstream of their SV40 promoter
2. Show that changing the amount of ATF6 and the number of ERSE can lead to differing levels of expression
3. Show that this can be achieved for both a secreted (SEAP) and intracellular (GFP) reporter protein
4. Show that our system can be used to control the expression of two reporter proteins, a secreted (SEAP) and intracellular (GFP) reporter protein simultaneously

3.4 – Materials and Methods

3.4.1 – Cell line and cell culture

CHOK1SV cells (Lonza Biologics, Slough, UK) were cultured in Erlenmeyer shake flasks (Corning, Surrey, UK) using CD-CHO medium (Gibco®, Life Technologies, Paisley, UK) supplemented with 6 mM L-glutamine (Gibco®, Life Technologies, Paisley, UK). Cells were cultured in 125 or 250 mL cap-vented Erlenmeyer flasks in a shaking incubator set at 37°C, 5% CO_2 and 140 rpm. Cells were passaged every 3-4 days. Viability (%) and viable cell concentration (cells mL^{-1}) were measured using the Vi-Cell automated cell counter (Beckman Coulter, Brea, CA, USA). The required amount of cells for a concentration of 0.2 $x10^6$ cells mL^{-1} in the new culture were centrifuged at 200 g for 5 minutes. The old medium was removed and the cells were resuspended in fresh medium and added to a new flask.

3.4.2 – DNA vectors used and vector engineering

The UPR vectors ATF6(90), ATF6(50), XBP1µ and XBP1s were created by Dr. S. Schlatter. All four genes are based on the mouse (*Mus musculas*) coding DNA sequences (CDS) and were inserted into the pcDNA3.1 (+) vector backbone (Life Technologies, Paisley, UK). The ATF6(90) vector encodes the transcriptionally inactive, full ER membrane inserted form of the ATF6 protein (Ye *et al.*, 2000; Haze *et al.*, 1999). ATF6(50) encodes the cleaved, transcriptionally active form of the protein (Ye *et al.*, 2000; Haze *et al.*, 1999). XBP1µ encodes for un-spliced XBP1 mRNA which is translated into the transcriptionally un-active, shorter (33 kDa) form of the XBP1 protein (Yoshida *et al.*, 2001a). The XBP1s vector encodes for spliced XBP1 mRNA and this translates to the transcriptionally active and longer (54 kDa) form of the protein (Yoshida *et al.*, 2001a).

6 different SEAP vectors were used in this chapter, SV40-SEAP (control), 1xERSE-SV40-SEAP, 3xERSE-SV40-SEAP, 6xERSE-SV40-SEAP, 9xERSE-SV40-SEAP and CMV-SEAP. The control vector SV40-SEAP is the Clonetech (Mountain View, CA, US) pSEAP2-control vector. To create the ERSE vectors, oligonucleotides containing the ERSE sequence, were inserted upstream of the SV40 promoter in the SEAP control vector using a

method described by Jobbagy *et al.*, (2002) for the unidirectional (same orientation) insertion of repeated DNA sequences. To create the CMV-SEAP vector the SV40 enhancer from the SV40-SEAP vector was first removed by restriction digest and the vector self-ligated. The SV40 promoter was then also removed by restriction digest. The CMV promoter was PCR amplified from the ATF6(50) vector using primers with restriction site overhangs which matched those used to remove the SV40 promoter. After digesting the CMV PCR fragment with the correct restriction enzymes it was ligated into the linear SEAP vector.

Six different GFP vectors were also used, SV40-GFP (control), 1xERSE-SV40-GFP, 3xERSE-SV40-GFP, 6xERSE-SV40-GFP, 9xERSE-SV40-GFP and CMV-GFP. These were created using the corresponding SEAP vector as a backbone for the insertion of a GFP CDS. The SEAP CDS was removed from the SEAP vectors by restriction digest. The GFP CDS was PCR amplified from the pTurboGFP-N vector (Evrogen, Moscow, Russia) using primers with restriction overhangs. The GFP PCR fragments then underwent restriction digest before being ligated into the corresponding vector backbone.

A negative control DNA vector containing no promoter or CDS was created by removing these elements by restriction digest from the SV40-SEAP vector and then ligating the sticky ends of the remaining vector backbone. This vector was used as a control to maintain the same DNA loads in all transfection experiments as well acting as a negative control for transfection.

Vector DNA was purified using both Miniprep and Maxiprep kits (Qiagen, Manchester, UK) according to the supplied protocol. Vector DNA concentration was measured using the NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, UK).

3.4.3 – Transient transfections

Transfections were carried out on CHOK1SV cells which had been in culture for at least 2-3 passages but no longer then 8 weeks. The transfection reagent used was Lipofectamine® 2000 (Life Technologies, Paisley, UK).

Cells on day 3 of culture were measured for cell concentration and viability. 1 mL of CHO cells at a concentration of 0.2 x 10^6 cells mL^{-1} were then plated on a 24 well plates (Nunclon™ Delta Surface microwell plates, Thermo Fisher Scientific, UK) in DMEM medium (Gibco®, Life Technologies, Paisley, UK) containing 10% FCS (foetal calf serum) (Gibco®, Life Technologies, Paisley, UK) and 6 mM L-glutamine. Plates were then placed in a static incubator at 37°C and 5% CO_2.

24 hours after plating the cells were transfected with vector DNA. For every transfection 3 µg of DNA and 3 µl of Lipofectamine® 2000 was used per well. The DNA and Lipofectamine® 2000 were diluted seperately in Opti-MEM media (Gibco®, Life Technologies, Paisley, UK) and were both prepared, mixed and added to the culture plates according to the manufactures supplied protocol and returned to the same incubator.

For SEAP vector transfections the culture plates were incubated for 48 hours post transfection before spinning the plates down (5 minutes at 200 g) and carefully removing the cell media for further analysis.

For GFP vector transfections the culture plates were analysed for GFP expression at 24 and 48 hours post transfection.

Transfection efficiency based on GFP expression was analysed using the Attune® Acoustic Focusing Cytometer and the Attune® Cytometric Software (Life Technologies, Paisley, UK).

3.4.4 – SEAP and GFP analysis

The AnaSpec SensoLyte® pNPP Secreted Alkaline Phosphatase Reporter Gene Assay Kit (*Colorimetric*) (Cambridge Bioscience, Cambridge, UK) was used to quantify SEAP protein. The assay was carried out according to the supplied protocol. SEAP samples were incubated in a water bath at 65 °C for 20 minutes to inactivate endogenous non-specific alkaline phosphatase.

GFP expression was measured using the FluoroSkan Ascent™ FL Microplate Fluorometer (Themro Scientific, UK). GFP was measured while the cells were still in their culture plates. GFP was excited at 485 nm and emission measured at 510 nm.

3.4.5 – Quantifying gene expression by real-time RT-qPCR

Cells were centrifuged at 200 g for 5 minutes and the cell medium was carefully removed. Extraction of mRNA was carried out using the RNeasy® Minikit in combination with the QIAShredder homogeniser according to the manufacturers' protocol (Qiagen, Manchester, UK).

mRNA concentration was measured using the NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, UK).

The TURBO DNA-*free*™ Kit (Applied Biosystems®, Life Technologies) was used to remove any contaminating DNA from mRNA samples and was used according to the supplied protocol. After removal of DNA, mRNA concentrations were measured again using the NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, UK).

mRNA was transcribed into DNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems®, Life Technologies) following the manufacturers protocol.

SEAP mRNA expression was quantified by RT-qPCR using the Applied Biosystems 7500 Fast Real-Time PCR system (Applied Biosystems®, Life Technologies) using an absolute quantification method (Pfaffl *et al.*, 2002). The SV40-SEAP vector was used as a standard.

SEAP primers were designed against the SEAP coding sequence and used at a concentration of 300 nM (SEAP FWD 5' – CCATATGTGGCTCTGTCCAA – 3', SEAP REV 5' – GTCTGGAAGTTGCCCTTGAC – 3').

The Power SYBR® Green PCR Mastermix (Applied Biosystems®, Life Technologies) was used according to the manufacturer's instructions. 1 µl of cDNA was added to each reaction. The following PCR cycle program was used: 1. 50°C for 2 minutes; 2. 95°C for 10 minutes; then 40 repeats of 3. 95°C for 15 seconds then 60°C for 1 minute. A dissociation step was added to the end of procedure to check for the specificity of amplification.

3.5 – Results and Discussion

3.5.1 – SEAP vector titrations

The SEAP vectors CMV-SEAP, SV40-SEAP and 9xERSE-SV40-SEAP were transfected over a range of DNA concentrations (0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg, 2.0 µg and 3.0 µg) to show the relative differences in their expression in CHOK1SV cells and whether the addition of ERSE alone had any significant impact on SEAP expression.

From figure 3.2 it can be seen that the CMV-SEAP vector produced more SEAP protein across the whole range of transfected amounts of DNA compared to both the SV40-SEAP and 9xERSE-SV40-SEAP vectors. Over the range of 0.03125 µg to 1.0 µg SEAP vector DNA the CMV-SEAP vector produced ≥1.5-fold more SEAP protein compared with the other vectors. SEAP expression from the CMV-SEAP vector plateaued at 1.0 µg of DNA. The SV40-SEAP and 9xERSE-SV40-SEAP vectors produced similar levels of SEAP protein for 0.03125 to 1.0 µg of transfected DNA. The SV40-SEAP vector produced slightly more SEAP than the 9xERSE-SV40-SEAP vector when 2.0 µg and 3.0 µg of SEAP DNA was transfected. However, unlike the CMV-SEAP vector the SV40-SEAP or 9xERSE-SV40-SEAP vectors protein expression did not plateau.

In summary, the results show that SEAP production is higher when a CMV promoter is used to initiate transcription. This was expected and it is already known that the CMV promoter is a stronger promoter in many cell lines when compared to the SV40 promoter (Zarrin *et al.*, 1999; Liu *et al.*, 1997; Foecking & Hofstetter, 1986). More importantly these results also showed that the addition of nine ERSE sequences upstream of the SV40 promoter alone had no positive effect on SEAP protein expression. These results were used to help guide further experiments.

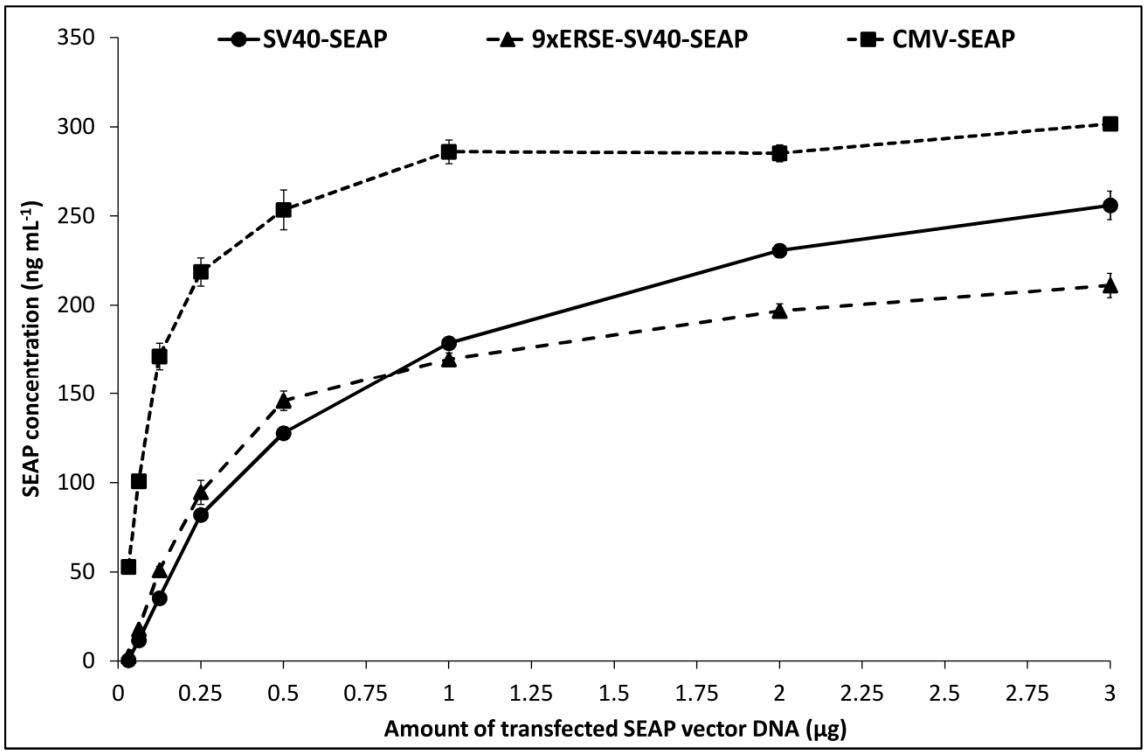

Figure 3.2 – Comparison of SEAP protein expression from three different SEAP DNA vectors over a range of transfected DNA amounts. SEAP protein expression for eight differing amounts (0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg, 2.0 µg and 3.0 µg) of three SEAP expression vectors (● SV40-SEAP, ▲ 9xERSE-SV40-SEAP, ■ CMV-SEAP) transiently expressed in CHOK1SV cells in 24 well plates. The total amount of transfected DNA was kept consistent in all transfections by using an empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. N = 3, error bars represent ± 1 S.D.

3.5.2 – The effect of UPR activators on SEAP expression

To test the effect of expressing UPR activators on SEAP expression the UPR vectors, ATF6(90), ATF6(50), XBP1μ and XBP1s were transiently co-expressed with three SEAP vectors, CMV-SEAP, SV40-SEAP and 9xERSE-SV40-SEAP (figure 3.3). See chapter 2, figure 2.1 or chapter 3, section 3.2.2 for more information on these UPR components and the UPR vectors. This experiment will help towards accomplishing the first aim set out in this chapter – can ATF6 transactivate vectors containing ERSE upstream of an SV40 promoter?

All SEAP vectors were co-transfected with an equal mass of all four of the UPR vectors. From figure 3.3 we can see that the SV40-SEAP vector control produced a similar level of SEAP expression when compared with the results in figure 3.2 (0.5 μg DNA). When the SV40-SEAP vector was co-transfected with the four UPR vectors no increase in SEAP expression was observed when compared to the control. ATF6(90), ATF6(50) and XBP1μ co-expression resulted in a significant decrease in expression compared to the control. The decrease was greater than 2-fold for the ATF6 vectors and around 1.6-fold for XBP1μ. Co-expression of XBP1s had no noticeable effect and SEAP expression was very similar to the control.

The control for the CMV-SEAP vector also produced similar results to that seen in figure 3.2 (0.5 μg DNA). When the UPR activator vectors were co-transfected with the CMV-SEAP vector, the effects were very similar and a roughly 1.3-fold decrease in SEAP expression was seen with all combinations.

For the 9xERSE-SV40-SEAP vector, figure 3.3 shows that the control again produced similar levels of SEAP expression when compared with figure 3.2 (0.5 μg DNA). Similar to results for SV40-SEAP there was a significant decrease in SEAP expression when ATF6(90) (-1.9-fold) and XBP1μ (-1.4-fold) were co-expressed. XBP1s again had no effect on SEAP expression when compared to the control. However, ATF6(50) caused a significant increase in SEAP expression (1.3-fold) when co-expressed with 9xERSE-SV40-SEAP. This shows that the ATF6(50) vector, which produces the transcriptionally active form of the ATF6 protein, is the only UPR activator that can increase SEAP expression from vectors containing ERSE sequences above that of its control.

The increase in SEAP expression seen when ATF6(50) was co-expressed with 9xERSE-SV40-SEAP was the only increase seen in this experiment. Although this was only a 1.3-fold increase, this is only when compared to the expression of the 9xERSE-SV40-SEAP vector under control conditions. To further understand the effects of the combination of ATF6(50) expression and ERSE sequences being present upstream of an SV40 promoter, comparison between the results of the SV40-SEAP, 9xERSE-SV40-SEAP and CMV-SEAP vectors is needed. SEAP expression was similar between CMV-SEAP and 9xERSE-SV40-SEAP when co-expressed with ATF6(50). This shows that inserting ERSE sequences in a vector containing a weaker promoter (SV40), can produce similar levels of SEAP expression to that of a vector using a stronger promoter (CMV) when both are in the presence of ATF6(50). When comparing the SEAP expression of ATF6(50) + 9xERSE-SV40-SEAP to the control CMV-SEAP the 9xERSE-SV40-SEAP vector produced less SEAP.

Comparison of the control for both SV40-SEAP and 9xERSE-SV40-SEAP showed that their levels of SEAP expression were extremely similar. This again shows that the presence of 9xERSE alone has no effect on SEAP expression from a vector using the SV40 promoter. However, when comparing SEAP expression between the SV40-SEAP and 9xERSE-SV40-SEAP vectors when ATF6(50) is co-transfected there is a 5.9-fold difference. This difference in SEAP protein expression is a better reflection on the effect of the insertion of 9xERSE sequences upstream of the SV40 promoter and the response to ATF6(50). This is also evidence towards showing that recombinant ATF6(50) is capable of transactivating a vector containing ERSE upstream of a SV40 promoter and completing the first aim set out in this chapter.

As well as discussing the increases in expression seen in this experiment it is also worth discussing the results where there were decreases SEAP expression with certain combinations of SEAP vectors and UPR activators.

It is expected when two DNA vectors are co-transfected that expression is reduced compared to if they were transfected alone. This is seen when the SV40-SEAP vector was co-expressed with ATF6(90), ATF6(50) and XBP1μ and the 9xERSE-SV40-SEAP vector is co-expressed with ATF6(90) and XBP1μ. The level of reduction in expression

can also be affected by the strength of the promoters in the different vectors. For example the CMV promoter is expected to negatively affect the expression of a gene from a weaker SV40 promoter more so than co-expression with another CMV driven gene (Huliak *et al.*, 2012). The UPR activator vectors are driven by a CMV promoter and this pattern can be seen from results of the different SEAP vectors in figure 3.3. This effect could be termed 'promoter interference', although it is poorly defined in the literature and could also be termed 'promoter competition' and 'transcriptional interference' amongst other names and seems to depend on the context of the research and the researchers (Huliak *et al.*, 2012; Curtin *et al.*, 2008; Shearwin *et al.*, 2005; Conte *et al.*, 2002; Eszterhas *et al.*, 2002; Hirschman *et al.*, 1988; Cullen *et al.*, 1984). In brief, for the purpose of this chapter 'promoter interference' shall be defined as the negative effects promoters on different vectors have on each other and there resulting expression. For a more detailed description of 'promoter interference' see chapter 4, section 4.2.1. The results in chapters 4 & 5 of this thesis also investigate this phenomenon further, but there will be further references to it in the remainder of this chapter.

'Promoter interference' did not occur for all combinations of SEAP and UPR activator vectors, e.g. - XBP1s co-transfection with SV40-SEAP or 9xERSE-SV40-SEAP. This shows that the effect of the gene being expressed should also be considered. The XBP1s vector produces the transcriptionally active form of the XBP1 protein. This could potentially induce the expression of endogenous XBP1/UPR target genes which encode proteins which are important in protein folding and secretion and enhance the cells protein synthesis capability and this has been demonstrated previously by XBP1s over-expression increasing recombinant protein production (Cain *et al.*, 2013; Campos-da-Paz *et al.*, 2008; Tigges & Fussenegger, 2006).

In summary, it was shown that out of the four UPR activators only ATF6(50) can increase the expression of a SEAP vector containing ERSEs upstream of an SV40 promoter. Therefore the remainder of this chapter will only involve co-expression of the ATF6(50) vector for further investigation of the potential of this tunable expression system.

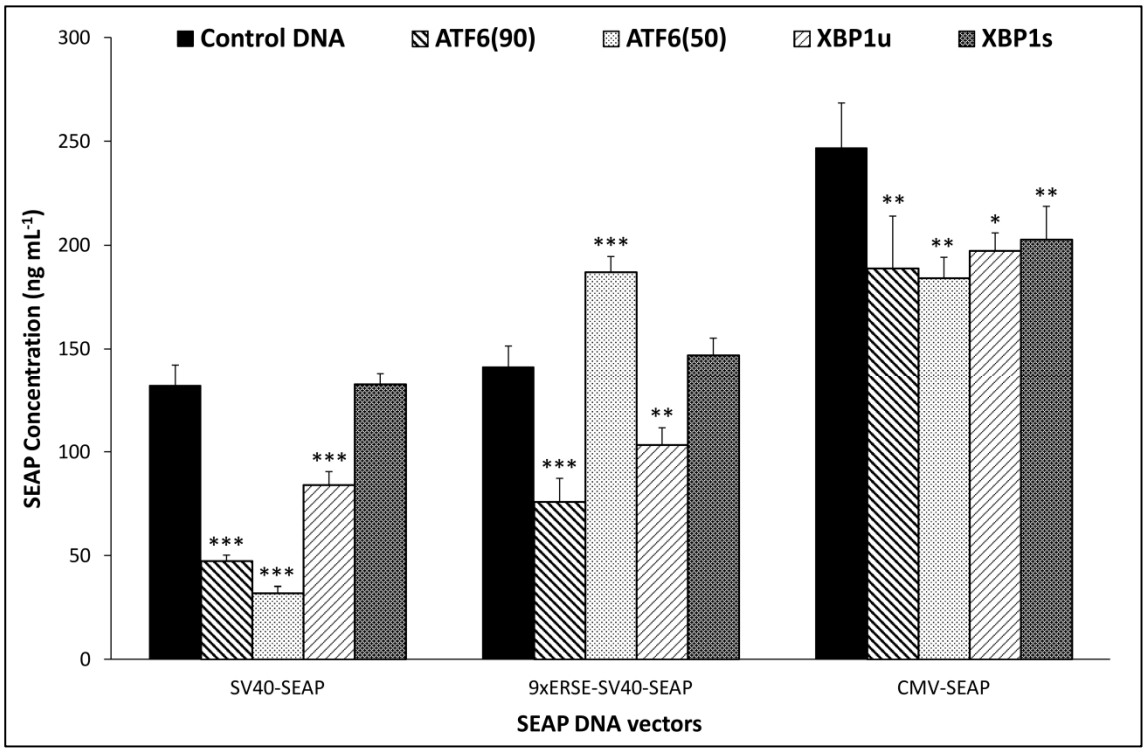

Figure 3.3 – The effect of ATF6(90), ATF6(50), XBP1µ and XBP1 on SEAP protein expression from three different SEAP DNA vectors. SEAP protein expression for 0.5 µg of SV40-SEAP, 9xERSE-SV40-SEAP and CMV-SEAP vector DNA co-transfected with the 0.5 µg of UPR activator vectors (ATF6(90), ATF6(50), XBP1u and XBP1s) in CHOK1SV cells. The total amount of transfected DNA was kept constant using the –ve control DNA vector. For further information on the SEAP, UPR activator and -ve control DNA vectors please refer to section "4.2.2" in this chapter. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. For each SEAP vector, mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, error bars represent + 1 S.D.

3.5.3 – The effect of titrating ATF6(50) on SEAP expression

The next experiment was carried out to investigate the effects of titrating the amount of ATF6(50) vector on SEAP expression. To do this different amounts of the ATF6(50) vector (0.03125 μg, 0.0625 μg, 0.125 μg, 0.25 μg, and 0.5 μg) were co-transfected with a constant amount of SEAP vector (0.5 μg). This was carried out for SV40-SEAP, 9xERSE-SV40-SEAP and CMV-SEAP. This will aid in fulfilling the first and second part of the aims/objectives set at the beginning of this chapter.

From figure 3.4 we can see that the controls for each SEAP vector again produced similar amounts of SEAP expression when compared to both figure 3.2 and 3.3. Increasing amounts of ATF6(50) had increasingly negative effects on SEAP expression from the SV40-SEAP vector. When compared to the control the difference in expression was significant for all amounts of ATF6(50) vector DNA, except for the smallest amount (0.03125 μg). For the CMV-SEAP vector, although some of the larger amounts of ATF6(50) did reduce the amount of SEAP expression, none of the decreases were statistically significant. As in figure 3.3 we can see that ATF6(50) co-expression with 9xERSE-SV40-SEAP again increased SEAP expression when compared to the control (figure 3.4). For all amounts of ATF6(50) the increases in SEAP expression were significant.

This again shows that a vector containing ERSE upstream of an SV40 promoter (9xERSE-SV40-SEAP) is capable of responding to the co-expression of the transcriptionally active form of ATF6 (ATF6(50)). When only the SV40 promoter is present (SV40-SEAP) SEAP expression is reduced when co-expressed with the ATF6(50) vector. Again this is likely to occur due to 'promoter interference'.

From table 3.1 we can see the fold differences when the amount of co-transfected ATF6(50) is adjusted. For the 9xERSE-SV40-SEAP vector we can see that increasing the amount of ATF6(50) vector from 0.03125 μg to 0.5 μg does increase the amount of SEAP expression but this plateaus at 0.125 μg with a 1.26-fold increase. In this instance this gives us only a very limited window in which to adjust the expression of SEAP and would make our system almost pointless for most applications. However, when you look at the effects on the SV40-SEAP vector you can see a decreasing step-

wise trend in SEAP expression when the amount of the ATF6(50) vector is increased. Overall there is a 3.81-fold decrease in SEAP expression when comparing the control and co-transfection of 0.5 μg of ATF6(50). When you compare this with the highest level of SEAP expression achieved when ATF6(50) (0.25 μg) was co-expressed with 9xERSE-SV40-SEAP there is a 5.35-fold increase from the lowest SEAP expression to the highest. This gives a much broader range in which to control the expression of SEAP. Table 3.2 shows this range of SEAP expression.

In summary, we have again shown that SEAP vectors containing ERSE upstream of an SV40 promoter are capable of responding to the co-transfection of ATF6(50). We can also see that by altering the amount of transfected ATF6(50) we change the level of SEAP expression. This can be achieved for both the 9xERSE-SV40-SEAP and the SV40-SEAP vector by using ATF6(50) and transactivation and 'promoter interference', respectively to improve the range of attainable SEAP expression. This provides further evidence towards achieving both the first and second aims/objectives of this chapter.

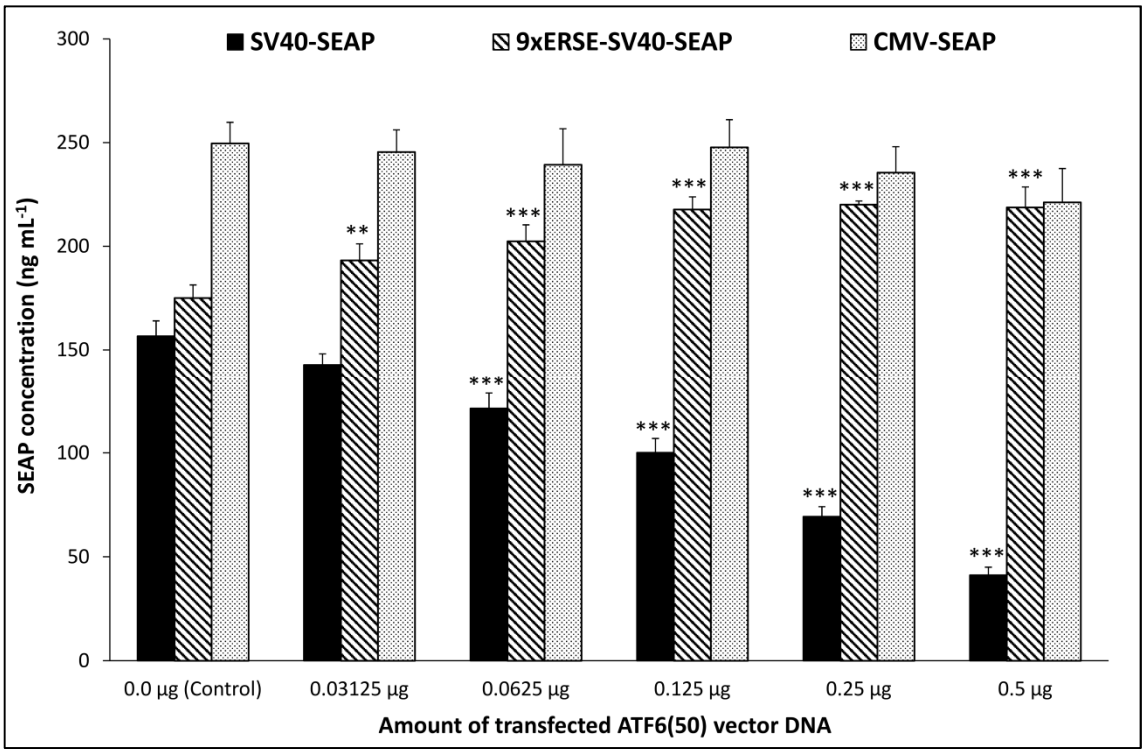

Figure 3.4 – The effect of differing amounts of ATF6(50) vector DNA on SEAP protein expression from 0.5 µg of three different SEAP DNA vectors. SEAP protein expression from 0.5 µg of SV40-SEAP, 9xERSE-SV40-SEAP and CMV-SEAP vector DNA when co-transfected with differing amounts of ATF6(50) vector DNA (0.0 µg (control), 0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg) in CHOK1SV cells. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. For each SEAP vector, mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, Error bars represent + 1 S.D.

Table 3.1 – Fold change in SEAP expression for each SEAP vector when compared to their control from figure 3.4

	SEAP vectors		
ATF6(50)	SV40-SEAP	9xERSE-SV40-SEAP	CMV-SEAP
0.0 µg (Control)	1.00[a]	1.00[a]	1.00[a]
0.031255 µg	-1.10[a]	1.10[a, b]	-1.02[a]
0.0625 µg	-1.30[b]	1.16[b, c]	-1.04[a]
0.125 µg	-1.57[c]	1.24[c]	-1.01[a]
0.25 µg	-2.26[d]	1.26[c]	-1.06[a]
0.5 µg	-3.81[e]	1.25[c]	-1.13[a]

The numbers in the table represent the fold change in SEAP protein expression from the control sample of the respective SEAP vectors. Calculated using the data in figure 3.4. For each SEAP vector, values with different letters differ significantly from each other (Tukey's test, $p < 0.05$). Comparisons between different SEAP vectors were not made.

Table 3.2 – Range of SEAP expression from figure 3.4

SEAP vector	ATF6(50)	Fold change
SV40-SEAP	0.5 µg	1.00[a]
SV40-SEAP	0.25 µg	1.68[b]
SV40-SEAP	0.125 µg	2.43[c]
SV40-SEAP	0.0625 µg	2.95[d]
SV40-SEAP	0.03125 µg	3.47[e]
SV40-SEAP	0.0 µg (Control)	3.81[e, f]
9xERSE-SV40-SEAP	0.0 µg (Control)	4.26[f, g]
9xERSE-SV40-SEAP	0.03125 µg	4.70[g, h]
9xERSE-SV40-SEAP	0.0625 µg	4.92[h, i]
9xERSE-SV40-SEAP	0.125 µg	5.29[i]
9xERSE-SV40-SEAP	0.25 µg	5.35[i]
9xERSE-SV40-SEAP	0.5 µg	5.32[i]

The numbers in the table represent the fold change in SEAP protein expression when compared to lowest expressing combination of SEAP vector and ATF6(50). Calculated using the data in figure 3.4. For each SEAP vector, values with different letters differ significantly from each other (Tukey's test, $p < 0.05$).

3.5.4 – The effect of different numbers of ERSE on SEAP expression

Along with the 9xERSE-SV40-SEAP vector, three other ERSE vectors were also engineered, 1xERSE-SV40-SEAP, 3xERSE-SV40-SEAP and 6xERSE-SV40-SEAP. In this experiment, we investigated whether altering the number of ERSE while keeping the amount of ATF6(50) vector constant would alter the expression of SEAP. As well as measuring SEAP protein expression, RNA samples were also taken to quantify the numbers of SEAP mRNA transcripts that were present. Similar to the last section, this will aid in providing evidence for the first and second aims/objectives.

From figure 3.5 – A. we can see that SV40-SEAP, 1xERSE-SV40-SEAP, 3xERSE-SV40-SEAP, 6xERSE-SV40-SEAP and 9xERSE-SV40-SEAP all produced similar levels of SEAP protein expression when transfected under control conditions (no ATF6(50)). The CMV-SEAP control again produced more SEAP protein when compared with vectors containing an SV40 promoter. This shows again that the addition of ERSE sequences upstream of the SV40 promoter alone does not improve SEAP protein expression.

The ERSE vectors all produced similar levels of SEAP mRNA under control conditions (figure 3.5 – B.). The SV40-SEAP control on average produced less SEAP mRNA when compared to the ERSE vector controls, but this difference was only significant for 6xERSE-SV40-SEAP. From these results it could be possible that the addition of ERSE upstream of an SV40 promoter might increase the transcriptional capacity of this promoter slightly. However, any difference seen at the mRNA level was not observed at the protein level (figure 3.5 – A.) and therefore the addition of ERSE alone does not seem to have a significant effect.

The CMV-SEAP vector produced on average 3.39-fold (± 1.10 (SD)) more SEAP mRNA compared to all the other SEAP vectors under control conditions (figure 3.5 – B.). This demonstrates the stronger transcriptional strength of the CMV promoter when compared to the SV40, but it also worth noting that mRNA stability and processing can be affected by the promoter used (Bregman *et al.*, 2011; Trcek *et al.*, 2011). The same difference in mRNA is not observed at the protein level (figure 3.5 – A.). The CMV-SEAP vector only produced on average 1.74-fold (± 0.05 (SD)) more SEAP protein then the other SEAP vectors. This shows that increases in SEAP mRNA and SEAP protein do

not necessarily correlate at a 1-to-1 ratio. To understand this it is worth considering the differences in SEAP mRNA and protein. SEAP protein is secreted into the cell media and will accumulate throughout culture. Protein also tends to have a greater half-life than mRNA (Vogel & Marcotte, 2012). The reported half-life of SEAP is ~500 hours (Schlatter *et al.*, 2002). Therefore it is safe to assume that the SEAP measurement will represent the overwhelming majority of SEAP protein produced where as the mRNA measurement due to its much shorter half-life (estimated at 10.5 hours (Weber *et al.*, 2007)) will only represent a proportion of the total SEAP mRNA produced throughout the experiment. It worth mentioning again that the different promoters used will also have an effect on the mRNA expression levels due to their transcriptional strength and the stability and processing of the mRNA produced (Bregman *et al.*, 2011; Trcek *et al.*, 2011).

When looking at the effect that ATF6(50) co-transfection had on SEAP expression, figure 3.5 – A. shows again that it is capable of increasing the expression of SEAP protein from vectors containing ERSE upstream of an SV40 promoter. The number of ERSE present in the vector also does have an effect as well. As the number of ERSE increased from 1xERSE to 9xERSE there was an increase in SEAP protein expression. The 9xERSE-SV40-SEAP vector produced the largest amount of SEAP protein of all the ERSE vectors but was only significantly different from the 1xERSE-SV40-SEAP vector. Table 3.3 shows the fold change in SEAP protein expression relative to the SV40-SEAP control. In this experiment it would seem that the difference between having 3xERSE, 6xERSE and 9xERSE and co-transfection of 0.125 µg of ATF6(50) is not significant.

When considering the amount of SEAP mRNA produced when the ERSE vectors were co-transfected with ATF6(50), similar to the protein results, there was an increase in SEAP mRNA compared to the SV40-SEAP control and from 1xERSE to 9xERSE. The fold changes in SEAP mRNA expression were greater than that seen at the protein level (table 3.3) but similar to the protein data 9xERSE-SV40-SEAP produced the most mRNA but this was only significantly different from the 1xERSE-SV40-SEAP vector.

From the data presented it would seem that increasing the numbers of ERSE present upstream of an SV40 promoter and the co-expression of ATF6(50) does lead to an

increase in SEAP mRNA and SEAP protein expression. However, the differences in expression are only significant when comparing the results of the 1xERSE-SV40-SEAP vector to the other three ERSE vectors. Although increases can be seen there was no significant difference in SEAP expression between the 3xERSE-SV40-SEAP, 6xERSE-SV40-SEAP and 9xERSE-SV40-SEAP vectors.

Although the range of control offered by changing the number of ERSE present in our vectors is limited, it has shown that increasing them does affect both SEAP mRNA and protein expression. The work presented so far is evidence for achieving the first and second aims and objectives set out at the beginning of this chapter:-

1. ATF6 is capable (ATF6(50)) of transactivating ERSE when placed upstream of an SV40 promoter. This has been shown by increases in both SEAP mRNA and protein when comparing the ERSE vectors with the SV40-SEAP vector (figure 3.3, 3.4 and 3.5).
2. Changing the amount of co-transfected ATF6(50) does result in different levels of SEAP expression. This is true for both an ERSE vector (9xERSE-SV40-SEAP - ↑ATF6(50) = ↑SEAP) and the SV40-SEAP vector (↑ATF6(50) = ↓SEAP) (figure 3.4). Changing the number of ERSE present while co-expressing ATF6(50) also had an effect on SEAP expression. An increase in both mRNA and protein was observed when increasing the number of ERSE from 1x to 3x to 6x and to 9xERSE (figure 3.5). 9xERSE gave the highest expression but was only significantly different from the 1xERSE vector.

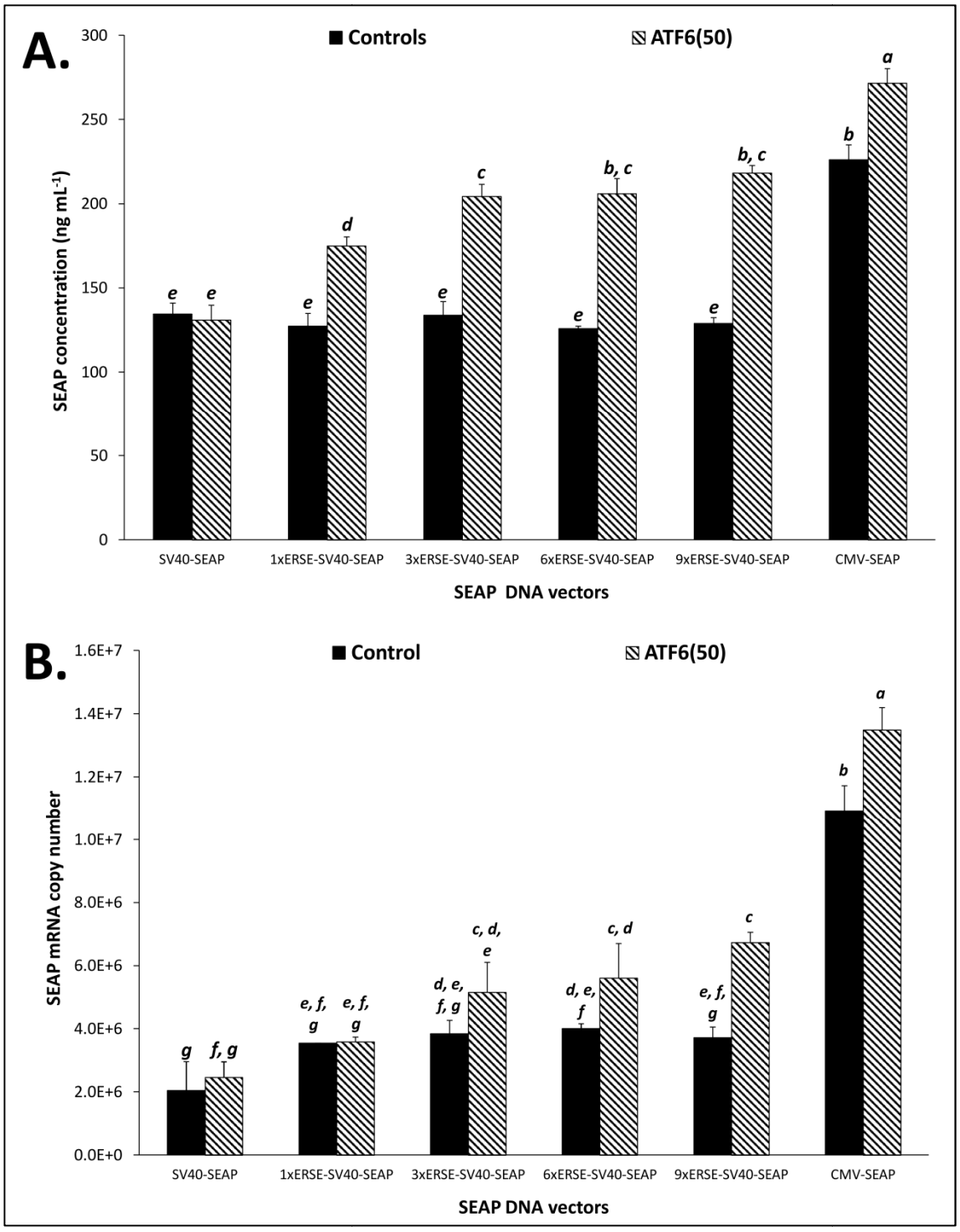

Figure 3.5 – The effect of different numbers of the ERSE sequence on SEAP protein and mRNA copy number when co-transfected with the ATF6(50) DNA vector. A. SEAP protein expression for 0.5 µg of SV40-SEAP, 1xERSE-SV40-SEAP, 3xERSE-SV40-SEAP, 6xERSE-SV40-SEAP, 9xERSE-SV40-SEAP and CMV-SEAP vectors when co-transfected with 0.125 µg ATF6(50) vector DNA in CHOK1SV cells. **B.** SEAP mRNA copy numbers for the different SEAP vectors taken from the same samples as used in **A**. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. For each SEAP vector, values with different letters differ significantly from each other (Tukey's test, $p < 0.05$). N = 3, Error bars represent + 1 S.D.

Table 3.3 – Fold change in SEAP protein and mRNA relative to the SV40-SEAP vector under control conditions

SEAP vector	SEAP protein fold change	SEAP mRNA fold change
SV40 SEAP (Control)	1.00[a]	1.00[a]
1xERSE-SV40-SEAP + ATF6(50)	1.30[b]	1.75[a, b]
3xERSE-SV40-SEAP + ATF6(50)	1.52[c]	2.52[b, c, d]
6xERSE-SV40-SEAP + ATF6(50)	1.53[c, d]	2.74[c, d]
9xERSE-SV40-SEAP + ATF6(50)	1.62[c, d]	3.29[d]

The numbers in the table represent the fold change in SEAP protein and mRNA expression when compared to the SV40-SEAP vector under control conditions. Calculated using the data in figure 3.5. For each SEAP vector, values with different letters differ significantly from each other (Tukey's test, $p < 0.05$).

3.5.5 – The effect of ATF6(50) and ERSE on GFP expression

In the previous sections it was demonstrated that ATF6(50) and different numbers of ERSE upstream of an SV40 promoter could be used to achieve differing levels of SEAP expression. Secreted proteins such as SEAP, as long as they have a long enough half life, will accumulate in the cell media, where as proteins which are not secreted and are only expressed intracellularly may accumulate to a lesser degree due to being retained within and degraded by the cell. GFP is a commonly used intracellular reporter protein and has a reported half-life of ~26 hours but this can vary depending on the variant used (Corish & Tyler-Smith, 1999). In this section investigation of whether the system can be used to control the expression of GFP in the same way as was shown with SEAP was carried out.

Similar to the work carried out using SEAP, six different GFP expression vectors were created. ERSE vectors containing 1x, 3x, 6x and 9xERSE upstream of an SV40 promoter and the GFP CDS were created to see the effect of differing numbers of ERSE. An SV40-GFP and CMV-GFP vector were also created. Both were used to compare the expression of the ERSE vectors, the former to show the effect of ERSE and the latter for comparison to a strong promoter. All GFP vectors were transfected separately with and without the co-transfection of the ATF6(50) vector.

From figure 3.6 (A. & B.) we can see that when the ERSE GFP vectors were transfected under control conditions there was no significant difference in GFP expression when compared to the SV40-GFP vector. Like the work carried out with the SEAP vectors, this showed again that the addition of ERSE upstream of an SV40 promoter alone does not affect expression. Likewise the CMV-GFP vector demonstrated the higher expression strength of the CMV promoter by producing significantly more GFP than the other vectors under control conditions at both time points.

When the ERSE GFP vectors were co-transfected with the ATF6(50) vector the 3xERSE, 6xERSE and 9xERSE-SV40-GFP vectors all showed increased expression compared to their controls at both 24 and 48 hours post-transfection (figure 3.6 – A. & B.). The 1xERSE-SV40-GFP vector produced the same amount of GFP when compared to its control. However, for the SV40-GFP vector there was a decrease in expression

compared to the control when co-transfected with ATF6(50). This was probably due to promoter interference from the strong CMV promoter in the ATF6(50) vector as mentioned earlier in this chapter. This shows that even a vector containing 1xERSE upstream of an SV40 promoter can respond to co-expression of ATF6(50). This is further evidence of ATF6(50) transactivating vectors containing ERSE upstream of an SV40 promoter.

The CMV-GFP vector remained largely unaffected by co-expression of ATF6(50). There were slight decreases in GFP expression when compared to the control, but this was only significant at 48 hours (figure 3.6 – A. & B.).

Taking a closer look at the effect of differing numbers of ERSE on GFP expression in the presence of ATF6(50) table 3.4 shows the fold changes in GFP expression compared with the lowest GFP production (SV40-GFP + ATF6(50)) from figure 3.6 for both 24 and 48 hours post transfection. From this table we can see that between the lowest and highest GFP expression levels achieved, comparing only the GFP vectors containing either an SV40 promoter or both an SV40 promoter and ERSE, that there was a 4.80- and 4.31-fold difference in GFP expression at 24 and 48 hours, respectively. The combination of SV40-GFP and ATF6(50) gave the lowest GFP expression while the 9xERSE-SV40-SEAP + ATF6(50) gave the highest. When increasing the number of ERSE from 1x to 9xERSE there were also increases in GFP expression when ATF6(50) was co-expressed. The differences in GFP expression were significant between the 1xERSE-SV40-GFP vector and the rest of the ERSE GFP vectors, as well as between the 3xERSE-SV40-GFP vector and the 6x and 9xERSE GFP vectors. This again shows that increasing the number of ERSE does lead to increases in protein expression when ATF6(50) is also co-transfected.

The results from the GFP vectors differed in two ways from the results seen in the similar experiment for the SEAP vectors (figure 3.5). Firstly the range of GFP expression shown from the lowest to the highest is greater than for SEAP expression, 4.31-fold and 1.62-fold, respectively at 48 hours. Secondly there were greater significant differences between the different ERSE GFP vectors at 48 hours when co-expressed with ATF6(50) (figure 3.5 – B.) than there was for the ERSE SEAP vectors

(figure 3.4 – A.). One simple explanation why there was differences seen between the ERSE SEAP and ERSE GFP vectors might be due to the amount of ATF6(50) used in both experiments. A lower amount of ATF6(50) DNA was used in the SEAP experiment then than the GFP experiment and we have already shown that increasing amounts of ATF6(50) can have a positive effect on expression from 9xERSE-SV40-SEAP vector (figure 3.4). This is also likely to occur with ERSE GFP vectors. Although this explains some of the difference it will not be the sole reason.

Another explanation could be the differences in the synthesis of GFP and SEAP from transcription right through to the production and expression of the mature protein. Firstly, their final destinations differ significantly, SEAP is secreted and GFP is retained within cells. Secondly, they have very different half-lives, ~500 and ~26 hours respectively for SEAP and GFP (Schlatter *et al.*, 2002; Corish & Tyler-Smith, 1999). Thirdly, the SEAP protein is twice as large as a GFP monomer, ~65 kDa compared to ~27 kDA, but GFP does normally exist as a dimer (Schlaeger *et al.*, 2003; Desai & Person, 1998). Fourthly, SEAP is a secreted glycoprotein (Schlatter *et al.*, 2001) whereas GFP has a beta-barrel structure encasing a chromophore (Ormo *et al.*, 1996; Yang *et al.*, 1996). Due to SEAP's more complex structure and extra processing steps (i.e. – glycosylation and secretion) SEAP maybe more of burden or take longer for the cell to produce the mature protein than it does for GFP.

In summary, we again showed that vectors containing ERSE upstream of an SV40 promoter can be transactivated by ATF6(50) co-expression and that increasing numbers of ERSE result in increasing expression of a reporter protein. We have now shown that this can be done for both SEAP and GFP expression and this fulfils the third aim/objective set out at the beginning of this chapter.

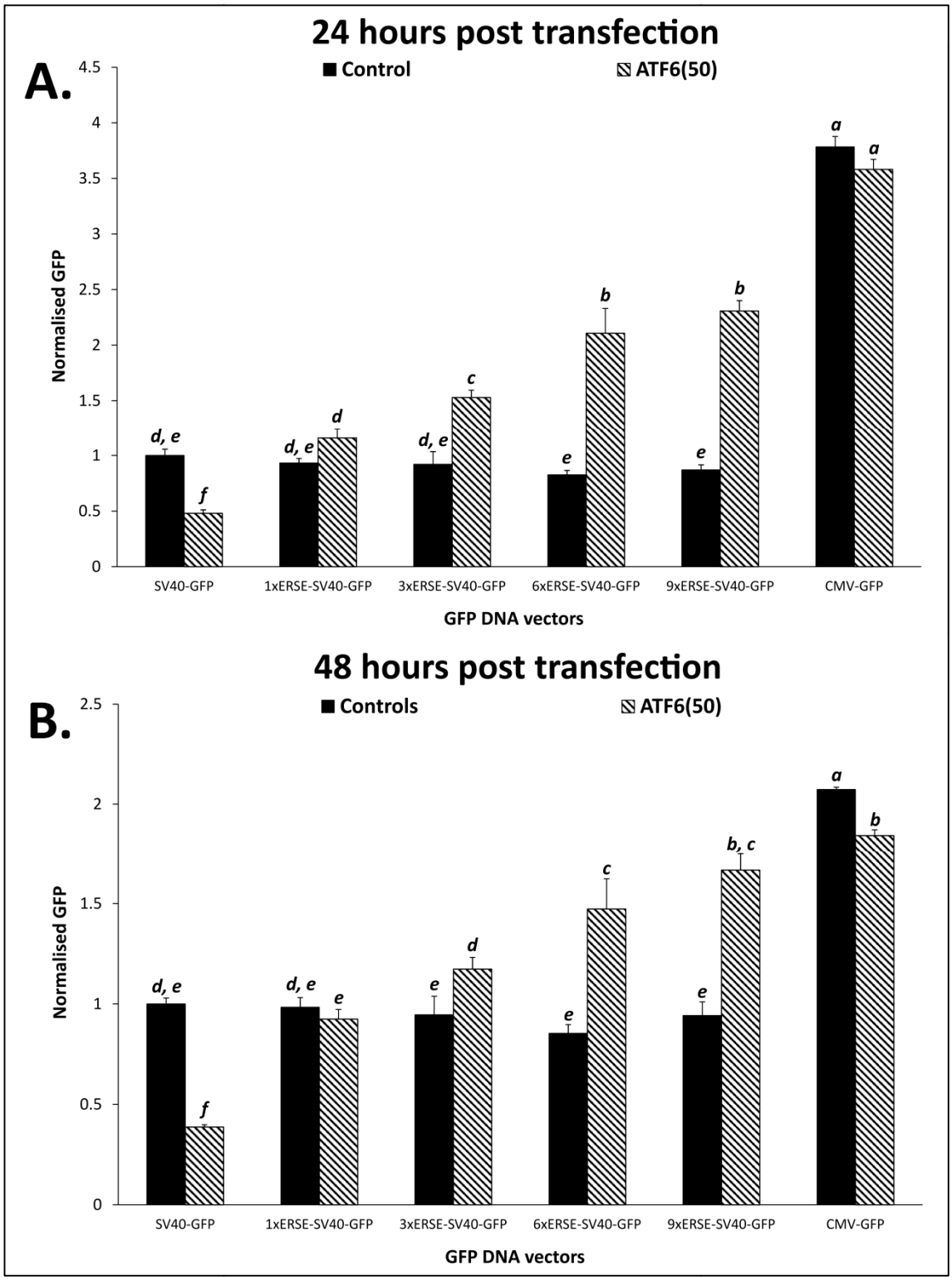

Figure 3.6 – The effect of ATF6(50) co-transfection and differing numbers of ERSE on the expression GFP. 0.5 µg of 6 different GFP expression vectors (SV40-GFP, 1xERSE-SV40-GFP, 3xERSE-SV40-GFP, 6xERSE-SV40-GFP, 9xERSE-SV40-GFP and CMV-GFP) were transfected separately in CHOK1SV cells under two different conditions. Controls contained the GFP vector and the –ve control DNA vector to normalise total DNA load. The GFP vectors were also co-transfected with 0.5 µg of ATF6(50). **A.** GFP expression 24 hours post-transfection. **B.** GFP expression 48 hours post-transfection. All GFP readings were normalised to the expression from the SV40-GFP vector under control conditions at the corresponding time point. For each GFP vector, values with different letters differ significantly from each other (Tukey's test, $p < 0.05$). N = 3, Error bars represent + 1 S.D.

Table 3.4 – Fold change in GFP expression when compared to the lowest GFP expression achieved for both 24 and 48 hours post-transfection

24 hours post-transfection		48 hours post-transfection	
GFP vector	Fold change	GFP vector	Fold change
SV40-GFP + ATF6(50)	1.00[a]	SV40-GFP + ATF6(50)	1.00[a]
SV40-GFP (Control)	2.08[b]	1xERSE-SV40-GFP + ATF6(50)	2.39[b]
1xERSE-SV40-GFP + ATF6(50)	2.41[b]	SV40-GFP (Control)	2.58[b, c]
3xERSE-SV40-GFP + ATF6(50)	3.19[c]	3xERSE-SV40-GFP + ATF6(50)	3.03[c]
6xERSE-SV40-GFP + ATF6(50)	4.39[d]	6xERSE-SV40-GFP + ATF6(50)	3.80[d]
9xERSE-SV40-GFP + ATF6(50)	4.80[d]	9xERSE-SV40-GFP + ATF6(50)	4.31[d]

The numbers in the table represent the fold change in GFP expression when compared to the SV40-GFP vector co-transfected with the ATF6(50) vector. Calculated using the data in figure 3.6. For each SEAP vector, values with different letters differ significantly from each other (Tukey's test, $p < 0.05$).

3.5.6 – Simultaneous and controllable expression of two proteins, SEAP and GFP, using ERSE and co-expression of ATF6(50)

To see if ERSE upstream of an SV40 promoter and co-expression of ATF6(50) can be used to control the levels of expression of two genes simultaneously, different combinations of the SEAP and GFP vectors were co-transfected together with ATF6(50). This involved nine different combinations of controlled SEAP and GFP expression. This included:-

- Low : low, low : medium, low : high, medium : low, medium : medium, medium : high, high : low, high : medium and high : high expression of both GFP and SEAP expression, respectively

From the work already presented in this chapter the following SEAP and GFP vectors were chosen:-

- SV40-SEAP & SV40-GFP – these were chosen because co-expression with ATF6(50), and the resulting negative effects previously seen, will give lower levels of expression of both of these reporter proteins

- 1xERSE-SV40-SEAP & 3xERSE-SV40-GFP – we expected that these vectors would give medium levels of expression for each reporter protein when co-expressed with ATF6(50). Vectors containing different numbers of ERSE were chosen for expression of SEAP and GFP because the differences between the numbers of ERSE present effect SEAP and GFP expression slightly differently

- 9xERSE-SV40-SEAP & 9xERSE-SV40-GFP – these two expression vectors gave the highest levels of SEAP and GFP expression when co-transfected with ATF6(50)

Initially every combination of these six SEAP and GFP vectors were co-transfected together (9 different combinations). 1 μg of SEAP and 1 μg of GFP was co-transfected along with three different amounts of ATF6(50) DNA (0.5, 0.75 and 1.0 μg). In total, 27 different combinations of SEAP and GFP vectors and ATF6(50) DNA amounts were tested. 48 hours post transfection the results were collected and analysed (data not shown). From this data, we could see that the ranges of expression of both SEAP and GFP were very different. SEAP expression from lowest to highest had a 3.77-fold range

while GFP expression had a 26.50-fold range. The highest levels of SEAP and GFP expression were similar to those seen earlier in this chapter. The results of both SEAP and GFP were compared together for every combination and combinations of SEAP and GFP vectors and the amount of ATF6(50) DNA were selected (see table 3.5) according to levels of expression we hoped to achieve and previously stated in this section. These combinations were then transfected again. Control transfections without ATF6(50) were also carried out for every combination of SEAP and GFP vector to show that both ATF6(50) was having an effect on the expression of both reporter proteins and show any effects these combinations of SEAP and GFP vectors had on each other's expression in the absence of ATF6(50).

Figure 3.7 – A. shows the results of the controls. We can see that SEAP expression from all the different combinations of SEAP and GFP vectors was roughly the same and it was also similar to SEAP expression shown previously in this chapter when the SV40-SEAP and ERSE SEAP vectors were transfected alone and without ATF6(50) DNA. The presence or absence of ERSEs and the presence of any of GFP vectors did not seem to effect SEAP expression, whereas for GFP there were clear differences in expression. There were both differences in expression between the three different GFP vectors and also differences for the same GFP vector depending on the co-transfected SEAP vector. Considering the average amount of GFP expression for each different GFP vector the SV40-GFP vector performed the worst under control conditions followed by the 3xERSE-SV40-SEAP vector while the 9xERSE-SV40-GFP vector produced the highest GFP expression. However, for each GFP vector the co-expressed SEAP vector also had an effect on GFP expression. When they were co-expressed with 9xERSE-SV40-SEAP, GFP expression was the lowest for each individual GFP vector. For both SV40-GFP and 3xERSE-SV40-GFP, GFP expression was highest when co-expressed with SV40-SEAP and slightly lower when co-expressed with 1xERSE-SV40-SEAP. The results for 9xERSE-SV40-GFP were different with co-expression of SV40-SEAP and 1xERSE-SV40-SEAP giving similar results, but this was still significantly higher compared to GFP expression when 9xERSE-SV40-SEAP was co-transfected and the highest GFP expression of all the GFP vectors. Only when the 9xERSE-SV40-GFP vector was co-transfected with SV40-SEAP and 1xERSE-SV40-SEAP was GFP expression at a similar level to that seen earlier

in this chapter when not co-expressed with a SEAP vector or ATF6(50). Both the SV40-GFP and 3xERSE-SV40-GFP vectors underperformed compared to early results.

These control results with the three different GFP vectors were unexpected and this was because from figure 3.6 we saw that in the absence of ATF6(50) there was very little difference in GFP expression between the SV40-GFP vector and the ERSE GFP vectors, which we had also seen with the SV40-SEAP and ERSE SEAP vectors (figure 3.5). We did not expect the presence of a SEAP vector to effect GFP expression so significantly, due to all vectors containing an SV40 promoter and the fact that the presence of ERSE alone did not seem to effect GFP expression previously. This shows that the ERSE present in the GFP vectors must have some effect when there is a co-transfected SEAP vector and even in the absence of ATF6(50) expression. The differences seen between the three GFP vectors and also the effects of the different SEAP vectors might be able to be explained by the number of ERSE present or their absence having an effect on expression from the GFP vectors when a SEAP vector is co-expressed. Having 9xERSE in the GFP vector did seem to benefit GFP expression when co-expressed with a SEAP vector and even 3xERSE was better for GFP expression than having none in this instance. GFP expression was also greater when the competing SEAP vector had less or a similar number of ERSE. Exactly why this is the case is unknown and why this only occurs for the GFP vectors and not the SEAP vectors is even harder to answer. Is it the differences between SEAP and GFP protein in terms of structure, processing, half-life and their final destinations that cause this major difference? Is SEAP selectively produced over GFP? Does the simultaneous expression of SEAP and GFP induce a UPR? If so, why does it not affect SEAP expression the same as GFP expression? Whatever the reason for this it is beyond the scope of this chapter and this thesis but is interesting none the less.

Table 3.5 shows the different combinations of GFP and SEAP vectors and also the amount of co-transfected ATF6(50) DNA used for the results in figure 3.7 – B.. From figure 3.7 – B. we can see the different levels of simultaneous GFP and SEAP expression. For GFP expression we can see that using either the absence or presence of ERSEs in these vectors and co-transfection of ATF6(50) we achieved three distinct levels of expression, namely low, medium and high expression. For all three levels of

expression the results within each group (low, medium and high) did not differ significantly while between groups they were significantly different. There were clear distinct differences in the GFP expression achieved between the low, medium and high expression groups. The highest levels of GFP expression were similar to those seen before for the 9xERSE-SV40-GFP vector when co-expressed with ATF6(50) but without a SEAP vector.

For SEAP expression we also saw differences between the low, medium and high expression groups (figure 3.7 – B.). The highest levels of SEAP expression were slightly higher than previous experiments when ERSE SEAP vectors were co-transfected with ATF6(50) and no GFP DNA was present. Within the low group, all SEAP expression was shown to be similar but significantly different from the medium and high groups. There were less clear differences between medium and high SEAP expression groups. Within each of these groups, there were no significant differences in SEAP expression, which is what we were aiming to achieve. However, between groups for some pairs of GFP and SEAP vectors, there was no significant difference between some medium and high SEAP expression conditions, which is not what we expected and we had aimed to achieve results similar to that seen for GFP expression. Ideally, SEAP expression for the medium group would be lower or the high group higher then it was or indeed both.

This lack of difference between high and medium SEAP expression conditions may be due to the range of SEAP that we managed to achieve being much smaller than the range of GFP expression. As mentioned earlier, the range in SEAP and GFP expression in the preliminary work for this section was 3.77-fold and 26.50-fold respectively. The ranges of expression from lowest to highest in figure 3.7 – B. were similar but slightly lower, 3.22-fold for SEAP and 24.21-fold for GFP expression. This means there was a much bigger window to work with for GFP expression then we did for SEAP. Also from the control results (figure 3.7 – A.), we saw how just the presence of different numbers of ERSE effected GFP expression when co-expressed with a SEAP vector which was not the case for SEAP expression. Both these differences would point to GFP expression being easier to affect and manipulate than SEAP expression. Work from earlier in this chapter also shows this.

Although the differences between the medium and high SEAP expression was not as great as we had aimed to achieve we have still shown how we can use different numbers of ERSE and different amounts of ATF6(50) to manipulate expression in a controllable fashion for two reporter proteins simultaneously. We set out to achieve nine different combinations of simultaneous SEAP and GFP expression and this was very nearly achieved.

In summary, ERSEs and ATF6(50) can be used to help control the level of two reporter proteins simultaneously, although GFP expression was more successfully manipulated then SEAP expression. This goes some way to fulfilling the final aim/objective we set out at the beginning of this chapter, but more work is needed to make this a more useful system for controlled mammalian expression.

Table 3.5 – The combinations of co-transfected GFP and SEAP vectors and the amounts of co-expressed ATF6(50) used to achieve the nine different levels of simultaneous and controlled GFP and SEAP expression in figure 3.7 – B.

Level of expression	GFP vector	SEAP vector	ATF6(50)
Low GFP : Low SEAP	SV40-GFP	SV40-SEAP	1.0 µg
Low GFP : Medium SEAP	SV40-GFP	1xERSE-SV40-SEAP	0.5 µg
Low GFP : High SEAP	SV40-GFP	9xERSE-SV40-SEAP	1.0 µg
Medium GFP : Low SEAP	3xERSE-SV40-GFP	SV40-SEAP	0.75 µg
Medium GFP : Medium SEAP	3xERSE-SV40-GFP	1xERSE-SV40-SEAP	0.75 µg
Medium GFP : High SEAP	3xERSE-SV40-GFP	9xERSE-SV40-SEAP	1.0 µg
High GFP : Low SEAP	9xERSE-SV40-GFP	SV40-SEAP	1.0 µg
High GFP : Medium SEAP	9xERSE-SV40-GFP	1xERSE-SV40-SEAP	1.0 µg
High GFP : High SEAP	9xERSE-SV40-GFP	1xERSE-SV40-SEAP	0.5 µg

The table shows the different combinations of co-transfected GFP and SEAP vectors (1.0 µg of each GFP and SEAP DNA were co-transfected) and the amount of co-expressed ATF6(50) DNA used in figure 3.7 – B. to achieve the nine different combinations of simultaneous GFP and SEAP expression.

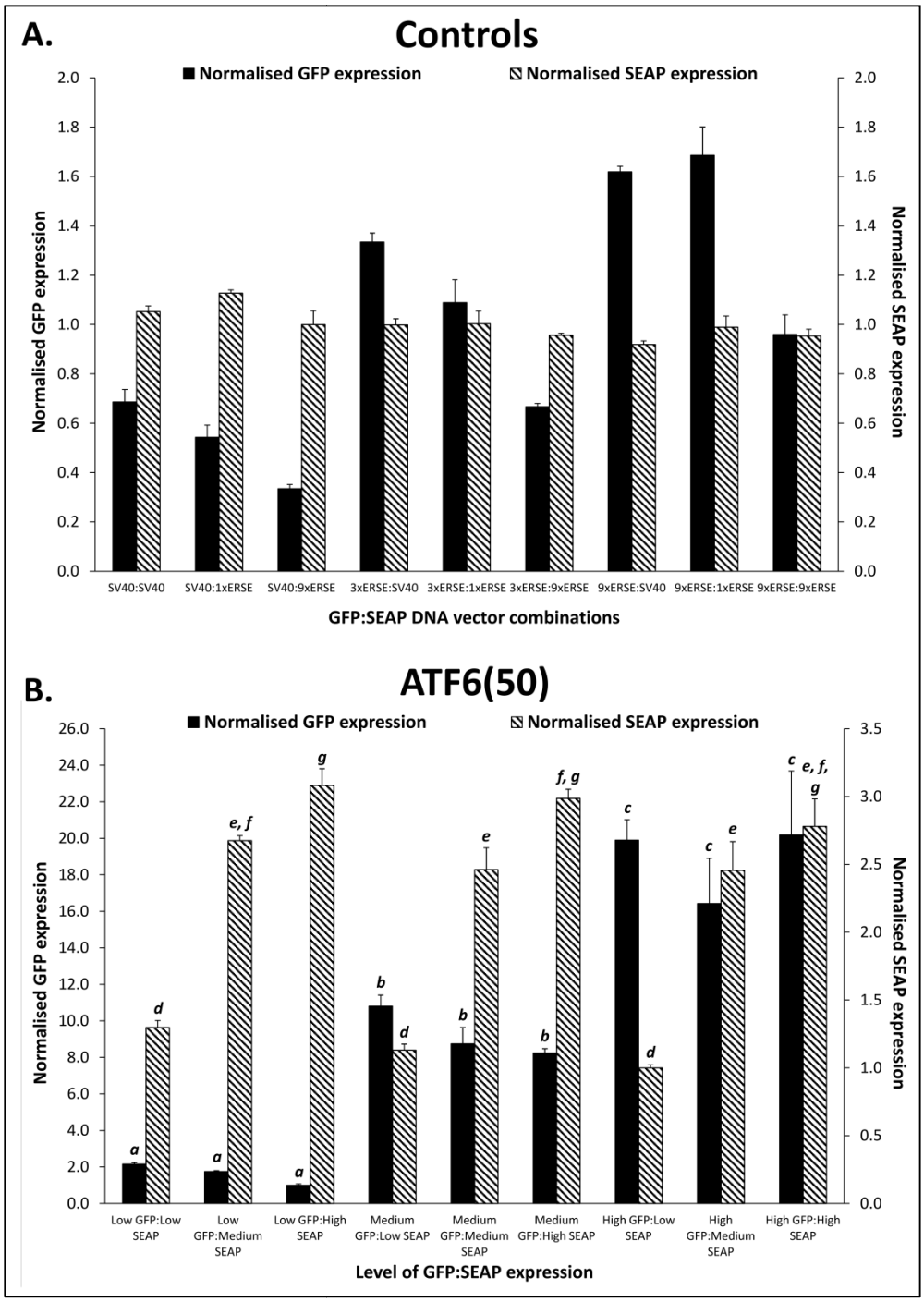

Figure 3.7 – Controllable and simultaneous expression of SEAP and GFP. 1.0 µg of SEAP vector and 1.0 µg of GFP vector DNA were co-transfected in the absence and presence of ATF6(50) DNA in CHOK1SV cells. Table 3.5 shows the different combinations of SEAP and GFP vectors used in each co-transfected pair as well as the amount of co-expressed ATF6(50). SEAP and GFP measurements were taken 48 hours post-transfection. **A.** Shows the results of the controls were no ATF6(50) DNA was present and transfection load was normalised using –ve control DNA. Both SEAP and GFP results were normalised separately to the average of all nine conditions for each reporter protein. The *y-axis* denotes the GFP and SEAP vectors used for each pair. **B.** Shows the results when ATF6(50) was co-transfected. Different amounts of ATF6(50) (see table 3.5) were used but total DNA load was kept constant using the –ve control DNA. SEAP and GFP expression were normalised separately to the lowest results seen for each reporter protein to show the range of expression. The *y-axis* denotes the level of expected GFP and SEAP expression for each pair. For SEAP and GFP results, values with different letters differ significantly from each other (Tukey's test, $p < 0.05$) (Differences between SEAP and GFP were not compared). N = 3, Error bars represent + 1 S.D.

3.6 – Further Discussion, Final conclusions and Future Directions

3.6.1 – Further discussion

At the beginning of this chapter we set out to produce a system where we used the number of ERSE and expression of its binding factor, ATF6(50), to control the expression of initially a single reporter protein and then two reporter proteins simultaneously. We showed how co-expression of ATF6(50) could transactivate both SEAP and GFP vectors which had ERSE sequences upstream of an SV40 promoter and increase expression above that of a control (no ATF6(50)). Due to the ATF6(50) vector containing the strong CMV promoter it had negative effects on expression from both SEAP and GFP vectors lacking ERSE and just containing the weaker SV40 promoter. This was thought to occur due to promoter interference which will be studied later in this thesis (chapter 4). Although a negative effect on expression, we took advantage of this and used it to increase the range of expression for both SEAP and GFP. The amount of ATF6(50) had an effect on either the increase or decrease in expression. We also showed how the number of ERSE had an effect on expression when ATF6(50) was present. Expression increased when the numbers of ERSE was increased although this was shown to work better when expressing GFP. When expressing SEAP and GFP ERSE vectors separately, the number of ERSE sequences had no effect when ATF6(50) was not co-transfected. At the end of the chapter we also showed how we could use ERSE and ATF6(50) to control expression of both SEAP and GFP simultaneously. This was for only nine different combinations of SEAP and GFP and the results for different levels of GFP expression were more distinct and significant. The number of different combinations of SEAP and GFP will need to be increased in the future as we develop this system further for tunable/controlled expression of one or more genes.

This was our first attempt at constructing a system for the tunable/controlled expression of genes in mammalian cells. However, this system still needs further development. The results for different levels of GFP expression were significantly different in the final results section (section 3.5.6 & figure 3.7 – B.) but the results for SEAP were not as significantly different as initially expected. The aim was for the high and medium levels of SEAP expression to be more distinct from each other. This may

have occurred because the range of SEAP expression was much lower than for GFP (3.22-fold and 24.21-fold respectively). Other experiments had also showed that SEAP expression was less amenable to the effect of different numbers of ERSE then GFP (figure 3.6 & 3.7) when ATF6(50) was co-expressed even though increases in expression were observed. It would seem that GFP expression was easier to manipulate then SEAP. It cannot be said with any certainty why there were such differences in controlling SEAP and GFP expression, although this was briefly discussed in a previous results section, but we will discuss it further in the next paragraph.

SEAP and GFP expression in terms of range and ease to control differed when expressed either separately or simultaneously when co-expressed with ATF6(50). Another major difference between them was their expression when the ERSE SEAP and SV40-SEAP vectors were co-transfected with ERSE GFP vectors and SV40-GFP in the absence of ATF6(50) (figure 3.7 – B.). We saw that all the SEAP vectors used in this experiment gave similar levels of SEAP expression regardless of the co-transfected GFP vector. However this was not the case for GFP, were both the number and absence of ERSE as well as the co-transfected SEAP vectors did have an effect. We already stated that SEAP and GFP are very different kinds of reporter proteins, but it was still an unexpected result. The presence of ERSE alone when GFP vectors were transfected by themselves showed them not to have an effect, so why did it seem they did once they were co-expressed with a SEAP vector but still in the absence of ATF6(50)? Surely this means there are very significant differences in the way both SEAP and GFP are produced in our cells.

One explanation could be that the dual expression of SEAP and GFP actually induces a UPR. If this is the case, then it would be expected that both SEAP and GFP vectors containing ERSE might be capable of responding to endogenous UPR activators such as endogenous XBP1s and more prominently ATF6(50). This is turn might result in increased transcription from vectors containing ERSE and therefore more mRNA available for translation. If this were true then it might be expected that both SEAP and GFP vectors containing ERSE should respond with increased SEAP and GFP protein expression. However, this might be where the differences between the proteins may affect their expression when in this context. SEAP is a secreted glyco-protein and

could be said to be a more complex protein to produce than GFP given the extra steps involving glycosylation and secretion. We have already shown that SEAP protein does not correlate at a one-to-one ratio with SEAP mRNA (figure 3.6). Increasing the amount of SEAP mRNA did not proportionally increase SEAP protein and this may be because the cells cannot handle or completely process this increase in SEAP mRNA. It could also be due to the secretary machinery of the cell being unable to process further increases in SEAP protein. This is known as 'secretion saturation' and it has been shown to occur in CHO cells previously (Schroder *et al.*, 1999; Schroder & Friedl, 1997) and in microbial systems as well (Smith & Robinson, 2002; Parekh *et al.*, 1995). This may be different for GFP, since it is not secreted extracellularly, the cells may be better able to handle increases in GFP mRNA and so increase GFP protein. This would mean that the correlation of GFP mRNA to GFP protein is closer to a one-to-one ratio and that even a small increase in GFP mRNA can be translated and reflected in the amount of protein expressed. This is just a hypothesis and further work such as measuring the amount of GFP mRNA present would need to be done to start to have a clearer understanding of the differences seen between SEAP and GFP protein expression. Although this does not completely answer why there were differences seen between the two proteins, and there are also other possible explanations, the fact that the work with SEAP seems less successful, SEAP is a secreted protein and future directions of the system being controlled expression of multiple genes for cell engineering it is likely that SEAP is not the right reporter protein to use to further develop this system. Another intracellularly expressed protein might be a better candidate.

In hindsight, SEAP is not the correct reporter protein to continue using for the future development of this system and alternatives will be mentioned later in this chapter along with how it can be improved and expanded into a more viable tunable/controllable system for the expression of one or multiple genes in mammalian cells. However, we have still showed it is possible using recombinant ATF6(50) expression and differing numbers ERSE upstream of an SV40 promoter driving expression of a reporter protein that we can control the level of expression and that this can also be done with two reporter proteins simultaneously.

3.6.2 – Final conclusions

The insertion of ERSE upstream of an SV40 promoter without co-transfection of ATF6(50) had no effect on either SEAP or GFP expression when these were expressed by themselves and compared to a vector just containing an SV40 promoter and no ERSE.

Vectors containing ERSE were transactivated and expression was increased by co-transfection of ATF6(50). Expression from vectors containing only the SV40 promoter was reduced by ATF6(50) expression. The amount of ATF6(50) DNA transfected had an effect, increased ATF6(50) DNA caused increased expression from ERSE vectors and decreased expression from SV40 only vectors. These positive and negative effects can both be used to increase the range of expression for our proposed system.

The number of ERSE also had an effect on the levels of expression when ATF6(50) was co-expressed. Increasing numbers of ERSE gave increasing levels of expression. This was shown better for GFP expression then it was for SEAP expression. The number of ERSE had no effect when ATF6(50) was not co-transfected and when SEAP and GFP vectors were transfected alone.

The same system of differing numbers of ERSE and ATF6(50) can used for manipulating the levels of both a secreted glycoprotein, SEAP, and an intracellular protein, GFP. Our system seemed to perform better for GFP compared to SEAP expression.

This same system for controlled expression can be used to control expression of SEAP and GFP simultaneously. However, GFP again performed better having a greater range of expression and more distinction between the three levels of expression we set out to attain which we simply termed low, medium and high.

These are the first steps towards developing a mammalian tunable/controlled expression system for one or multiple genes and although we achieved the majority of the goals we set out to achieve there is still much room for improvement, expansion and refinement of our system to be a valuable tool for not only the biopharmaceutical industry but many areas of biological research and biotechnology. This will be discussed in the next section.

3.6.3 – Future directions

Although this chapter has shown what can be achieved using a activating transcription factor and its binding sequence to control expression of either a single reporter protein or two reporter proteins simultaneously further work needs to be done to develop our system into a viable one for tunable/controlled expression.

Firstly, we should consider applications for this kind of system. Improved biopharmaceutical production is a vast research area studied by both academic and industrial researchers. Within this research area there have been a vast number of strategies that have been studied to help increase or lower the cost of production from bioreactor design, media formulation, cell line selection, and improved DNA production vectors etc. Cell line engineering has also been attempted to improve the productivity of CHO cells and other cell lines important in this area. Most basic strategies involve the over-expression of a single protein and although some have been successful many more have failed. Although the ideas behind these strategies may be rational, the complexity of biological systems means that manipulating the expression of a single protein is unlikely to have a dramatic effect. There will be a need in the future for strategies were more than one protein is expressed, two/three or probably even greater numbers. It is also likely that the different proteins will need to be expressed at different levels which are quite difficult to achieve with completely constitutive promoters. This is what we hope to achieve with further development of this system. We want to be able to discretely control or tune the expression of multiple proteins simultaneously in mammalian cells. Firstly this would be done in CHO cells and used to improve performance of an industrially important cell line and most likely for a so called 'difficult to express' biopharmaceutical. This system can also be classed as mammalian synthetic biology and we expect that it could be used for applications outside of the biopharmaceutical area. For this to happen, though, we need to further develop, expand and improve our system.

The work in this chapter used the reporter proteins, SEAP and GFP, although we did show we can manipulate the expression of both simultaneously, in hindsight and also thinking about the future development of our system this is not the ideal combination to show the full potential of this system. The limitations showed when manipulating

SEAP expression, mostly the range of its expression, and the fact it is a secreted glycoprotein means it is unlikely to reflect the type of protein that this system will be used for controlling expression. GFP being an intracellular protein is more suited and future work should initially involve the expression of other fluorescent proteins (RFP and YFP) and other intracellular reporter proteins. Using both RFP and YFP along with GFP would make the initial further development of the system easier, in terms of being able to easily measure their expression, and also expanding the number of controlled proteins. Eventually proteins that actually have more of a direct effect on cells would need to be tested, but this would be a good starting point for a proof of concept.

To further expand this system the number of ERSE placed upstream of an SV40 promoter should also be increased past nine copies until we reach a point where additional copies no longer have an additive effect. Further titrations of ATF6(50) with both GFP and other fluorescent proteins also needs to be carried out to better understand the effects when different numbers of ERSE are present in these vectors.

Optimizing this system for the expression of multiple proteins may be benefitted by using a DOE approach. As long as we fully understand the limits or boundaries of the design space correctly and the interactions and the expressions of multiple proteins can be modelled sufficiently using this approach, this will reduce the number of experiments required to show our system works along with giving some predictive power. In turn this will reduce the timeframe of development. However DOE is not simple and we would have to make sure we approached it in the right manner.

Further development might also include the use of a synthetic transcription factor to replace ATF6(50) which does not bind endogenous transcription factor binding sites but a novel binding site instead of the ERSE. This would add orthogonality and remove any off system effects that may currently occur with ATF6(50) which are presently unknown. However this is something that would not be done initially and it may even be the case that over-expression of ATF6(50) might not affect the cell too greatly when there are vectors containing ERSE for it to bind, acting as transcription factor sponges, and thus reducing the available ATF6(50) to affect endogenous targets.

If the vision of our system is achieved it would be a novel method of controlling the expression of multiple genes in mammalian cells in cell engineering strategies and would be an extremely useful tool for not only the biopharmaceutical industry but also for biotechnology and biological research in general.

Chapter 4 – A Study of Promoter Interference between Three Commonly used Constitutive Promoters for Biopharmaceutical Production in CHO Cells

4.1 – Abstract

The SV40 early, human CMV major immediate-early and the human EF1α promoters are all commonly used in the production and research of therapeutic proteins. However, when more than one promoter is expressed in a system as will occur in many cell engineering strategies, there is competition between promoters which can have negative effects on the expression from the promoters involved. This phenomenon can be termed promoter interference and this can occur between even physically unlinked promoters.

The expression of these three promoters in CHO-S cells along with their abilities to negatively affect the expression of another competing promoter was compared. SEAP was used as a model glycoprotein and GFP in place of an active intracellular effector protein. Vectors containing CMV and EF1α promoters produced similar and higher levels of SEAP expression compared to an SV40 containing vector. Co-expression of SEAP and GFP vectors utilizing different promoters showed that a SV40 containing SEAP vector was most negatively affected by the presence of a competing GFP vector. This depended on the competing promoter and a SV40 had the least, effect while the CMV had the greatest negative effect closely followed by the EF1α promoter. Expression from a CMV-SEAP vector was the least affected by a competing promoter. The level of interference of a competing promoter seemed to be related to its transcriptional strength since the CMV promoter was the most transcriptionally active it was also the most interfering followed by the EF1α promoter with SV40 promoter being the least.

These promoters are poorly suited for use in cell engineering strategies were multiple genes are needed to be expressed and that the development of independent promoters which do not overly interfere with each other's expression will be needed in the future. For the development of such promoters a more in depth understanding of why even physically unlinked promoters either interfere or compete for expression is needed.

4.2 – Introduction

In chapter three a phenomenon termed “promoter interference” was discussed briefly due to its effects on the expression of co-expressed proteins. From this work it could be seen that when an expression vector using a weaker promoter such as the SV40 was co-transfected with another containing the stronger CMV promoter, the amount of expression from the latter was reduced (figure 3.3 & 3.4). Although this effect is not a new discovery and rather just an accepted consequence when mixing more than one promoter in a system there are only a limited number of published studies on it, especially in terms of the context we encountered it.

There is great interest in biopharmaceutical research in understanding CHO cells and the processes surrounding recombinant therapeutic protein production in mammalian cells. A large proportion of the work involves the expression of recombinant proteins which might include therapeutic proteins, reporter proteins (SEAP, luciferase and GFP), and effector proteins for potential cell engineering strategies (e.g. – over-expression of UPR proteins and molecular chaperones (Pybus *et al.*, 2013)). Sometimes this will involve the co-expression of a single protein, but when a new cell engineering strategy is being tested then two or more recombinant proteins maybe over-expressed simultaneously (e.g. – reporter and effector protein). Initially this will likely involve the over-expression of a reporter protein such as SEAP, which is often used in this area of academic research due it being an easily expressed model glycoprotein and the ease and low cost of measuring its expression compared to therapeutic glycoproteins. An effector protein will also be co-expressed with the reporter protein. This would likely be done first in a transient manner due to the time consuming nature of creating stable cell lines. The effect of the cell engineering strategy on the expression of the reporter protein will then be measured after a given amount of time post-transfection. Depending on the effects of the effector protein further studies may be carried out to understand and develop the approach further. Sometimes however, some strategies do not work or are unviable and this can happen for many reasons. This could include incorrect expression of the recombinant proteins (meaning either too high/low or not optimal expression) and negligible, non-target or negative effects of recombinant proteins on the cell expression system.

Another point to consider is that when co-expressing two proteins this will probably involve, at least initially, the co-transfection of two DNA vectors. As well as the two recombinant proteins competing at the translational level there will also be prior competition at the transcriptional level. If two strong DNA promoters are used, this may not become a problem if high level expression of each protein is acceptable. However, if different strength promoters are used because there is a need for different levels of expression, negative interactions may occur between them resulting in the reduction in expression of either one of them or both. This may make it very difficult to get the optimum levels of expression required for certain cell engineering strategies to work. Changing the ratios of transfected DNA may help overcome this problem but some promoters such as the CMV have been shown to negatively affect expression from other promoters even when present in small amounts (Huliak *et al.*, 2012).

In chapter four it was shown that the strong CMV promoter reduced expression from a DNA vector containing the weaker SV40 promoter. This was thought to occur through 'promoter interference'. The research on this phenomenon has studied it in a few different scenarios but never from a cell engineering for biopharmaceutical production perspective. In this chapter we aim to investigate this phenomenon from this perspective by comparing the expression of three promoters commonly used in biopharmaceutical research and production (SV40, CMV and EF1α) and how there expression changes when two are present in the same system. This will be done by co-transfecting SEAP vectors with GFP vectors. An explanation of why these reporter proteins were chosen is given in the aims and objectives of this chapter (section 4.3).

The next two sub sections give a brief description of what promoter interference is and the origins, structure and uses of the SV40, CMV and EF1α promoters.

4.2.1 – Brief Background – 'Promoter interference'

Promoter interference is basically defined as the "perturbation of one transcriptional unit by another" (Huliak *et al.*, 2012; Curtin *et al.*, 2008; Eszterhas *et al.*, 2002). Promoter interference can also be called transcriptional interference or promoter competition (Curtin *et al.*, 2008; Shearwin *et al.*, 2005; Conte *et al.*, 2002; Eszterhas *et al.*, 2002; Hirschman *et al.*, 1988; Cullen *et al.*, 1984). The different names used to describe this perturbation of one transcriptional unit by another can reflect the context or the area of the research. The definition at the beginning of this paragraph and repeated in the previous sentence is very broad because there is more than one way to perturb transcription from a promoter and this can occur in many different scenarios. This could occur in the context where a DNA enhancer could activate multiple genes but only selectively activates the promoter of a single gene (Conte *et al.*, 2002; Choi & Engel, 1988; Hirschman *et al.*, 1988). Other examples involve either the blocking of the binding of RNA polymerase (RNAP) or its progression along DNA and this can occur when promoters are in close proximity or the sequences of genes overlap (Shearwin *et al.*, 2005; Callen *et al.*, 2004; Prescott & Proudfoot, 2002; Wang *et al.*, 1998; Adhya & Gottesman, 1982). It could also occur through a process called transcriptional squelching, where there is competition for limited amount of a transcriptional regulator/s (Huliak *et al.*, 2012; Cahill *et al.*, 1994a; Prywes & Zhu, 1992). These occurrences can occur naturally in chromosomes and in integrated and non-integrated transgene DNA expression vectors (Curtin *et al.*, 2008; Shearwin *et al.*, 2005; Conte *et al.*, 2002; Eszterhas *et al.*, 2002; Choi & Engel, 1988; Hirschman *et al.*, 1988; Cullen *et al.*, 1984). Although the term 'interference' may make this phenomenon sound somewhat of a negative and unwanted effect, and in some cases it such as certain examples of controlled expression of multiple transgenes (Curtin *et al.*, 2008; Thompson & Myatt, 1997; Ingelbrecht *et al.*, 1991), it can actually be used for gene regulation (Shearwin *et al.*, 2005; Callen *et al.*, 2004; Nasser *et al.*, 2002; Wang *et al.*, 1998).

The overwhelming majority of the research published on promoter interference describes cases when this occurs for promoters or transcriptional regulatory sequences which are *in cis*, meaning that they are either on the same chromosome or the same

DNA expression vector. Very little work has been carried out on promoter interference on promoters which are physically unlinked or *in trans*. In chapter 4 we encountered promoter interference and this occurred *in trans* when two different expression vectors were co-transfected.

The only research carried out on physically unlinked promoters, which is in the same context, as the interference we encountered was conducted by Huliak *et al.*, (2012). They were interested in the interpretation of transiently expressed reporter gene experiments that are frequently used to analyse promoters and transcriptional regulators. Most of these experiments involve the expression of a reporter protein, an internal control (e.g. - GFP) and/or *trans*-acting factor such as transcriptional activator/repressor. Generally it is thought that the expression of the reporter protein is a good relative indicator of the strength of the transcriptional units involved in the assay. In many cases it is, but there is known inherent limitations with this kind of experiment and situations when the reporter protein does not accurately reflect the strength of a promoter/regulator. This could occur when the reporter protein itself is toxic to the cell or when the internal control and/or *trans*-acting factor interfere with reporter protein activity (Huliak *et al.*, 2012; Liu *et al.*, 1999). Alternatively, negative interactions may occur between two physically unlinked promoters and could be another reason why reporter protein expression might not reflect the true transcriptional capacity of given regulatory element, or in other words promoter interference. Huliak *et al*., (2012) showed that when the often utilized CMV promoter was present, even in only small amounts, it had a severe negative impact on expression from other promoters. This negative impact was seen for both basal and activated promoter activities and even when a CMV promoter was present in a vector lacking a recombinant gene to express it still had a negative effect. The take home message was that because the CMV is so commonly utilised in many vectors that are used in transient reporter gene assays that its impact on promoters when co-expressed should be taken into account when interpreting the results. A hypothetical example would be when a GFP vector with a CMV promoter is used as an internal control in such an assay. A newly designed synthetic promoter which is driving expression of a reporter protein may seem according to reporter protein activity to be

performing disappointingly poorly. However, if the internal control is removed from the experiment the promoter performs better due to the removal of the negative interference of the CMV promoter.

In the study Huliak *et al.*, (2012) co-expressed the CMV promoter with the human immunodeficiency virus (HIV), human T-cell leukaemia virus type-I (HTLV-T), NF-κB-responsive and p53-responsive promoters. However, apart from the CMV none of these promoters are routinely used in vectors for recombinant therapeutic protein production. The work in this chapter will investigate the negative interactions of three promoters (SV40, CMV and EF1α) which are routinely used in biopharmaceutical production. The next section will give a brief description of the three promoters studied in this chapter.

4.2.2 – Brief Background – SV40, CMV and EF1α promoter

4.2.2.1 – SV40 early promoter

The SV40 (Simian virus 40) early promoter originally came from the polyomavirus, SV40, which was initially found in Rhesus monkey cells (Eddy *et al.*, 1961; Sweet & Hilleman, 1960). It is responsible for the transcription of so called early genes which are expressed soon after productive viral infection and before viral genome replication (Wildeman, 1988). Research on the promoter has been very important in understanding eukaryotic transcription (Cooper, 2000; Banerji *et al.*, 1981; Benoist & Chambon, 1981)

The promoter is composed of a TATA box with three 21 bp repeats (GC boxes) and two 72 bp repeats (also known as the SV40 enhancer) upstream (Byrne *et al.*, 1983). However, it has been shown that although the enhancer is essential for good transcription from the promoter its position relative to it is not important and can still enhance transcription when placed downstream of the promoter and gene CDS (Cooper, 2000; Kadesch & Berg, 1986; Berg *et al.*, 1984).

The SV40 promoter/enhancer has been used for recombinant gene expression in many different areas of biological research. It can function in many different cell lines, gives

constitutive expression and although not overall the strongest promoter in terms of transcription it still gives relatively good expression. In the biopharmaceutical industry the SV40 promoter is not commonly used to drive expression of a therapeutic protein because stronger/better alternatives are available, but it is still used for expression of selection genes for the creation of stable cell lines. For example the GS (glutamate synthetase) expression system (Bebbington *et al.*, 1992) (Lonza Biologics, Slough, UK) and the MTX (methotrexate) DHFR (dihydrofolate reductase) expression system (Kaufman & Sharp, 1982) use the SV40 promoter.

The SV40 promoter used in this chapter and the rest of this thesis is slightly different to the wild-type promoter (figure 4.1 – A. & table 4.1). It is composed of the three 21 bp repeats and a TATA box which are upstream of the CDS. Downstream of the CDS there are two 72 bp repeats and another three 21 bp repeats. The downstream elements are in the opposite orientation to the upstream elements of the SV40 promoter. This promoter originally came from the pSEAP2-Control vector (Clontech, Mountain View, CA, US).

4.2.2.2 – Human CMV major immediate-early promoter

The human CMV major immediate-early promoter is another commonly used viral promoter. It originates from the human CMV (cytomegalovirus), which is part of the herpesviruses family (Jordan, 1983). *In vivo* the promoter controls the transcription of immediate-early genes of the virus (Stenberg *et al.*, 1985).

In the virus as well as the core promoter other DNA regulatory sequences exist, such as the upstream regulatory sequence, a unique region and a strong enhancer upstream of the core promoter followed by the first exon and then intron A downstream (Chapman *et al.*, 1991; Nelson *et al.*, 1987; Hennighausen & Fleckenstein, 1986; Boshart *et al.*, 1985; Stinski & Roehr, 1985).

The CMV is the most commonly used promoter in the biopharmaceutical industry and in customized and commercial vectors due to its constitutive and high expression in many cell lines. In nearly all DNA vectors containing the CMV promoter, they compose of at least the strong enhancer upstream of the core promoter which contains a TATA

box. However, different companies and researchers use CMV promoters which also contain different combinations of all the other CMV DNA regulatory sequences. The majority of studies using these different combinations only used one or two of the sequences as well as different vector backbones, cell lines and reporter proteins and therefore the effects are probably cell type specific and also possibly reporter protein specific (Mariati *et al.*, 2010; Lee *et al.*, 2007; Xia *et al.*, 2006; Xu *et al.*, 2002; Simari *et al.*, 1998; Ghazal & Nelson, 1991; Nelson *et al.*, 1987; Foecking & Hofstetter, 1986).

The CMV promoter used in this chapter and in the rest of this thesis consists of the core promoter and enhancer (figure 4.1 – B. & table 4.1) and this was mainly due to the vector where this promoter originated (pcDNA3.1 (+), Life Technologies, Paisley, UK).

4.2.2.3 – Human EF1α promoter

The EF1α promoter is a mammalian promoter and controls the transcription of the ubiquitously expressed EF1α (eukaryotic translation elongation factor 1 alpha) (Locus tag: RP11-505P4.2) gene which encodes the alpha sub-unit of the elongation factor-1 complex which delivers aminoacyl tRNAs to ribosomes (Wakabayashiito & Nagata, 1994; Kim *et al.*, 1990; Mizushima & Nagata, 1990; Uetsuki *et al.*, 1989).

The EF1α promoter is composed of a TATA box, a 200 bp upstream 5' flank containing two regulatory elements (EFP1 and EFP2) and the downstream first exon and intron of the EF1α gene (Wakabayashiito & Nagata, 1994; Mizushima & Nagata, 1990). These elements were found to be essential for promoter activity (Wakabayashiito & Nagata, 1994). Another characteristic of this promoter is that it contains a 5' TOP (5' terminal oligopyrimidine tract) element (Shibui-Nihei *et al.*, 2003; Avni *et al.*, 1997) and it is found at the transcriptional start site (beginning of exon 1) (Shibui-Nihei *et al.*, 2003; Uetsuki *et al.*, 1989). A 5' TOP starts with a cytidine residue, and then between 4 and 14 pyrimidine residues (Meyuhas, 2000; Avni *et al.*, 1997; Biberman & Meyuhas, 1997). The EF1a promoter has a C residue followed by five T's (Uetsuki *et al.*, 1989). The 5' TOP forms part of the 5' cap of the resulting mRNA and it has been shown to be

important in translational regulation in a positive cell growth dependent manner (Hamilton *et al.*, 2006; Meyuhas, 2000; Biberman & Meyuhas, 1997).

This promoter, like the CMV, has been shown to give high levels of constitutive expression across many mammalian cell lines (Qin *et al.*, 2010; Gopalkrishnan *et al.*, 1999; Goldman *et al.*, 1996; Mizushima & Nagata, 1990; Uetsuki *et al.*, 1989) and it has been used for high level expression of recombinant therapeutic proteins and other biotechnology applications. It is an alternative to the CMV due to its ability to maintain its expression for longer time durations in some cell lines were the CMV can become silenced (Ramezani *et al.*, 2000; Gopalkrishnan *et al.*, 1999; Nakai *et al.*, 1998; Song *et al.*, 1998; Ye *et al.*, 1998; Guo *et al.*, 1996).

The EF1α promoter (figure 4.1 – C. & table 4.1) was taken from the pEF-GFP vector (Plasmid – 11154, http://www.addgene.org/11154/) which was originally constructed in the lab of Dr. S. Sugano (University of Tokyo) (Kim *et al.*, 1990).

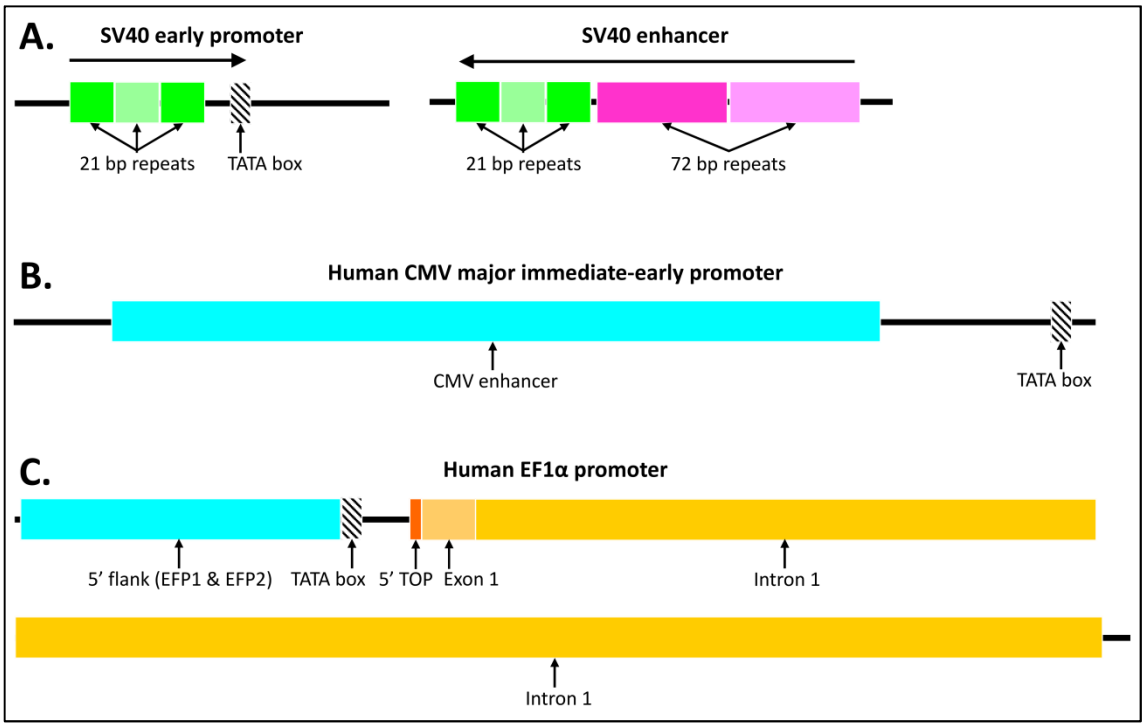

Figure 4.1 – Basic structure of the three promoters used in this chapter. A. SV40 early promoter – The SV40 (Simian virus 40) early promoter is composed of a promoter region (209 bp) downstream of the CDS and an upstream enhancer region (246 bp). The promoter region contains three 21 bp repeats followed by a TATA box. The enhancer region also consists of three 21 bp repeats as well as two 72 bp repeats. The arrows indicate the orientation (5' → 3') of the different elements. This promoter originated from the pSEAP2-control vector (Clontech, Mountain View, CA, US). **B. Human CMV major immediate-early promoter** – The human CMV (Cytomegalovirus) major immediate early promoter (589 bp) is composed of the strong CMV enhancer region and a TATA box. This promoter originates from the pcDNA3.1 (+) vector (Life Technologies, Paisley, UK). **C. Human EF1α promoter** – The human EF1α (eukaryotic translation elongation factor 1 alpha) promoter (1189 bp) is composed of a 5' flank consisting of the EFP1 and EFP2 regions, then a TATA box followed by the first exon and the first intron. At the start of exon 1 is a 5' TOP (5' terminal oligopyrimidine tract) element. This promoter originated from the pEF1-GFP vector (Plasmid – 11154, http://www.addgene.org/11154/), which was originally constructed in the lab of Dr. S. Sugano (University of Tokyo), (Kim *et al.*, 1990).

Table 4.1 – Length and origin of the three promoters studied in this chapter

Promoter	Length (bp)	Origin
SV40 early promoter + enhancer	209 + 246 = 455	Viral
CMV major immediate-early	589	Viral (human)
EF1α	1189	Mammalian (human)

Table shows the lengths and origins of the three promoters studied in this chapter. The SV40 (Simian virus 40) early promoter originally came from the pSEAP2-control vector (Clontech, Mountain View, CA, US). The human CMV (Cytomegalovirus) major immediate-early promoter originally came from pcDNA3.1(+) vector (Life Technologies, Paisley, UK). The human EF1α (eukaryotic translation elongation factor 1 alpha) promoter originally came from the pEF1-GFP vector (Plasmid – 11154, http://www.addgene.org/11154/), which was originally constructed in the lab of Dr. S. Sugano (University of Tokyo), (Kim *et al.*, 1990).

4.3 – Aims and Objectives

The expression and interactions of three promoters (SV40, CMV and EF1α) that are commonly used in both biotechnology and biopharmaceutical research and production are compared. This was done using two reporter proteins, SEAP and GFP. These were chosen due the main research interest of our lab being biopharmaceutical production. Most simple engineering strategies implemented to increase therapeutic recombinant protein production involve the expression of a secreted recombinant glycoprotein and an intracellular expressed effector protein. SEAP acts as our model glycoprotein while GFP will act as our effector protein. Although GFP is not actually an effector protein per se it is being used because it should not have a dramatic effect on the cellular pathways downstream of protein synthesis. Therefore it should allow us to look more directly at the interactions of two promoters in the same system while still being relevant to a cell engineering strategy.

The aims/objectives of this chapter are as follows:

1. Compare the relative expression of the SV40, CMV and EF1α promoters in CHO-S cells to see if the results are comparable with those from the literature. This will be done at both the protein and mRNA level
2. Compare the expression when more than one promoter is co-expressed in the same system in different scenarios:-
    a. When SEAP DNA is kept constant but the amount of GFP DNA is titrated
    b. When GFP DNA is kept constant but the amount of SEAP DNA is titrated
    c. When SEAP and GFP DNA are co-expressed at a one-to-one ratio in terms of DNA vector copy number. Comparisons will be made at both the protein and mRNA level

We hope understanding how these promoters effect each other will help us make better rational decisions for selecting promoters when the expression of more than one recombinant protein is required.

4.4 – Materials and Methods

4.4.1 – Cell line and cell culture

CHO-S cells (Life Technologies, Paisley, UK) were cultured in Erlenmeyer shake flasks (Corning, Surrey, UK) using CD-CHO medium (Gibco®, Life Technologies, Paisley, UK) supplemented with 8 mM L-glutamine (Gibco®, Life Technologies, Paisley, UK). Cells were cultured in 125 or 250 mL cap-vented Erlenmeyer flasks in a shaking incubator set at 37°C, 5% CO_2 and 140 rpm. Cells were passaged every 3-4 days. Viability (%) and viable cell concentration (cells mL^{-1}) were measured using the Vi-Cell automated cell counter (Beckman Coulter, Brea, CA, USA). The required amount of cells for a concentration of 0.2 $x10^6$ cells mL^{-1} in the new culture were centrifuged at 200 g for 5 minutes. The old medium was removed and the cells were resuspended in fresh medium and added to a new flask.

4.4.2 – DNA vectors used and vector engineering

Six different reporter vectors were used in this chapter, SV40-SEAP, CMV-SEAP, EF1α-SEAP, SV40-GFP, CMV-GFP and EF1α-GFP. The origin and construction of SV40-SEAP, CMV-SEAP, SV40-GFP and CMV-GFP were described earlier in chapter 3, section 3.4.2. The EF1α-SEAP and EF1α-GFP vectors were created using the respective SV40 vectors. The SV40 promoter and enhancer were removed via restriction digest. The EF1α promoter was cloned by PCR from the pEF-GFP vector (Plasmid – 11154, http://www.addgene.org/11154/) (Kim *et al.*, 1990) using primers containing overhanging restriction digest sites that matched those in the acceptor vector. After restriction digest the PCR fragments were ligated into the SEAP and GFP vectors.

A control vector was also used for normalising DNA concentrations in all transfections as well as acting as a negative control for transfection. The construction of this vector was also described in chapter 3, section 3.4.2.

Vector DNA was purified using both Miniprep and Maxiprep kits (Qiagen, Manchester, UK) according to the supplied protocol. Vector DNA concentration was measured using the NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, UK).

4.4.3 – Transient transfections

Transfections were carried out on CHO-S cells which had been in culture for at least 2-3 passages but no longer then 8 weeks. Transfection was carried out using PEI (*Polyethylenimine) as the transfection reagent.*

Cells on day 3 of culture were measured for cell concentration and viability. 0.5 mL of CHO-S cells at a concentration of 1.0 x 10^6 cells mL^{-1} were then plated on a 24 well plates (Nunclon™ microwell plates, Thermo Fisher Scientific, UK) in CD-CHO medium with 8 mM L-gluatmine. Plates were then placed in a static incubator at 37°C and 5% CO_2.

One to two hours after plating the cells were transfected with vector DNA. In total 2.0 µg of DNA was transfected in each well. The PEI protocol used for transfection was developed and optimised in our lab for CHO-S cells by Dr. B. Thompson and Dr. O. Mozley. For each well two tubes were set up. One containing 6.5 µl of 150 mM NaCl, 2.35 µl of 300 mM NaCl and 2.35 µl of stock PEI solution (1 mg mL^{-1} in nuclease-free water) (11.2 µl in total). The other containing 2.0 µg of vector DNA and then made up to 22.2 µl using 150 mM NaCl. The two tubes where then added together and the DNA and PEI allowed to complex for one minute before being added to the cells in the culture plate. After DNA addition the plates were gently shook for around one minute before returning to the same incubator. Master mixes were prepared when PEI transfections were scaled up for triplicate samples.

For SEAP vector transfections the culture plates were incubated for 48 hours post transfection before removing the samples from the plate and spinning them down (5 minutes at 200 g) and carefully removing the cell media for further analysis.

Transfection efficiency based on GFP expression was analysed using the Attune® Acoustic Focusing Cytometer and the Attune® Cytometric Software (Life Technologies, Paisley, UK).

4.4.4 – SEAP analysis

The AnaSpec SensoLyte® pNPP Secreted Alkaline Phosphatase Reporter Gene Assay Kit (*Colorimetric*) (Cambridge Bioscience, Cambridge, UK) was used to quantify SEAP protein. The assay was carried out according to the supplied protocol. SEAP samples were incubated in a water bath at 65 °C for 20 minutes to inactivate endogenous non-specific alkaline phosphatase.

4.4.5 – Quantifying gene expression by real-time RT-PCR

Cells were centrifuged at 200 g for 5 minutes and the cell medium was carefully removed. Extraction of mRNA was carried out using the RNeasy® Minikit in combination with the QIAShredder homogeniser according to the manufacturers' protocol (Qiagen, Manchester, UK).

mRNA concentration was measured using the NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, UK).

The TURBO DNA-*free*™ Kit (Applied Biosystems®, Life Technologies) was used to remove any contaminating DNA from mRNA samples and was used according to the supplied protocol. After removal of DNA, mRNA concentrations were measured again using the NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, UK).

mRNA was transcribed into DNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems®, Life Technologies) following the manufacturers protocol.

SEAP mRNA expression was quantified by RT-qPCR using the Applied Biosystems 7500 Fast Real-Time PCR system (Applied Biosystems®, Life Technologies) using an absolute quantification method (Pfaffl *et al.*, 2002). The SV40-SEAP vector was used as a standard.

SEAP primers were designed against the SEAP coding sequence and used at a concentration of 300 nM (SEAP FWD 5' – CCATATGTGGCTCTGTCCAA – 3', SEAP REV 5' – GTCTGGAAGTTGCCCTTGAC – 3').

The Power SYBR® Green PCR Mastermix (Applied Biosystems®, Life Technologies) was used according to the manufacturer's instructions. 1 µl of cDNA was added to each reaction. The following PCR cycle program was used: 1. 50°C for 2 minutes; 2. 95°C for 10 minutes; then 40 repeats of 3. 95°C for 15 seconds then 60°C for 1 minute. A dissociation step was added to the end of procedure to check for the specificity of amplification.

4.4.6 – Fitted SEAP inhibition curves

In section 4.5.3, SEAP titrations are carried out with a fixed amount of a competing GFP vector. The raw data was then used to produce fitted curves using the Michaelis-Menten equation as a model:-

$$V = \frac{V_{max}\,[S]}{K_m + [S]}$$

Where V equals the concentration of SEAP protein produced (ng mL^{-1}), S equals the amount of SEAP DNA transfected (µg), V_{max} is the maximum concentration of SEAP protein produced and K_m is the amount of transfected SEAP DNA (S) required to produce half of the maximum concentration of SEAP protein (half V_{max}).

This was done in Excel using the Solver Add-in program through iterations of changing the predicted values of K_m and V_{max} until the sum of the square differences between the calculated and actual SEAP concentrations was at a minimum. The quality of the fit was then analysed by working out the R-squared value between actual and calculated SEAP concentrations. All fits had an R-squared value of ≥ 0.99.

The predicted V_{max} and K_m were then used to draw curves in Excel.

K_i values were calculated from the predicted K_m values using the follow equation:-

$$K_i = \frac{K_{m\,control}[I]}{K_{m\,inhibitor} - K_{m\,control}}$$

$K_{m\ control}$ is the K_m value under control conditions, [I] is the concentration of the inhibitor (or competing vector) present and $K_{m\ inhibitor}$ is the K_m value when in the presence of the inhibitor (or competing vector).

4.5 – Results and Discussion

4.5.1 – Relative expression of the SV40, CMV and EF1α promoters

To show the relative strengths of the SV40, CMV and EF1α promoters their corresponding SEAP DNA vectors were transfected into CHO-S cells over a range of eight different DNA concentrations (0.015625 μg, 0.03125 μg, 0.0625 μg, 0.125 μg, 0.25 μg, 0.5 μg, 1.0 μg and 2.0 μg) and SEAP protein and mRNA expression measured 48 hours post-transfection.

From figure 4.2 and table 4.1 it can be seen that the SV40-SEAP vector produced less SEAP protein and mRNA then both the other two vectors across the whole range of transfected DNA amounts. This was expected due the SV40 promoter being known to be a weaker promoter then both the CMV and EF1α promoters in many cell lines (Qin *et al.*, 2010; Zarrin *et al.*, 1999; Liu *et al.*, 1997; Thompson *et al.*, 1990; Foecking & Hofstetter, 1986). Similar titrations in chapter 3 (figure 3.2) also showed the SV40-SEAP vector to produce less SEAP protein over a range of transfected DNA amounts then the CMV-SEAP vector. However, the differences seen in expression in this chapter were not as large as those seen chapter 3. This is probably due to both the different transfection protocols and the different CHO cell lines used.

The CMV-SEAP and EF1α-SEAP vectors produced very similar amounts of SEAP protein for the majority of DNA concentrations (figure 4.2 – A. & table 4.2). Only at the lowest two DNA amounts did the CMV-SEAP vector produce slightly more than the EF1α-SEAP vector. When looking at the amount of SEAP mRNA produced from vectors containing either the CMV or EF1α promoter it can be seen that the CMV produced more mRNA at all amounts of transfected DNA then the EF1α (figure 4.2 – B. & table 4.2). The differences in mRNA expression were larger over the range of 0.015625 – 0.125 μg of DNA. For the larger amounts of transfected DNA (≥ 0.25 μg) the differences in SEAP mRNA expression became increasingly smaller. The different patterns seen in mRNA expression between CMV-SEAP and EF1α-SEAP compared to SEAP protein expression (figure 4.2 – A. & B.) might be due to the differences in mRNA stability or processing and this could be caused by the presence of the 5' TOP element within the EF1α promoter (Shibui-Nihei *et al.*, 2003; Avni *et al.*, 1997).

These results confirm that the SV40 promoter, as expected, was indeed the weakest of the three promoters tested. It produced the least protein and mRNA across all amounts of transfected DNA. The SEAP protein results for both the CMV and EF1α promoter's vectors were very similar but the CMV promoter did produce more SEAP mRNA for most of the transfected amounts of DNA. Both the CMV and EF1α promoters are known to be high expressing promoters so their similarities was not unexpected and the differences seen between their SEAP mRNA and protein expression may be due to the stability or the processing of the transcripts that each promoter produces.

From this data the three different promoters can be ranked accordingly:-

- Overall SEAP protein expression – CMV = eEF1a > SV40
- SEAP mRNA expression – CMV > eEF1a > SV40

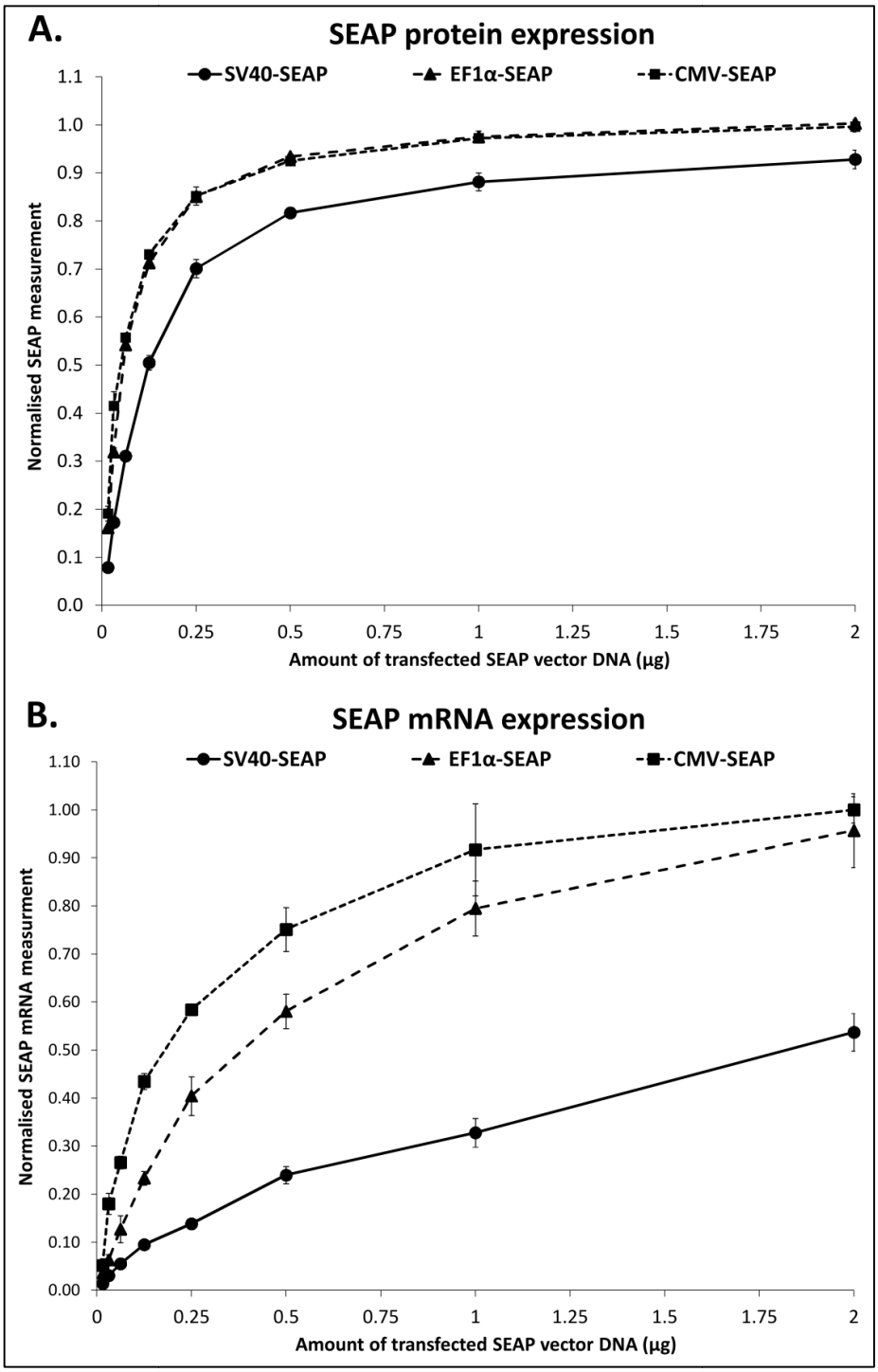

Figure 4.2 – Comparison of SEAP protein and mRNA expression from SEAP DNA vectors utilising three different promoters. **A.** SEAP protein expression for eight differing amounts (0.015625 µg, 0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 10 µg and 2.0 µg) of SEAP vector DNA from three SEAP vectors (● SV40-SEAP, ▲ EF1α-SEAP, ■ CMV-SEAP) utilising three different promoters transiently expressed in CHO-S cells. SEAP protein expression was normalised to the average SEAP measurement for 2.0 µg of CMV-SEAP. **B.** SEAP mRNA expression for the different SEAP vectors taken from the same samples as used in A. SEAP mRNA expression was normalised to the average SEAP measurement for 2.0 µg of CMV-SEAP. The total amount of transfected DNA was kept consistent in all transfections by using an empty –ve control vector. Media and cell samples were collected 48 hours post-transfection and analysed for SEAP protein and mRNA expression. N = 3, error bars represent ± 1 S.D.

Table 4.2 – SEAP protein and mRNA expression represented as a percentage of the measurements for 2.0 µg of transfected CMV-SEAP DNA

	SEAP DNA vectors					
	SV40-SEAP		EF1α-SEAP		CMV-SEAP	
Amount of DNA	Protein	mRNA	Protein	mRNA	Protein	mRNA
0.015625 µg	7.9%	1.4%	16.2%	3.6%	19.2%	5.2%
0.03125 µg	17.3%	3.0%	32.0%	6.3%	41.7%	18.05
0.0625 µg	31.2%	5.5%	54.4%	12.7%	55.9%	26.6%
0.125 µg	50.7%	9.5%	71.5%	23.3%	73.3%	43.5%
0.25 µg	70.3%	13.8%	85.4%	40.4%	85.5%	58.4%
0.5 µg	82.0%	24.0%	93.7%	58.1%	92.8%	75.1%
1.0 µg	88.5%	32.8%	97.9%	79.5%	97.6%	91.7%
2.0 µg	93.1%	53.7%	100.7%	95.7%	100.0%	100.0%

The numbers in the table represent the percentage of SEAP protein and mRNA expression when compared to 2.0 µg of transfected CMV-SEAP DNA. Calculated using the data in figure 4.2.

4.5.2 – Promoter interference between the SV40, CMV and EF1α promoters: The effect of titrating GFP vectors utilizing different promoters on SEAP protein expression from constant amounts of SEAP vectors also utilizing different promoters

To investigate the effects on expression when more than one promoter is used in the same system all combinations of the SEAP and GFP vectors using the different promoters were transfected into CHO-S cells. The amount of SEAP DNA was kept constant (0.0625 μg) while the amount of GFP DNA was titrated (0.015625 μg, 0.03125 μg, 0.0625 μg, 0.125 μg, 0.25 μg, 0.5 μg, 1.0 μg). SEAP expression was then measured 48 hours post-transfection to see the effects of different levels of GFP expression driven by different promoters had on SEAP expression.

4.5.2.1 – SV40 promoter

The SV40-SEAP vector was co-transfected with the SV40-GFP, CMV-GFP and EF1α-GFP vectors. SV40-SEAP was also transfected alone to determine the control level of expression in this experiment. All SEAP measurements where then compared to this control.

Figure 4.3 shows that for all GFP vectors co-transfected with the SV40-SEAP vector there were decreases in SEAP expression from the control. The decreases in SEAP protein expression increased with increasing amounts of co-transfected GFP DNA. GFP vectors containing the CMV or EF1α promoters caused the biggest decreases in SEAP expression. For all amounts of co-transfected GFP DNA the CMV and EF1α GFP vectors had roughly similar effects. When these vectors were at a one-to-one ratio to the SV40-SEAP vector (in terms of μg's of DNA) SEAP expression was reduced to about 60% of the control. SEAP had been reduced to around 50% when 0.125 μg of CMV-GFP and EF1α-GFP were co-transfected. For the largest amount of co-transfected GFP (1.0 μg) SEAP expression had reduced to around 10-15%.

Co-expression of the SV40-GFP vector with SV40-SEAP also caused a reduction in SEAP protein expression but not to the same extent as that seen for the other two promoters (figure 4.3). SEAP expression was reduced to 90% of the control when

these vectors were co-transfected at a one-to-one ratio. A near 50% reduction was only seen when 1.0 µg of SV40-GFP DNA was co-transfected.

From this data we can see that GFP vectors containing either the CMV or EF1α promoters had a much greater negative effect on SEAP expression from SV40-SEAP vector then the SV40-GFP vector.

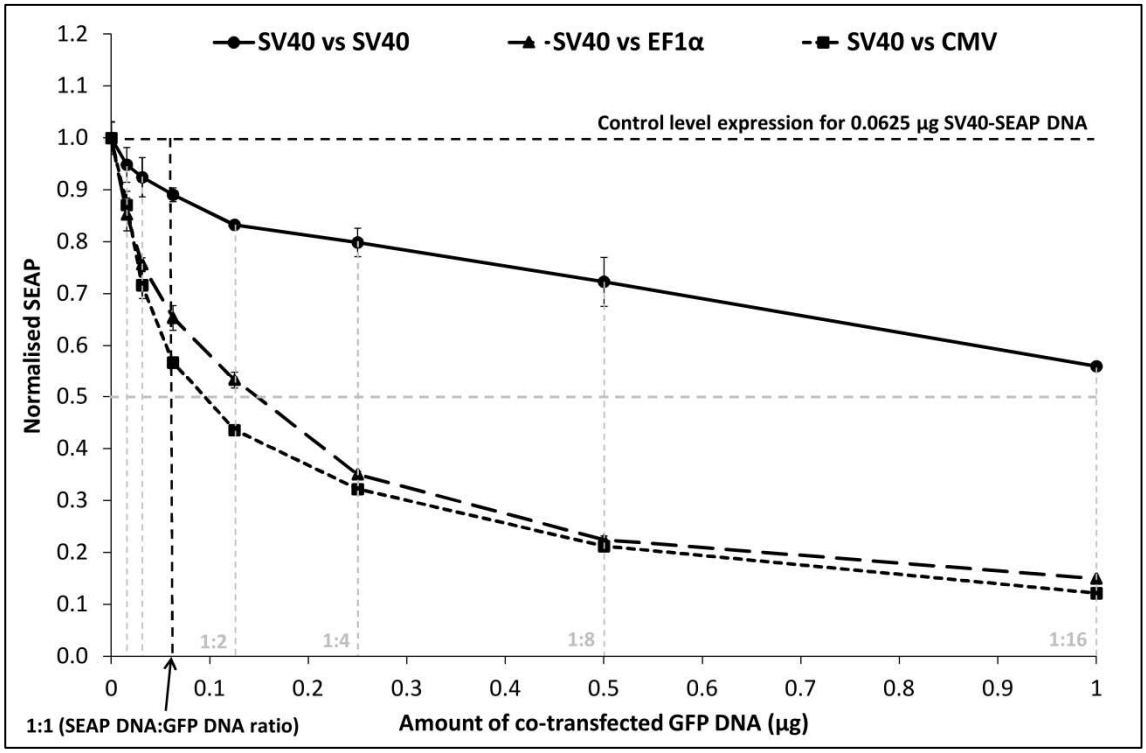

Figure 4.3 – The effect of a competing promoter from another vector on the expression of SEAP protein from the SV40-SEAP vector. SEAP expression for a fixed amount of SV40-SEAP DNA (0.0625 µg) co-transfected with differing amounts (0.015625 µg, 0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg) of GFP expressing vectors utilising three different promoters (● SV40-GFP, ▲ EF1α-GFP, ■ CMV-GFP) in CHO-S cells. The total amount of transfected DNA was kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to a control sample (0.0625 µg of SV40-SEAP co-transfected with –ve control DNA only). The ratios in grey represent the ratio of SEAP DNA to GFP DNA (µg). N = 3, error bars represent ± 1 S.D.

4.5.2.2 – EF1α promoter

The EF1α-SEAP vector was also co-transfected with the SV40-GFP, CMV-GFP and EF1α-GFP vectors. EF1α-SEAP was also transfected alone to determine the control level of expression in this experiment. All SEAP measurements where then compared to this control.

Figure 4.4 shows the effects of co-transfecting different amounts of SV40-GFP, CMV-GFP and EF1α-GFP on SEAP expression from the EF1α-SEAP vector. Co-transfection of SV40-GFP mostly had no effect on SEAP expression. For the largest amount of co-transfected SV40-GFP DNA (1.0 μg), which was 16 times the amount of EF1α-SEAP, there was only a reduction to around 90% of the control.

The results were much more significant when the EF1α-SEAP vector was co-transfected with GFP vectors utilizing the CMV or EF1α promoter (figure 4.4). Both the CMV-GFP and EF1α-GFP vectors had a similar effect on reducing SEAP expression but the CMV-GFP reduced the level of SEAP to a slightly greater extent for every amount of co-transfected GFP DNA. When the CMV-GFP and EF1α-GFP vectors were co-transfected at a one-to-one ratio with the SEAP vector, SEAP protein expression was reduced to around ~78% and ~88%, respectively. SEAP protein expression was reduced to below 50% when 0.5 μg of these two GFP vectors were co-transfected. When 1.0 μg of CMV-GFP and EF1α-GFP was co-transfected with EF1α-SEAP, SEAP expression dropped to between 20-30%.

This data shows that a GFP vector containing an SV40 promoter negligible effect on SEAP expression from a vector containing an EF1α promoter. Only when there was 16 times more GFP DNA then SEAP DNA was there a reduction seen and this was to only to around 90% of the control. GFP vectors utilizing either the CMV or EF1α promoters had a much greater effect and were a much more negative influence on SEAP expression. Although CMV and EF1α GFP vectors showed a similar pattern in reducing SEAP expression the CMV-GFP vector generally affected SEAP slightly more for all co-transfected amounts of GFP DNA.

The results in figure 4.4 and 4.3 have similar patterns in so much as the SV40 promoter had the least negative effects while the CMV and EF1α promoters had similar effects with the CMV-GFP reducing SEAP expression slightly more than EF1α-GFP.

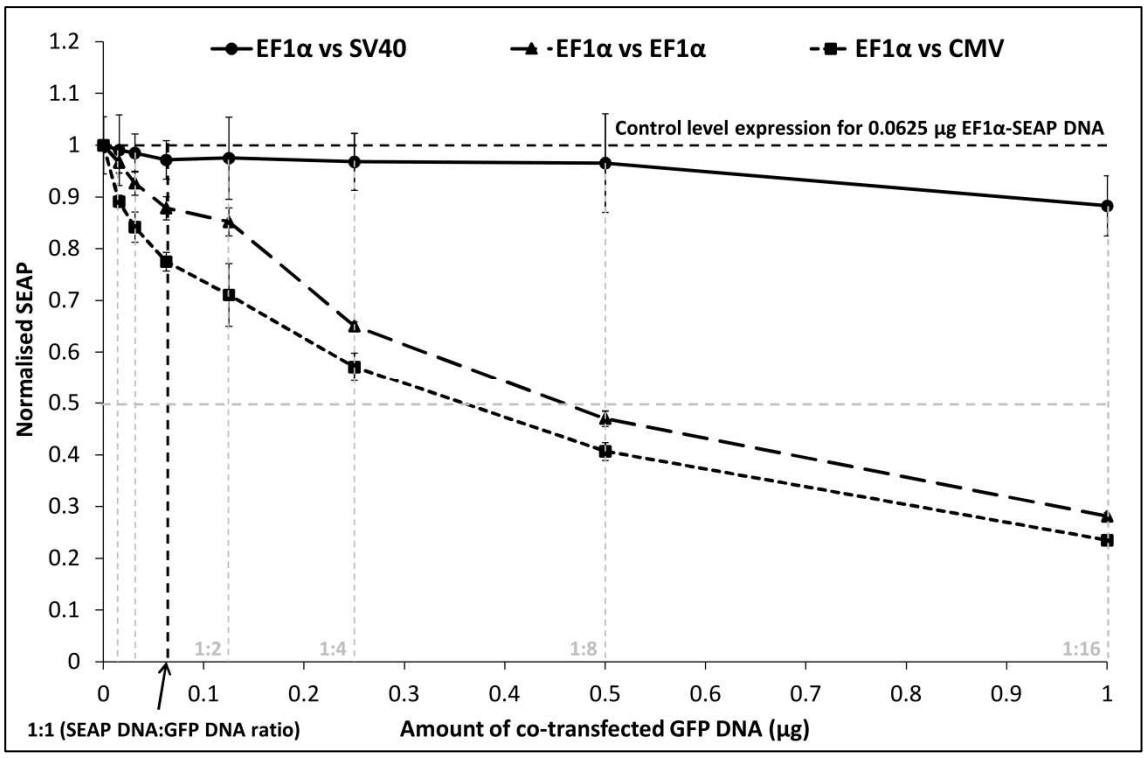

Figure 4.4 – The effect of a competing promoter from another vector on the expression of SEAP protein from the EF1α-SEAP vector. SEAP expression for a fixed amount of EF1α-SEAP DNA (0.0625 µg) co-transfected with differing amounts (0.015625 µg, 0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg) of GFP expressing vectors utilising three different promoters (● SV40-GFP, ▲ EF1α-GFP, ■ CMV-GFP) in CHO-S cells. The total amount of transfected DNA was kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to a control sample (0.0625 µg of EF1α-SEAP co-transfected with –ve control DNA only). The ratios in grey represent the ratio of SEAP DNA to GFP DNA (µg). N = 3, error bars represent ± 1 S.D.

4.5.2.3 – CMV promoter

The SV40-GFP, CMV-GFP and EF1α-GFP vectors were co-transfected with the CMV-SEAP vector. CMV-SEAP was also transfected alone to determine the control level of expression in this experiment. All SEAP measurements where then compared to this control.

Figure 4.5 shows the effects of co-transfecting the three GFP vectors on SEAP expression from CMV-SEAP. Similar to the EF1α-SEAP vector (figure 4.4) the SV40-GFP vector had no negative effect on SEAP expression from CMV-SEAP across the whole range of DNA amounts co-transfected.

Co-transfection of CMV-GFP and EF1α-GFP vectors again had negative effects on SEAP expression (figure 4.5). Both GFP vectors had similar effects on SEAP expression from the CMV-SEAP vector from the lowest amount of co-transfected DNA to 0.0625 μg DNA. At this one-to-one ratio SEAP expression was around 90% of the control. After this, although both the CMV and EF1α GFP vectors still had negative effects, the levels of reduction in SEAP expression significantly diverged between these two vectors. A ~50% reduction in SEAP was achieved when 0.5 μg of CMV-GFP DNA was co-transfected and this decreased to around 25% when 1.0 μg was co-transfected. The results of CMV-GFP co-transfection with CMV-SEAP were quite similar to the results for EF1α-SEAP and CMV-GFP in figure 4.4. When more than 0.0625 μg of EF1α-GFP DNA was co-transfected with CMV-SEAP (figure 4.5) there was less negative effect when compared to CMV-GFP. SEAP expression did not drop below 50% for any amount of co-transfected EF1α-GFP. When the largest amount of EF1α-GFP (1.0 μg) was co-expressed SEAP expression was around 60% of the control.

In summary the results from figure 4.5 showed that SV40-GFP had no effect on SEAP expression while EF1α-GFP did it was not to the same extent as the GFP vector utilizing the CMV promoter.

In comparison to results in figures 4.3 and 4.4 the SV40-GFP vector again had the least negative effects on SEAP expression. GFP vectors containing CMV and EF1α promoters had more of negative effects but this time the CMV-GFP vector decreased SEAP expression more significantly then EF1α-GFP.

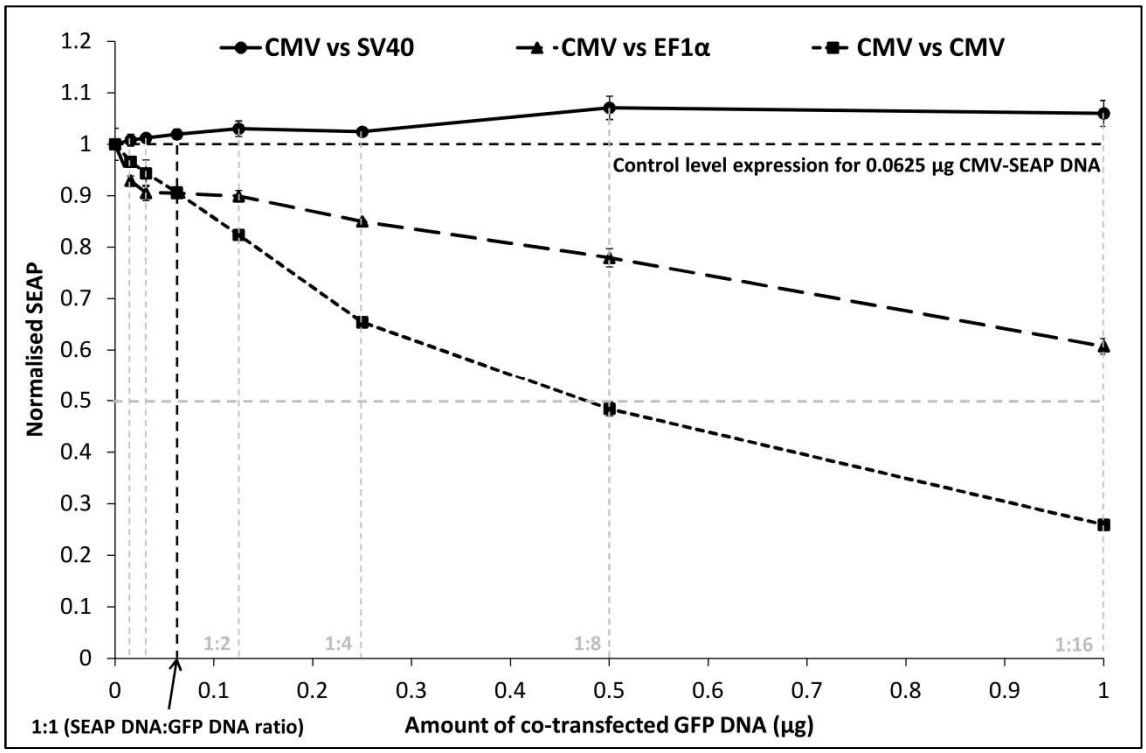

Figure 4.5 – The effect of a competing promoter from another vector on the expression of SEAP protein from the CMV-SEAP vector. SEAP expression for a fixed amount of CMV-SEAP DNA (0.0625 µg) co-transfected with differing amounts (0.015625 µg, 0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg) of GFP expressing vectors utilising three different promoters (● SV40-GFP, ▲ EF1α-GFP, ■ CMV-GFP) in CHO-S cells. The total amount of transfected DNA was kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to a control sample (0.0625 µg of CMV-SEAP co-transfected with –ve control DNA only). The ratios in grey represent the ratio of SEAP DNA to GFP DNA (µg). N = 3, error bars represent ± 1 S.D.

4.5.2.4 – Comparisons and conclusions

Table 4.3 shows the collated results from figure 4.3, 4.4 and 4.5 in terms of fold change from their control. From these set of experiments we can see that expression of SEAP was most negatively affected by a competing vector/promoter when it was under the control of the SV40 promoter. All three GFP vectors significantly affected SEAP expression but there was a clear difference between the effects of SV40-GFP and both CMV-GFP and EF1α-GFP across the whole range of co-transfected DNA amounts. When the GFP vectors were titrated against the EF1α-SEAP and CMV-SEAP vectors only the CMV and EF1α GFP vectors caused a significant reduction in SEAP expression. The SV40-GFP had very little negative effects on SEAP expression from the EF1α vector and no negative effects on the CMV-SEAP vector. Both the CMV-GFP and EF1α-GFP vectors negatively affected SEAP in similar ways for both SV40-SEAP and EF1α-SEAP but this

was not the case when co-transfected with CMV-SEAP. The CMV-GFP vector had greater negative effects on CMV-SEAP expression with increasing amounts of GFP DNA when compared to EF1α-GFP.

This fits in with the strengths of the promoters used. In figure 4.2 it was shown that for the same amount of DNA (0.0625 μg) the SV40 vector produced around half the amount of SEAP protein (~57%) then that of the CMV and EF1α SEAP vectors which had very similar expression. Greater differences were seen when the levels of mRNA were measured with SV40 producing around 20% and EF1α producing ~50% of that produced from the CMV-SEAP vector.

The different results seen when co-transfecting the GFP vectors utilizing different promoters is the interesting result, but when the amount of SEAP DNA is kept constant it is not unexpected that increasing the amount of a competing vector has a negative effect on SEAP expression. However, to show that these results are not simply due to increasing the amount of co-transfected GFP DNA we also need to look at the effects of a constant amount of co-transfected GFP DNA on a range of SEAP DNA amounts. This is carried out in the next section.

Table 4.3 – Summary of the fold change in SEAP protein expression of the results from figures 4.3, 4.4 and 4.5

	SV40-SEAP vector			EF1α-SEAP vector			CMV-SEAP vector		
	Promoter of the competing GFP vector			Promoter of the competing GFP vector			Promoter of the competing GFP vector		
Ratio of SEAP DNA: GFP DNA	SV40	EF1α	CMV	SV40	EF1α	CMV	SV40	EF1α	CMV
1:0.25	-1.05	-1.17	-1.15	-	-	-1.12	-	-1.08	-
1:0.5	-1.08	-1.32	-1.46	-	-1.08	-1.19	-	-1.10	-1.06
1:1	-1.12	-1.53	-1.76	-	-1.14	-1.29	-	-1.11	-1.10
1:2	-1.20	-1.88	-2.30	-	-1.17	-1.41	-	-1.11	-1.21
1:4	-1.25	-2.85	-3.11	-	-1.54	-1.75	-	-1.18	-1.53
1:8	-1.38	-4.46	-4.71	-	-2.13	-2.45	-	-1.28	-2.07
1:16	-1.78	-6.68	-8.24	-1.13	-3.55	-4.25	-	-1.65	-3.86

The numbers in the table represent the fold change in SEAP protein expression compared to the control of the respective SEAP vectors. "-" indicates where there was no negative fold change in SEAP protein expression when compared to the control sample. The fold changes were from the data in figure 4.3, 4.4 and 4.5.

4.5.3 – Promoter interference between the SV40, CMV and EF1α promoters: The effect of a fixed amount of GFP DNA from vectors utilizing different promoters on SEAP titration curves

In the last set of experiments it was seen that with increasing the amount of a competing vector/promoter that SEAP expression would decrease. Although this result was not unexpected, it was the different responses seen depending on the two promoters present that was the interesting result. To further investigate the effects of two competing promoters and to show it is not just an effect of increasing the amount of GFP vector present we kept GFP DNA constant (0.25 μg) while titrating the amount of SEAP DNA (0.015625 μg, 0.03125 μg, 0.0625 μg, 0.125 μg, 0.25 μg, 0.5 μg, 1.0 μg and 1.75 μg) similar to section 4.5.1, figure 4.2. The same combinations of co-transfected GFP and SEAP vectors as in section 4.5.2 were used.

The results in section 4.5.1 (figure 4.2) seemed to resemble curves for enzymatic reactions. Due to this similarity, the new curves produced in this section were treated as such and the competing GFP vectors considered inhibitors. The raw data generated was used to produce fitted curves using the Michaelis-Menten equation as a model (see section 4.4.6 for a description of this was done). This allowed us to make estimates for K_m (concentration required for half the maximum expression) values and allow us to compare the effect of different GFP vectors with different competing promoters on SEAP expression. Lineweaver-Burk (LB) plots were also made (by plotting the reciprocals of both DNA concentration and SEAP expression against each other) for further graphical representation of the effects of GFP vectors and the different promoters utilized. K_i values were also calculated and these reflect the level of inhibition a GFP vector has on a SEAP vector allowing for further comparison of the effects of the different promoters.

4.5.3.1 – SV40 promoter

Figure 4.6 shows the effects of co-transfecting a fixed amount (0.25 μg) of SV40-GFP, CMV-GFP and EF1α-GFP on titrations of SV40-SEAP DNA. All three GFP vectors caused a change in the shape of the fitted curves when compared to the control (figure 4.6 – A.). Co-expression of the CMV-GFP vector caused the biggest shift in the curve by reducing the amount of SEAP produced for all transfected amounts when compared to the control. EF1α-GFP also caused a shift in the curve but not to the same extent as CMV-GFP while SV40-GFP had the least effect of all of the GFP vectors.

These results correlate with those in the previous section (4.5.2.1) that GFP vectors containing CMV and EF1α promoters had the greatest negative impact on SEAP expression from a vector utilizing the SV40 promoter. The LB plots (figure 4.6 – B.) also show that co-transfection of CMV-GFP caused the greatest inhibition on SEAP expression from the SV40-SEAP vector, EF1α-GFP the second greatest level of inhibition and SV40-GFP the least. This plot also suggests that competitive inhibition between the vectors/promoters is occurring due to the fact that lines differ in their steepness but all cross the *y* axis at the same point. However, the LB plot has limitations and it cannot be used as proof of competitive inhibition.

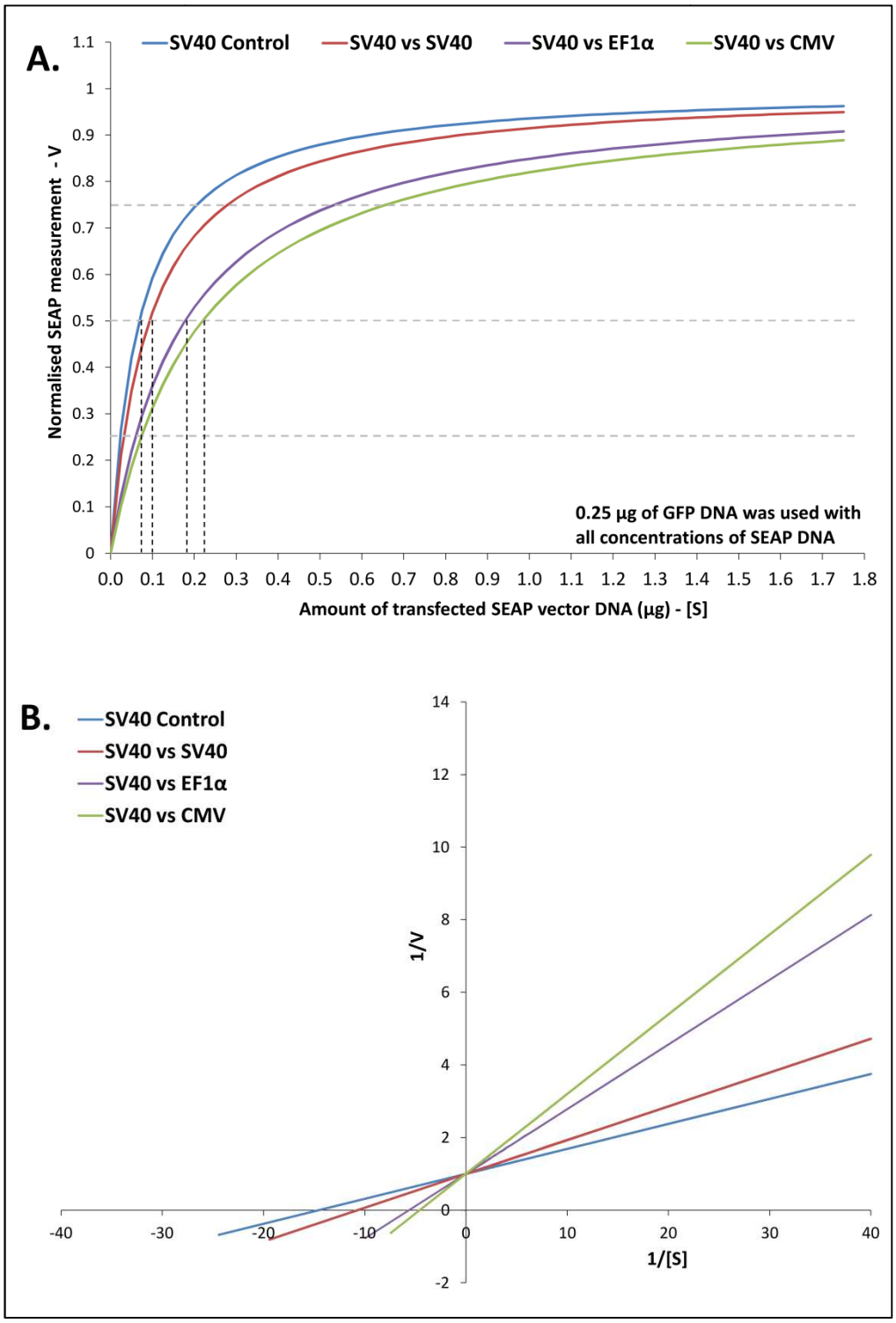

Figure 4.6 – Fitted plots showing the effect of competing promoters on SEAP protein expression from the SV40-SEAP vector. A. The graph shows the effects of co-transfecting 0.25 µg of three different GFP vectors utilising three different promoters (SV40-GFP, EF1α-GFP, CMV-GFP) with differing amounts (0.015625 µg, 0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg and 1.75 µg) of the SV40-SEAP DNA vector in CHO-S cells. The control contained no GFP DNA. The total amount of transfected DNA was kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to SEAP expression from 1.75 µg SV40-SEAP DNA under control conditions. The initial raw data was then fitted using the Michaelis-Menten model and plotted. All fitted lines had an $R^2 \geq 0.99$. **B.** Lineweaver-Burk (or double reciprocal) plot of the data in A.

4.5.3.2 – EF1α promoter

As with the SV40-SEAP vector the three GFP vectors were then co-transfected with the EF1α-SEAP vector. Figure 4.7 shows the results and it can be seen that the CMV-GFP again had biggest effect of inhibiting SEAP expression, EF1α-GFP the second and SV40-GFP has the least effect on SEAP expression. The shifts in the curves (figure 4.7 – A.) were not as pronounced as those seen for the SEAP vector utilizing the weaker SV40 promoter (figure 4.6 – A.).

Again the levels of reduction in SEAP expression for the GFP vectors using the three different promoters match what has previously been shown in this chapter. The LB plot (figure 4.7 – B.), presents the data in another graphical form to show that the CMV driven GFP vector had the greatest effect of inhibiting SEAP expression followed by the EF1α driven GFP vector and then the SV40. The changes in steepness of the lines in the LB plot (figure 4.6 – B.) were not as great as those seen for the SV40-SEAP vector (figure 4.6 – B.).

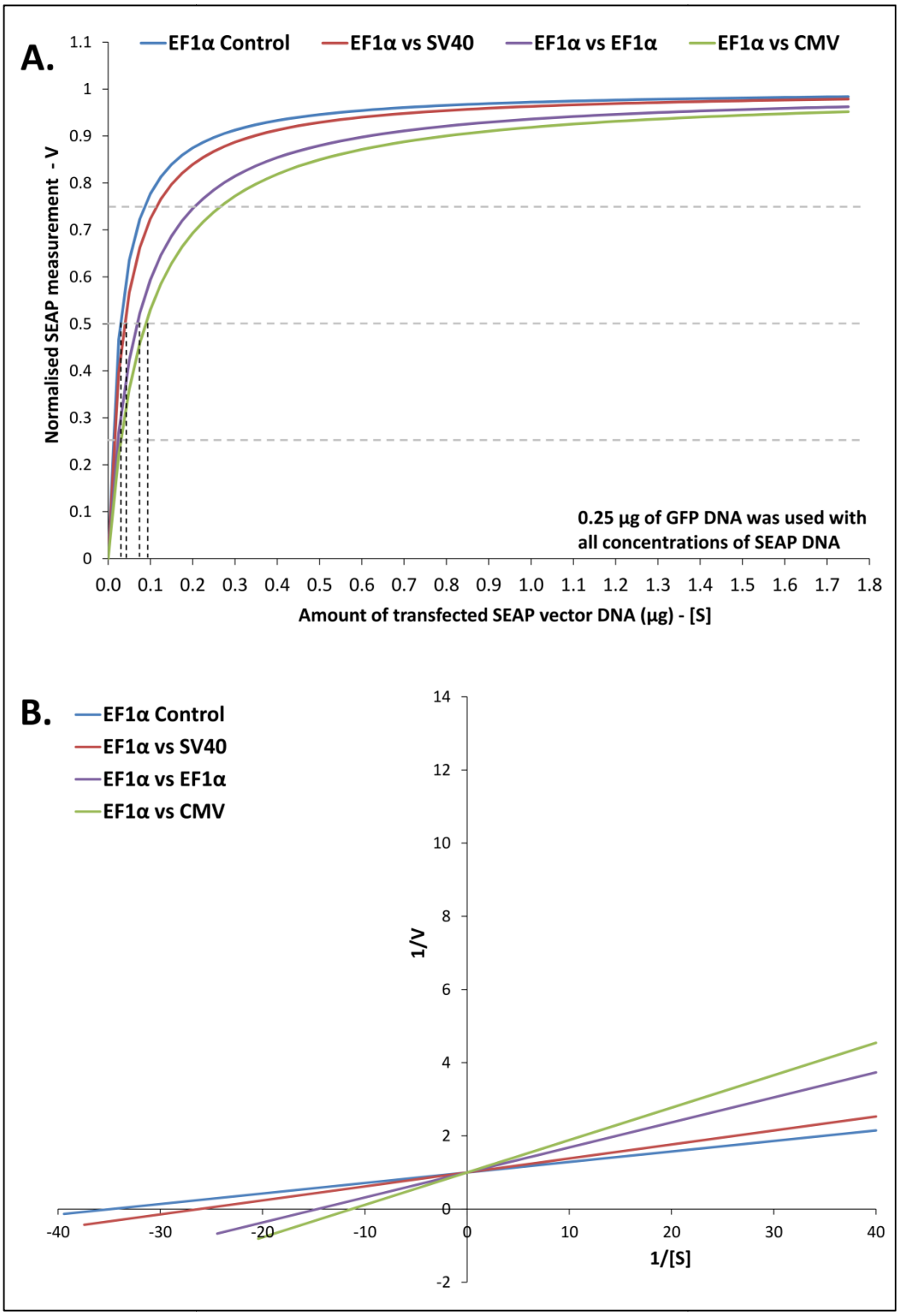

Figure 4.7 – Fitted plots showing the effect of competing promoters on SEAP protein expression from the EF1α-SEAP vector. **A.** The graph shows the effects of co-transfecting 0.25 µg of three different GFP vectors utilising three different promoters (SV40-GFP, EF1α-GFP, CMV-GFP) with differing amounts (0.015625 µg, 0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg and 1.75 µg) of the EF1α-SEAP DNA vector in CHO-S cells. The control contained no GFP DNA. The total amount of transfected DNA was kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to SEAP expression from 1.75 µg EF1α-SEAP DNA under control conditions. The initial raw data was then fitted using the Michaelis-Menten model and plotted. All fitted lines had an $R^2 \geq 0.99$. **B.** Lineweaver-Burk (or double reciprocal) plot of the data in A.

4.5.3.3 – CMV promoter

Figure 4.8 shows the effects of co-transfecting a fixed amount (0.25 μg) of SV40-GFP, CMV-GFP and EF1α-GFP on titrations of CMV-SEAP DNA. Again the CMV-GFP vector had the greatest negative impact on SEAP expression, but this time both the GFP vectors using the EF1α and SV40 promoter had little to no effect (figure 4.8). The shift in the curve caused by co-transfection of CMV-GFP (figure 4.8 – A.) was to a smaller extent when compared to the effect of CMV-GFP on the other two SEAP vectors (figure 4.6 & 4.7).

As for both the other two SEAP plasmids the results for this experiment using CMV-SEAP match what has been previously been shown in this chapter. Namely, that a larger negative effect on SEAP expression is seen when a GFP vector utilizing a CMV promoter is co-transfected then with the SV40 and EF1α promoters. The LB plot for CMV-SEAP shows this as well (figure 4.8 – B.). Conversely, this data also shows that a SEAP vector utilizing a CMV promoter is less prone to having its expression effected then the SV40 and EF1α promoters. Which makes sense since any interaction between the SEAP and GFP vectors will occur in both directions, i.e. – CMV-SEAP will also be effecting the expression of competing GFP vectors.

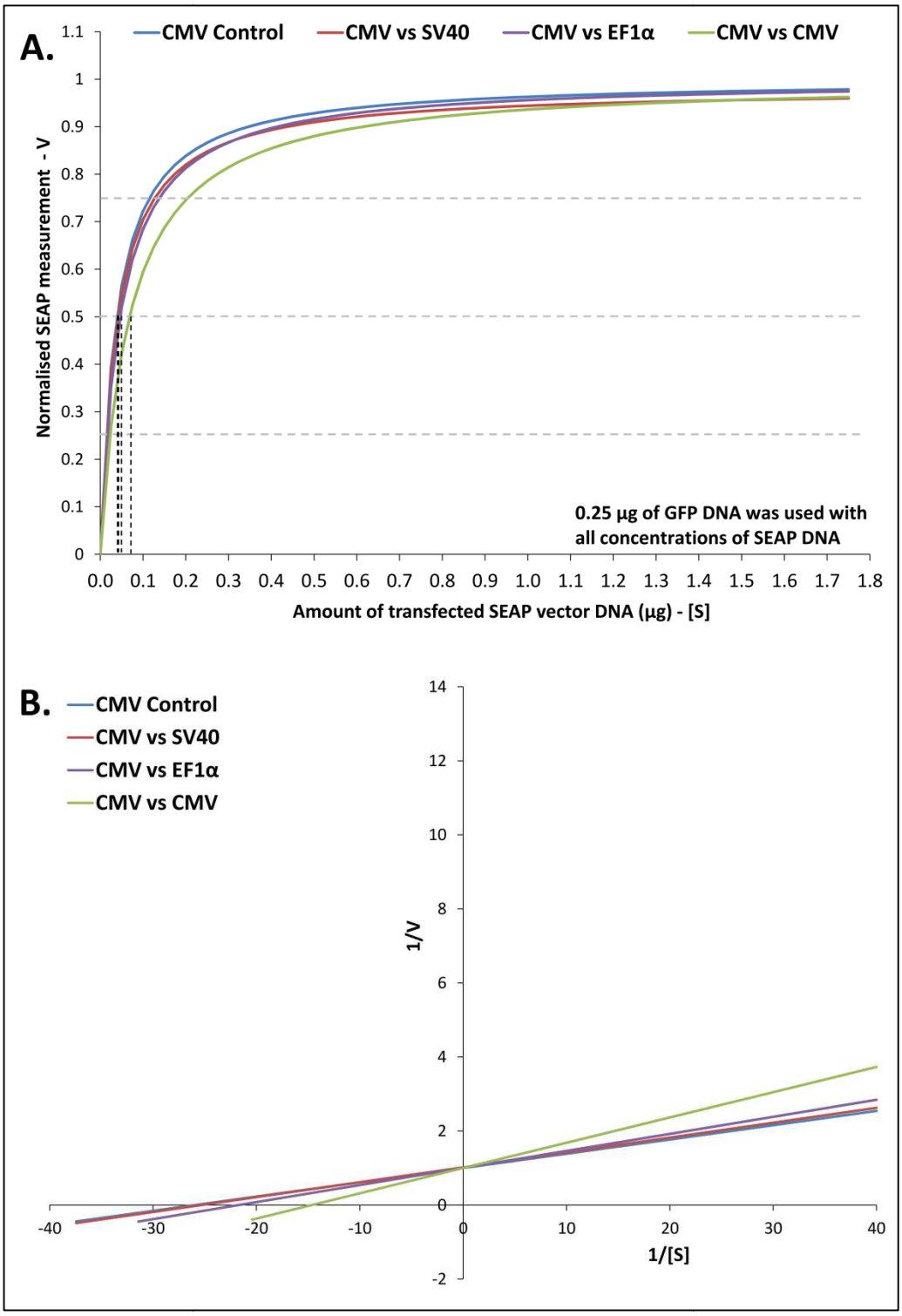

Figure 4.8 – Fitted plots showing the effect of competing promoters on SEAP protein expression from the CMV-SEAP vector. A. The graph shows the effects of co-transfecting 0.25 µg of three different GFP vectors utilising three different promoters (SV40-GFP, EF1α-GFP, CMV-GFP) with differing amounts (0.015625 µg, 0.03125 µg, 0.0625 µg, 0.125 µg, 0.25 µg, 0.5 µg, 1.0 µg and 1.75 µg) of the CMV-SEAP DNA vector in CHO-S cells. The control contained no GFP DNA. The total amount of transfected DNA was kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to SEAP expression from 1.75 µg CMV-SEAP DNA under control conditions. The initial raw data was then fitted using the Michaelis-Menten model and plotted. All fitted lines had an $R^2 \geq 0.99$. **B.** Lineweaver-Burk (or double reciprocal) plot of the data in A.

4.5.3.4 – Comparisons and conclusions

Table 4.4 shows the fold changes, when compared to their respective controls, in the estimated K_m values from the fitted data presented in figures 4.6, 4.7 and 4.8. The K_m in this situation is the amount of SEAP DNA required to produce half the maximum SEAP protein expression (V_{max}) for each individual SEAP expression vector. This is another way to show how the SEAP titration curves changed when a competing GFP vector was co-transfected. From table 4.4, we can see that the presence of CMV-GFP brought about the biggest fold changes in K_m for all three SEAP vectors used while SV40-GFP had the least effect.

Estimated K_i values were also calculated from the fitted data presented in figures 4.6, 4.7 and 4.8 and are shown in table 4.5. The K_i value reflects the level of inhibition that a competing GFP vector has on a SEAP vector. The lower the K_i value the greater the level of inhibition. For both the SV40-SEAP and EF1α-SEAP vectors the K_i values for each respective competing GFP vector were roughly similar with the SV40-GFP vector having the highest K_i value and CMV-GFP having the lowest. For the CMV-SEAP vector the K_i values for all GFP vectors were higher than for the other two SEAP vectors but again the SV40-GFP vector had the highest K_i while the CMV-GFP vector had the lowest.

The work so far suggests that a vector containing a CMV promoter has a greater negative influence on the expression of a competing vector when compared to both the EF1α and SV40 promoter. In turn, the EF1α promoter has a greater negative influence when compared to the SV40 promoter. The level of this negative interference could be due to the strengths of the promoters. From figure 4.2 – B., it was shown that the CMV-SEAP vector generally produced more mRNA then EF1α-SEAP which produced more than the SV40-SEAP vector. Since it is the promoters in these vectors that control transcription and therefore mRNA production from the vectors the CMV can be seen as the strongest in terms of transcriptional activity. Therefore, it is likely to out-compete transcriptionally weaker promoters for important factors which are needed for transcription. These include things such as general transcription factors, shared activating transcription factors and RNAP II (RNA polymerase II). Some of these factors may only be present in limited amounts and if a stronger promoter is

better at recruiting these then there will be less left for a weaker promoter and transcription from this promoter may be lower than it would be when it is present by itself.

Although all the evidence so far points to the CMV promoter being the most interfering promoter of the three promoters used in this chapter we should also consider that all the vectors are not the same length (in bp), meaning that although the same amount in micrograms of each vector has been used in all related experiments if the co-expressed SEAP and GFP vectors differ in size there will not be equal numbers in terms of DNA vector copy number. This will have an effect on competition between two vector/promoters. In the next section this is taken into account and the GFP and SEAP vectors are co-transfected at a 1:1 ratio in terms of DNA vector copy number.

Table 4.4 – Summary of the fold change in the estimated K_m values from the results in figures 4.6, 4.7 and 4.8

	SEAP vectors		
GFP vectors	SV40-SEAP	eEF1a-SEAP	CMV-SEAP
SV40-GFP	+1.35	+1.33	+1.02
eEF1a-GFP	+2.59	+2.39	+1.20
CMV-GFP	+3.19	+3.08	+1.77

The numbers in the table represent the fold change in estimated K_m values compared to the control sample of the respective SEAP vectors. The fold changes come from the data for figures 4.6 (SV40-SEAP), 4.7 (EF1α-SEAP) and 4.8 (CMV-SEAP).

Table 4.5 – Estimated K_i values from the results in figures 4.6, 4.7 and 4.8

	SEAP vectors		
GFP vectors	SV40-SEAP	eEF1a-SEAP	CMV-SEAP
SV40-GFP	0.71	0.75	10.35
eEF1a-GFP	0.16	0.18	1.27
CMV-GFP	0.11	0.12	0.32

Estimated K_i values from the results presented in figures 4.6 (SV40-SEAP), 4.7 (EF1α-SEAP) and 4.8 (CMV-SEAP).

4.5.4 – Promoter interference between the SV40, CMV and EF1α promoters: The effect on SEAP expression of a competing promoter when SEAP and GFP vectors utilizing different promoter combinations are co-expressed at a one-to-one ratio in terms of vector copy number

Table 4.6 shows the lengths of the different SEAP and GFP vectors used in this chapter. From the table we can see that the longest GFP and SEAP vectors both contain the EF1α promoter and the shortest both contain the CMV promoter. At the end of the last section, we mentioned that the lengths of the vectors may have an impact in terms of the number of copies of each vector being transfected and how this could have an effect on competition between two vectors in the same system. In this section, we account for this and transfect SEAP and GFP vectors at a one-to-one ratio in terms of DNA vector copy number.

The amount in micrograms for 1.66 x 10^{11} copies of each SEAP and GFP vector was worked out by rearranging the following equation :-

Equation 5.1

$$\text{Number of DNA vector copies} = \frac{(\text{amount of vector (ng)} \times (6.022 \times 10^{23}))}{(\text{length of vector (bp)} \times (1 \times 10^{9}) \times 650)}$$

6.022 x 10^{23} = Avogadro's number, 1 x 10^{9} = conversion to ng and 650 = one mole of bp equals 650 g (the average weight of a single bp equals 650 Daltons)

Using the amounts that were calculated for 1.66 x 10^{11} copies of each vector all the SEAP vectors were co-transfected with every GFP vector as in the previous sections. 48 hours post-transfection media and cell samples were collected for analysis of SEAP protein and mRNA expression. SEAP mRNA was also analysed during this experiment to see what effect a competing vector/promoter had on transcription and not just solely on SEAP protein expression.

Table 4.6 – Length in base pairs (bp) of the SEAP and GFP vectors used in this chapter

SEAP expression vectors		GFP expression vectors	
Vector	Length in bp	Vector	Length in bp
CMV-SEAP	5045	CMV-GFP	4347
SV40-SEAP	5115	SV40-GFP	4409
EF1α-SEAP	5596	EF1α-GFP	4898

Numbers in the table represent the length in base pairs (bp) of the different DNA expression vectors.

4.5.4.1 – SV40 promoter

Figure 4.9 shows the effect on SEAP protein and mRNA expression from the SV40-SEAP vector when co-expressed at a 1:1 ratio, in terms of vector copy number, with SV40-GFP, CMV-GFP and EF1α-GFP. Co-transfection of SV40-GFP had no effect on SEAP protein, while both EF1α-GFP and CMV-GFP had both significantly reduced it compared to the control. CMV-GFP reduced SEAP protein expression slightly more than EF1α-GFP, about ~67% and ~77% compared to the control respectively.

All three GFP vectors significantly reduced SEAP mRNA expression when compared to the control. SV40-GFP reduced it to ~77% of the control, EF1α-GFP reduced it to ~27% and CMV-GFP reduced it to ~19% (figure 4.9). From this we can see that the reductions in SEAP mRNA are much greater but that this is not translated directly to the amount SEAP protein produced.

In comparisons to the previous work in this chapter SEAP expression from the SV40-SEAP vector was again most negatively affected by the co-transfection of the CMV-GFP vector, closely followed by the EF1α-GFP vector and least affected by the SV40-GFP.

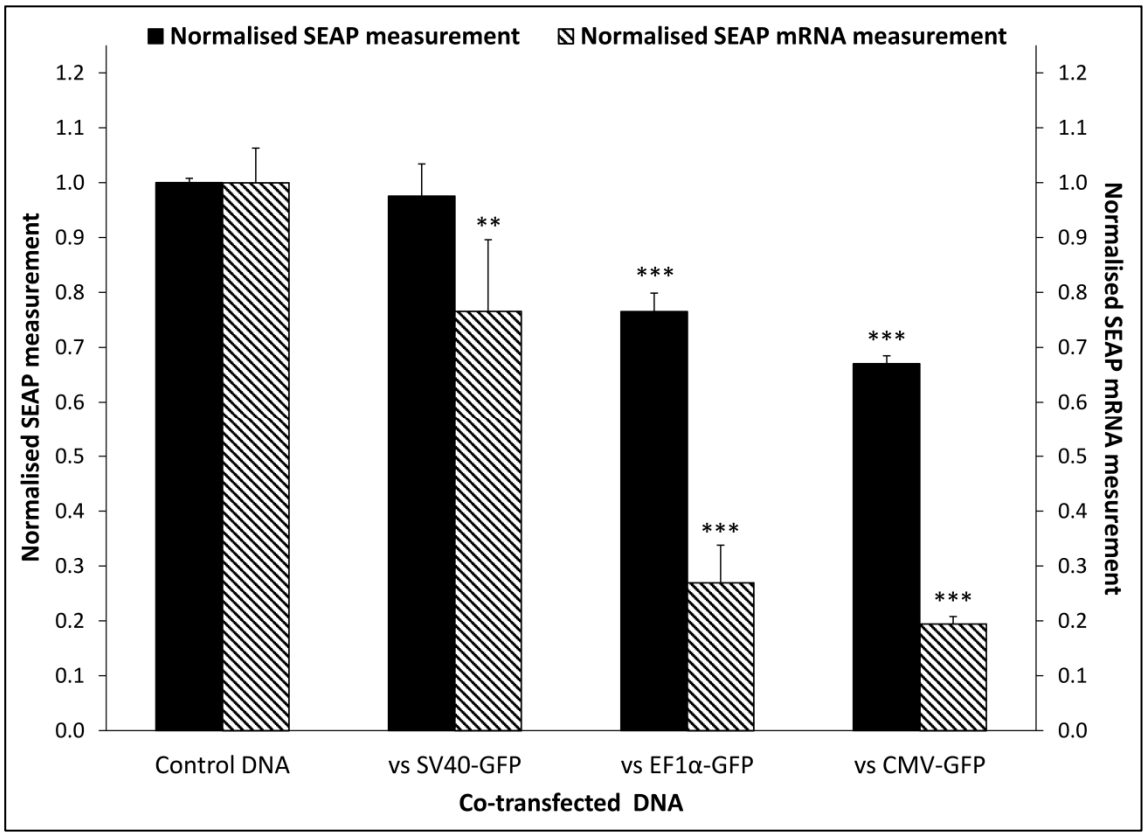

Figure 4.9 – Graph showing the effect on SEAP protein and mRNA expression from the SV40-SEAP vector when co-transfected at a 1:1 ratio (vector copy number) with three different GFP vectors. The effects of co-transfection of SV40, EF1α and CMV-GFP with the SV40-SEAP vector at a 1:1 ratio in CHO-S cells. The 1:1 ratio is in terms of DNA vector copy number rather than total amount of DNA in µg's. 1.66×10^{11} copies of the SEAP and GFP DNA vectors were used. The total amount of transfected DNA was still, however, kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to the control sample. Mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, error bars represent + 1 S.D.

4.5.4.2 – EF1α promoter

The SV40-GFP, CMV-GFP and EF1α-GFP vectors were this time co-transfected at a one-to-one ratio with EF1α-SEAP. Figure 4.10 shows the effects of these GFP vectors on SEAP protein and mRNA expression. Only the CMV-GFP vector caused a significant decrease in SEAP protein expression but this was only a ~8% decrease from the control. Neither GFP vector containing the SV40 or EF1α promoter had any significant effect on SEAP protein expression from the EF1α-SEAP vector.

The results for SEAP mRNA were quite different from the protein results (figure 4.10). This time all three GFP vectors caused a significant and large reduction relative to the protein results. The CMV-GFP vector reduced mRNA expression the most too around 34% of the control. SEAP mRNA was reduced to ~45% by the EF1α-GFP vector and to around 77% of the control for SV40-GFP. This correlates with the results for SV40-SEAP (figure 4.9) in that the order of GFP vectors effecting SEAP mRNA expression the greatest was the CMV then EF1α and then the SV40-GFP vector. Also it showed again that the effects on SEAP mRNA are not necessarily translated to the protein expression level.

Compared to the SV40-SEAP vector (figure 4.9) all GFP vectors had a less negative effect on SEAP protein/mRNA expression from the EF1α-SEAP vector. This correlates with the rest of the work carried out previously in this chapter.

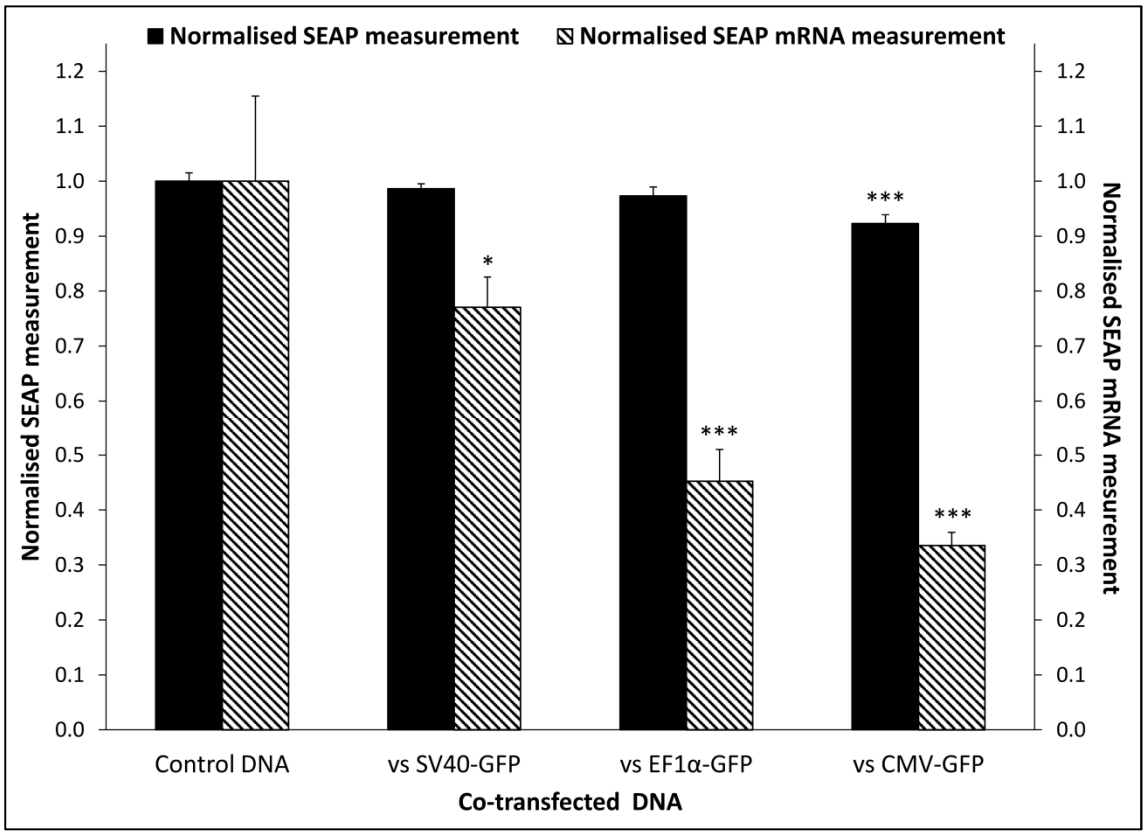

Figure 4.10 – Graph showing the effect on SEAP protein and mRNA expression from the EF1α-SEAP vector when co-transfected at a 1:1 ratio (vector copy number) with three different GFP vectors. The effects of co-transfection of SV40, EF1α and CMV-GFP with the EF1α-SEAP vector at a 1:1 ratio in CHO-S cells. The 1:1 ratio is in terms of DNA vector copy number rather than total amount of DNA in µg's. 1.66×10^{11} copies of the SEAP and GFP DNA vectors were used. The total amount of transfected DNA was still, however, kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to the control sample. Mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, error bars represent + 1 S.D.

4.5.4.3 – CMV promoter

Again the three GFP vectors utilizing the SV40, CMV and EF1α promoters were co-transfected at a one-to-one ratio but this time with the CMV-SEAP DNA vector. The results are shown in figure 4.11. We can see that none of the GFP vectors had any significant effect on SEAP protein expression.

When looking at the results for SEAP mRNA, the SV40-GFP vector had no negative effect on mRNA expression from the CMV-SEAP vector (figure 4.11). The CMV-GFP vector did cause a significant decrease in SEAP mRNA and reduced it to ~66% of the control. When the EF1α-GFP vector was co-transfected SEAP mRNA was reduced, although it was not found to be statistically significant, to around 87% of the control. Again this is similar to the results for the other two SEAP vector in so much as the SV40-GFP vector has the least negative effect on SEAP mRNA expression and the CMV-GFP vector the most. We can also see once again that reductions in SEAP mRNA were not translated to the protein level.

When compared to the results for both SV40-SEAP (figure 4.9) and EF1α-SEAP (figure 4.10) we can see that the effects of a co-transfected GFP vector had less of a negative impact on SEAP expression from the CMV-SEAP vector especially in regards to SEAP mRNA (figure 4.11).

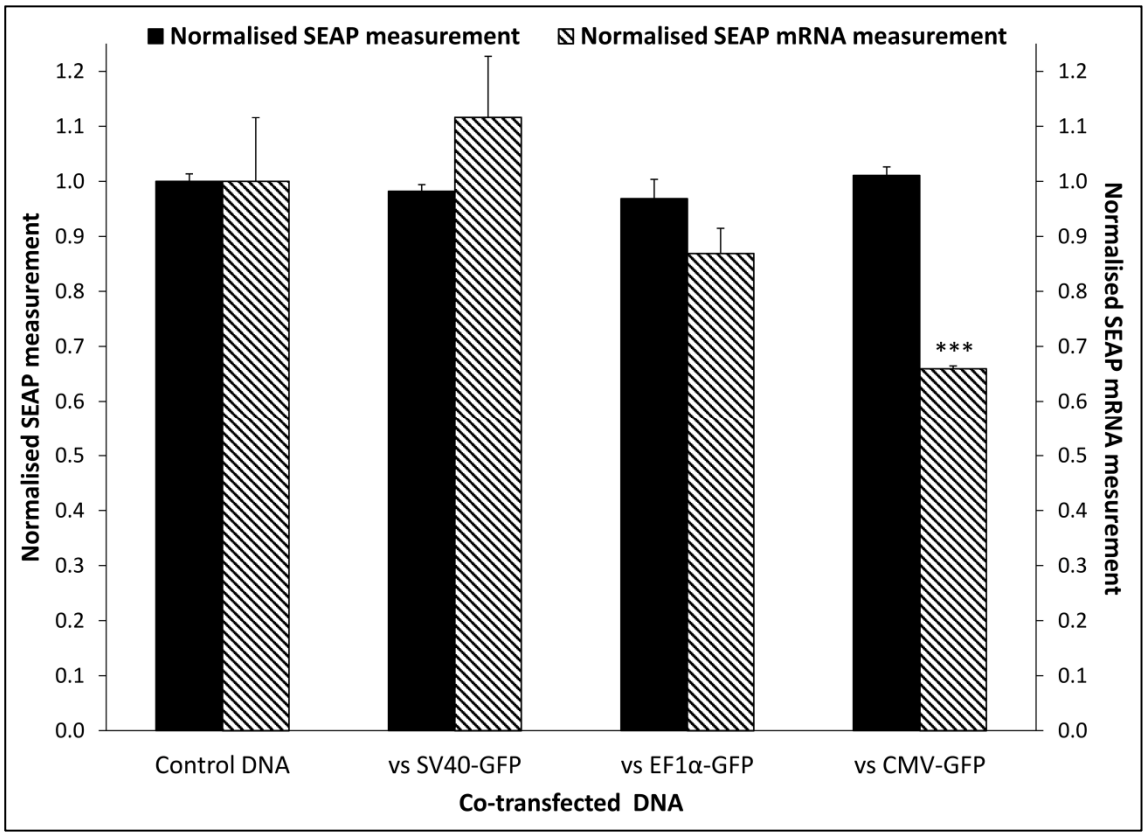

Figure 4.11 – Graph showing the effect on SEAP protein and mRNA expression from the CMV-SEAP vector when co-transfected at a 1:1 ratio (vector copy number) with three different GFP vectors. The effects of co-transfection of SV40, EF1α and CMV-GFP with the CMV-SEAP vector at a 1:1 ratio in CHO-S cells. The 1:1 ratio is in terms of DNA vector copy number rather than total amount of DNA in µg's. 1.66×10^{11} copies of the SEAP and GFP DNA vectors were used. The total amount of transfected DNA was still, however, kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP protein expression. All results were normalised to the control sample. Mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, error bars represent + 1 S.D.

4.5.4.4 – Comparisons and conclusions

Table 4.7 collates the results of figure 4.9, 4.10 and 4.11 and shows them as a fold decrease from their respective control. This again shows how in this experiment that the GFP vector containing the CMV promoter had the most negative effect on both SEAP protein and mRNA expression from all the SEAP vectors while the SV40-GFP vector had the least effect. The CMV-GFP vector also had a much greater effect on expression from the SV40-SEAP vector then on the other two SEAP vectors as well as having a greater effect than the other two GFP vectors. This pattern of negative interference has been seen in the other sections of this chapter.

Interestingly, apart from the SV40-SEAP vector the effects on SEAP protein expression were not as great as was expected. Figure 4.12 and table 4.8 show side-by-side comparisons of SEAP protein expression from figures 4.3-4.5 when SEAP and GFP vectors were co-transfected in equal amounts of DNA in terms of mass compared with those from figures 4.9-4.11 when they were co-transfected in equal amounts (equimolar) in terms of vector copy number. When the SEAP vectors were co-transfected with an equal mass of GFP vectors there was a greater decrease in SEAP expression across all combinations then when compared with the results for equal vector copy numbers. The most probable explanation for this is that for all equal mass co-transfections there was more GFP vector present in terms of vector copy number than SEAP vector (see numbers in the brackets in table 4.8). This is due to all the GFP vectors being smaller in terms of the number of bp's than all SEAP vectors (table 4.6). It would be expected that having more copies of a vector present, it would compete more and have a greater effect on SEAP protein expression. This fits in with the full results from figures 4.3-4.5 and table 4.3, were increasing masses of GFP DNA had increasing effects on SEAP expression. It would be expected that as you change the amount in terms of mass of both SEAP and GFP vectors and also there ratio in terms of vector copy number that the effects of a competing GFP vector on SEAP expression would change. Although we would still expect that a CMV promoter would have a greater negative impact than the other two promoters under comparable conditions.

Even though in figures 4.9-4.11 there were smaller or no decreases in SEAP protein expression compared with figures 4.3-4.5, there were significant decreases in SEAP

mRNA expression. Possible explanations why these decreases in mRNA were not translated into decreases in protein could be that for both the CMV and EF1α promoters there is an excess production of mRNA which is not being actively translated or produces an excess of nascent protein and that the post-translational machinery is limiting the quantity of fully mature protein being produced. Even when the amount of mRNA production from the SEAP vectors utilizing these promoters is reduced it may reduce the amount of excess mRNA rather than the mRNA that is being actively translated or reduce the amount of excess nascent protein. It is has been shown previously that translational/post-translational mechanisms can be become limiting in a transient expression system (Mason *et al.*, 2012) but from our data we cannot say for certain whether this occurred in this case.

Alternatively the SV40 promoter is weaker and produces less mRNA and possibly produces either no excess or a smaller amount of excess mRNA or nascent protein then the other promoters. Its SEAP mRNA expression is also more negatively affected by the competing GFP vectors and therefore, it is more likely that the amount of SEAP mRNA available for translation is decreased and becomes a limiting factor for SEAP protein expression when compared to the other promoters.

In summary, from this section we can see that once again, as in other sections in this chapter that the GFP vectors containing the CMV promoter has the most negative effects on a competing vector and that SV40-GFP had the least. Expression from the SV40-SEAP was the most negatively affected by all GFP vectors and CMV-SEAP the least affected. Changes in SEAP protein expression were not as large as those seen in other sections in this chapter however decreases in SEAP mRNA expression did follow a similar pattern to the previous SEAP protein results.

Table 4.7 – Fold decrease in SEAP protein and mRNA expression when SEAP and GFP vectors are co-transfected at a one-to-one ratio in terms of DNA vector copy number

	SV40-SEAP		EF1α-SEAP		CMV-SEAP	
GFP vector	Protein	mRNA	Protein	mRNA	Protein	mRNA
SV40	-	-1.31	-	-1.30	-	-
EF1α	-1.31	-3.71	-	-2.21	-	-1.15
CMV	-1.49	-5.15	-1.08	-2.98	-	-1.52

The numbers in the table represent the fold decrease in SEAP protein and mRNA expression when compared to the control for the respective SEAP vector. SEAP and GFP vectors were co-transfected at a one-to-one ratio in terms of DNA vector copy number (1.66×10^{11} copies of each vector). Fold changes come from the data in figures 4.9 (SV40-SEAP), 4.10 (EF1α-SEAP) and 4.11 (CMV-SEAP)..

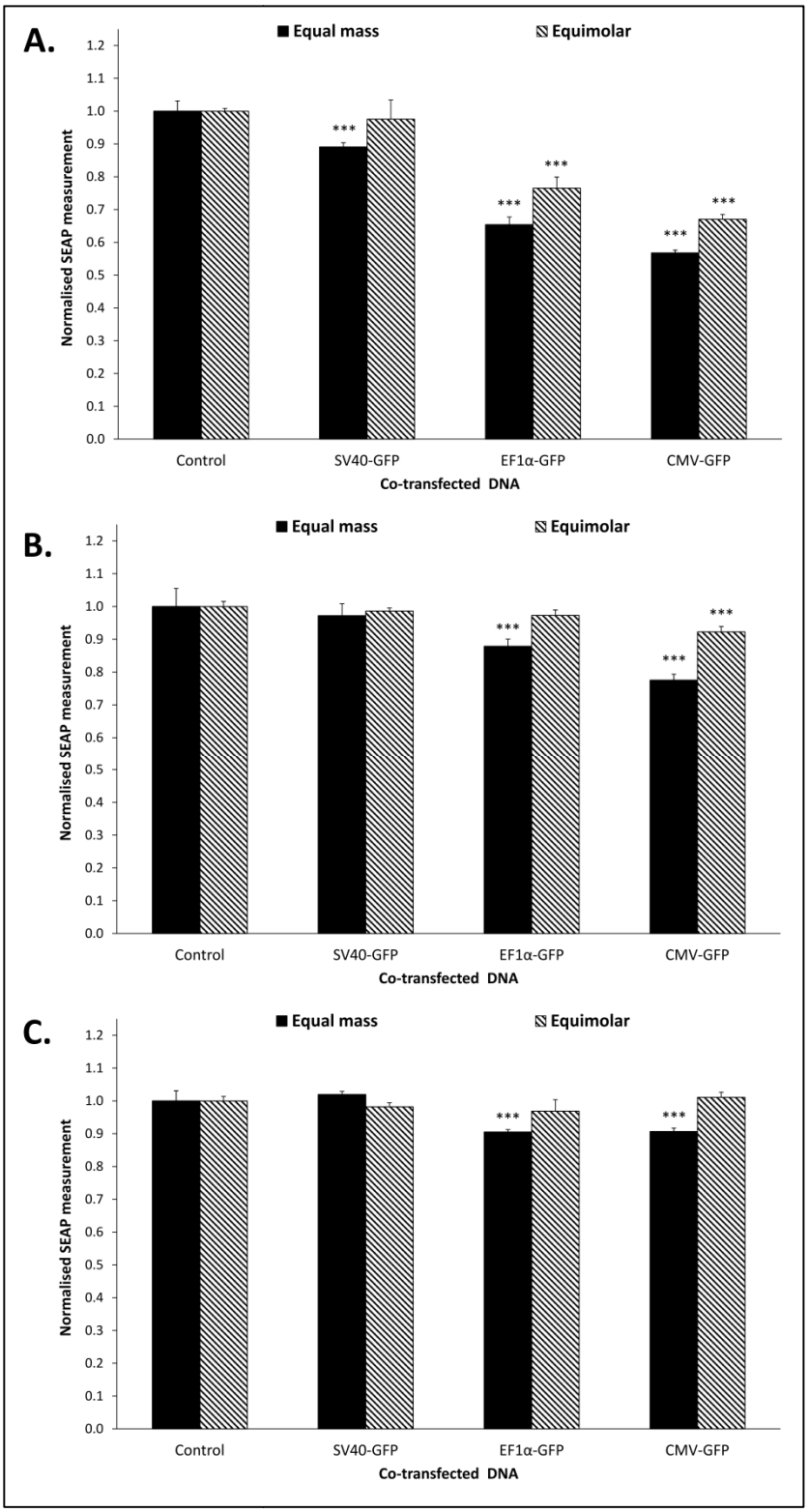

Figure 4.12 – Side-by-side comparison of the effects on SEAP expression when the three SEAP vectors are each co-transfected at a one-to-one ratio with the three GFP vectors in terms of both the amount of DNA in mass and vector copy number. A. SV40-SEAP. **B.** EF1α-SEAP**. C.** CMV-SEAP**.** Equal mass relates to the results in section 4.5.2, figures 4.3-4.5 and shows SEAP expression when 0.0625 µg of both the SEAP and GFP vectors were co-transfected. Equimolar relates to the results in section 4.5.4, figures 4.9-4.11 and shows SEAP expression when 1.66 x 10^{11} copies of both the SEAP and GFP vectors were co-transfected. All results were normalised to the control sample. Mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, error bars represent + 1 S.D.

Table 4.8 – Side-by-side comparison of the fold change in SEAP expression when the three SEAP vectors are each co-transfected at a one-to-one ratio with the three GFP vectors in terms of both the amount of DNA in mass and vector copy number

	SV40-SEAP		EF1α-SEAP		CMV-SEAP	
GFP vector	Equal mass	Equi-molar	Equal mass	Equi-molar	Equal mass	Equi-molar
SV40	-1.12 (1.16)	-	- (1.27)	-	- (1.14)	-
EF1α	-1.53 (1.04)	-1.31	-1.14 (1.14)	-	-1.11 (1.03)	-
CMV	-1.76 (1.18)	-1.49	-1.29 (1.29)	-1.08	-1.10 (1.16)	-

The numbers in the table represent the fold decrease in SEAP protein expression when compared to the control for the respective SEAP vector. Fold changes come from the data collated in figure 4.12 and from figures 4.3 & 4.9 (SV40-SEAP), 4.4 & 4.10 (EF1α-SEAP) and 4.5 & 4.11 (CMV-SEAP). SEAP and GFP vectors were co-transfected at a one-to-one ratio in terms of both DNA mass (0.0625 µg – equal mass)) and vector copy number (1.66×10^{11} copies of each vector - equimolar). For equal mass the numbers in the brackets represent how much extra GFP vector was co-transfected in terms of DNA vector copy number. In all equal mass instances more GFP vector was transfected than SEAP vector due to the all GFP vectors being smaller in terms of the number of bp's than the all the SEAP vectors.

4.6 – Further Discussion, Final Conclusions and Future Directions

4.6.1 – Further discussion

The results at the beginning of this chapter (figure 4.2 & table 4.2) looked at the relative strengths of the three promoters used. We saw that both SEAP vectors utilizing the CMV and EF1α promoters produce similar amounts of SEAP protein across a range of transfected DNA amounts. This was greater than was seen for a SEAP vector using the SV40 promoter. The SEAP mRNA expression for the same range of transfected DNA amounts painted a slightly different picture in so much that the SEAP vectors using the CMV and EF1α promoters were not as equal anymore with the vector containing a CMV promoter producing more SEAP mRNA for the majority of transfected DNA amounts. The SV40 promoter vector produced lower amounts of SEAP mRNA then both the other two promoters. SEAP mRNA expression had a much more linear relationship across the whole range of increasing DNA amounts then SEAP vectors using the CMV and EF1α promoters. The differences in SEAP mRNA expression seen between the three different promoters are not reflected completely by the levels of SEAP protein expression. This shows that increases in mRNA are not always followed by equal increases in expression of the protein. This could possibly be due to the capacity of these CHO-S cells to produce more mRNA being greater than their capacity to produce more protein when both systems are pushed to their functional limits.

Another possible reason that the differences in SEAP mRNA between the different promoters were not reflected at the protein level could be differences in the mRNA transcripts that they each produce. While each promoter produces mRNA that is translated into SEAP protein and all three mRNAs will have the same Kozak sequence (important in initiation of translation) and polyadenylation signal (3' end of the mRNA), all of which are important for further processing of mRNA, their 5' ends do differ. The 5' end of eukaryotic mRNA usually has a 5' cap (RNAP II transcripts) which is important for creating mature mRNA as well as nuclear export, prevention of mRNA degradation, 5' proximal intron excision and promotion of translation (Cowling, 2010; Shuman, 2002; Shatkin & Manley, 2000). As mentioned earlier the EF1α promoter produces

mRNA with a 5' TOP element at the 5' cap and therefore can be called a TOP mRNA (Shibui-Nihei *et al.*, 2003; Avni *et al.*, 1997; Uetsuki *et al.*, 1989). The translation of TOP mRNA is thought to be somewhat dependent on the growth state of the cell. TOP mRNA, unlike non-TOP mRNA, show an association with sub-polysomes in non-growing cells and polysomes in growing cells, meaning TOP mRNA is more actively translated in growing cells (Hamilton *et al.*, 2006). Therefore SEAP mRNA from the EF1α-SEAP vector will be processed differently by the cell when compared to SEAP mRNA from the SV40 and CMV-SEAP vectors. This may possibly explain why the CMV-SEAP and EF1α-SEAP vectors produced similar levels of SEAP protein even though their levels of SEAP mRNA expression differed.

Even though both the SV40 and CMV promoters do not produce TOP mRNA the 5' ends of the SEAP mRNA they produce will still differ from each other. The information sheet for pSEAP2-Control (Clontech, Mountain View, CA, US) states there are four major transcriptional start points for the SV40 promoter (Khalili *et al.*, 1986) in this vector while information for the CMV promoter (GenBank: X03922.1) only suggests one major transcriptional start point. This will have an effect on the mRNA produced with the SV40 promoter possibly producing multiple mRNA with slightly different 5' sequences and the CMV promoter possibly producing SEAP mRNA with a more consistent 5' sequence. This might have an effect on both transcription and processing of the mRNA produced from the SV40 and CMV promoters and why the larger differences in SEAP mRNA expression between the two promoters is not reflected in differences in SEAP protein expression.

Looking at the data from figure 4.2, it can said that the CMV promoter is the most transcriptionally active promoter due to producing more SEAP mRNA across the whole range of transfected SEAP vector amounts then the other promoters. However, its mRNA might be said to be less translationally active then the others. To clarify, both the SEAP vectors utilizing the SV40 and EF1α promoters, produced more SEAP protein compared to the amount of SEAP mRNA they produced when compared to the CMV-SEAP vector. This can be seen in table 4.2 were differences in SEAP mRNA expression for both SV40 and EF1α compared to CMV are much greater than the respective SEAP

protein expression. Again this could be down to differences in the 5' structure of their respective SEAP mRNA.

In summary the three different promoters could be ranked accordingly:-

- SEAP protein expression – CMV = EF1α > SV40
- SEAP mRNA expression – CMV > EF1α > SV40

This work is in agreement with previously published literature that both the CMV and EFα promoters are stronger than the SV40 promoter in many cell lines (Qin *et al.*, 2010; Zarrin *et al.*, 1999; Liu *et al.*, 1997; Thompson *et al.*, 1990; Foecking & Hofstetter, 1986). However, a comparison of these three specific promoters expression in the industrially relevant CHO-S cell line has not been published before.

The rest of the chapter looked at the influence of having two expression vectors and therefore two promoters in the same system, potential interference between the two promoters and the effects on SEAP expression. Interestingly, a similar pattern to that seen for the transcriptional strength of the promoters emerged when looking at the negative effects each promoter had on the expression of a competing promoter. This pattern was that the GFP vector containing the CMV promoter had the bigger negative influence on SEAP expression then any of the other GFP vectors and this influence was seen for all three of the SEAP vectors tested. The EF1α-GFP vector had the second biggest negative influence while the SV40-GFP the least. This was shown at both the SEAP protein and mRNA level.

Looking at the results of all the promoter interference experiments (figures 4.3 – 4.12 & table 4.3 – 4.8) together the SEAP expression from the vector containing the SV40 promoter was the most affected by the co-expression of a GFP vector and a competing promoter in all the experiments. The EF1α-SEAP was the second most effected and the CMV-SEAP vector was the least affected by a competing vector/promoter. This would suggest that the strength of a promoter in our case also had an effect on how much it is affected by the presence of another promoter.

However, it is impossible to separate the effects of a certain promoter driving SEAP expression being affected by the presence of a certain promoter driving GFP

expression whose expression will also be affected. The interference that will occur between different promoters is not unidirectional as has already been said early during this chapter. Interference between the SEAP and GFP vectors will occur both ways. So although the CMV-SEAP vector may seem to be less affected by co-expression of competing GFP vector it might be more correct to say that the CMV-SEAP is better at competing with any GFP vector containing any promoter for its own expression but that this does vary depending on the competing promoter. In fact in our case it may be more appropriate to call this "perturbation of one transcriptional unit by another" (Huliak *et al.*, 2012; Curtin *et al.*, 2008; Eszterhas *et al.*, 2002) "promoter competition" rather than promoter interference. Promoter competition suggests more of a bi-directional effect then promoter interference does.

In terms of the effects seen with the different promoters it was not unexpected or surprising that the SV40 promoter was the worst promoter at competing for its own expression and negatively affecting expression from another promoter due to being clearly the weakest of the three studied. What was interesting was the difference seen between the CMV and EF1α promoters which were shown to be more clearly matched in terms of strength shown by both SEAP protein and mRNA expression. When these promoters were used to drive GFP expression they had similar effects on SEAP expression from both the SV40-SEAP and EF1α-SEAP vector, although the CMV-GFP vector always had a slightly greater effect, but when co-expressed with the CMV-SEAP vector there affects differed significantly. The GFP vector containing the EF1α promoter had less of negative impact then CMV-GFP in all the interference/competition experiments. This suggests that although their differences in transcriptional strength may not be great, there is something about the CMV promoter that does allow it to out-compete other promoters and maintain higher levels of expression. This is best shown in section 4.5.4 were the changes in SEAP mRNA were also shown alongside the changes in SEAP protein expression. Although changes in SEAP protein were not as great as expected, this will discussed again later, for both the SV40-SEAP and EF1α-SEAP vectors there were significant reductions in SEAP mRNA when co-expressed with any competing GFP vector. Even the SV40-GFP vector reduced SEAP mRNA compared to the controls for both these SEAP vectors. However

when looking at the results for the CMV-SEAP vector the SV40-GFP had no negative effect while the EF1α-GFP vector only caused a small decrease in SEAP mRNA expression. Only another CMV promoter significantly reduced the mRNA expression from the CMV-SEAP vector. There must be something about the regulation of the CMV promoter and its structure in terms of binding sites for factors involved in transcription which makes it capable of out competing the other two promoters for expression. The fact that the CMV promoter has the ability to compete well with other promoters makes sense when you consider its origin is viral. It would be advantageous of a virus to be able to out-compete not only endogenous promoters such as the EF1α for expression but also other viruses such as the SV40. Exactly what would give it this ability can only be hypothesized in terms of the results in this chapter. It could be through a direct competition for general transcription factors and transcriptional machinery or competition for other transcription factors that can bind more than one of these promoters (this includes competition for factors that positively affect expression from a CMV promoter as well as sequestering transcription factors which may not directly enhance CMV expression but are required by other the promoters).

So in terms of the level of potential promoter interference/competition the promoters can be ranked accordingly:-

- Promoter interference/competition – CMV > EF1α > SV40

In section 4.5.4 all combinations of SEAP and GFP vectors were co-transfected together but at roughly a one-to-one ratio in terms of DNA vector copy number. This was done to help normalise for the effects of the different sizes of the vectors and potentially make competition between two vectors and therefore two promoters initially more equal. As well as measuring SEAP protein expression SEAP mRNA was also quantified to see what effects a competing promoter had on SEAP mRNA expression. As in previous sections in this chapter we expected to see reductions in SEAP protein expression especially in the presence of the CMV and EF1α-GFP vectors. Although this was observed for the SV40-SEAP vector (figure 4.9), little to no effect on SEAP protein expression was observed for the EF1α-SEAP (figure 4.10) and CMV-SEAP (figure 4.11). Fortunately mRNA samples were also taken which did show when quantified for SEAP

mRNA expression that although levels of SEAP protein did not change significantly the levels of SEAP mRNA did decrease significantly. This matched what we had previously shown (Promoter interference/competition – CMV > EF1α > SV40) and also that competition between promoters occurs at the level of transcription.

Another thing this showed was that the cells in our experiments were capable of producing similar levels of SEAP from the same vector even when the amount of SEAP mRNA was reduced by a competing promoter (figure 4.9 – SV40-SEAP + SV40-GFP, figure 4.10 – EF1α-SEAP + SV40-GFP or EF1α+GFP and figure 4.11 – CMV-SEAP + CMV-GFP). This would seem to suggest that there is an excess or surplus of SEAP mRNA compared to how much mRNA can be actively translated into fully mature SEAP protein. This seems best demonstrated by the difference seen between the SV40-SEAP and the other two SEAP vectors. The SV40-SEAP vector was already shown to produce significantly less SEAP mRNA compared to the other two SEAP vectors and between 0.9 – 1.0 μg of SEAP DNA was transfected depending on the size of vector in section 4.5.4. In figure 4.2 – B. we saw that for 1.0 μg of SV40-SEAP DNA the expression of SEAP mRNA was significantly less than 50% of that produced from the other two SEAP vectors but SEAP protein expression was ~90%. This shows that extra SEAP mRNA is not all translated into SEAP protein and that possibly there is in fact quite a substantial excess. Only when the amount of SEAP mRNA was significantly reduced compared to its control did the level of SEAP protein begin to drop. For the SV40-SEAP vector this occurred when the amount of SEAP mRNA was reduced to 20 – 30% of the control by the CMV and EF1α-GFP vectors. This is likely to also be dependent on the promoter used. Although CMV-GFP reduced EF1α-SEAP mRNA expression to ~30% of the control and only a small but significant drop in SEAP protein was observed and this is probably because of the large amounts SEAP mRNA produced by the EF1α promoter meant even a reduction of this size did not significantly affect the amount of SEAP protein it was capable of producing.

An interesting final point to make is that this promoter interference/competition may already be occurring unwittingly and to the benefit of stable CHO cell lines which have already been and are still being created to produce a therapeutic protein. The CMV and EF1α promoters are routinely used to drive the expression of a therapeutic protein

while the SV40 promoter is utilized in the same vector to drive the expression of a selection gene. If promoter interference/competition is occurring between promoters in this situation, which it is likely that will, then it is adding further selection pressure and possibly contributing to the selection of high producing CHO cell clones.

4.6.2 – Final conclusions

In this chapter the relative strengths of three different promoters, SV40, CMV and EF1α, to produce SEAP mRNA and protein over range of transfected DNA amounts in CHO-S cells were shown. In terms of the relative amounts of SEAP protein and mRNA these promoters were ranked accordingly:-

- SEAP mRNA expression – CMV > EF1α > SV40
- SEAP protein expression – CMV = EF1α > SV40

This is in agreement with previously published work for other cell lines (Qin *et al.*, 2010; Zarrin *et al.*, 1999; Liu *et al.*, 1997; Thompson *et al.*, 1990; Foecking & Hofstetter, 1986). This also shows that both the CMV and EF1α promoters can be used for similarly high levels of recombinant protein expression in CHO-S cells.

It was also showed via co-transfections of SEAP and GFP vectors that promoter interference/competition can occur between two physically unlinked promoters. The level of competition or interference seemed to be related to the strengths of the two promoters present. SEAP protein expression was most negatively affected when a GFP vector containing a CMV promoter was co-transfected and least negatively affected by co-transfection of an SV40-GFP vector. Reductions in SEAP expression were greater for the SV40-SEAP vector and the least for the CMV-SEAP vector.

- Level of promoter interference/competition inflicted – CMV-GFP > EF1α-GFP > SV40-GFP
- Level of reduction in SEAP protein seen – SV40-SEAP > EF1α-SEAP > CMV-SEAP

Reductions in SEAP mRNA were also seen which showed the same pattern of promoter competition and showed that some competition must be occurring at the

transcriptional stage. However, this does not rule out competition at other stages of SEAP production as well.

Although the strengths of the EF1α and CMV promoters did not differ greatly the promoter competition effects did. The CMV was shown to be stronger at competing with another promoter but the reasons for this still need to be further investigated (more information on this is provided in the next section and the next chapter).

Finally these results show why the CMV and EF1α promoters and the SV40 promoter are used for the expression of therapeutic proteins and selection genes respectively but also why they might not be ideally suited for expression in cell engineering strategies. All three promoters are constitutively expressed but their expression can be effected by not only the presence of different promoters but also by the presence of the same promoter and this will affect our ability to tightly control the expression of multiple genes in the same system. For future progress of cell line engineering strategies which could be used for the improvement of not only CHO cells lines producing therapeutic proteins but also in other areas of biological research there will be a need for promoters which can produce differing levels of expression and which function somewhat independently and do not overly interfere with each other's expression. To be able to achieve this we will first need a better understanding of why two physically unlinked promoters interfere or compete for expression.

4.6.3 – Future directions

Future directions should include further investigation into what gives promoters their strength and why promoters interfere and compete with each other. In the next chapter this will be investigated by taking a bioinformatics approach to studying the promoters used in this chapter. A collection of programs will be used (Genomatix Software Suite, Genomatix Software, Munich, Germany) to analyse the potential transcription factor binding sites (TFBS) present in the promoters. A comparison between the promoters may give clues to the regulation of the different promoters and why they interfere with expression of a competing promoter.

It is also possible that this information can be used to help design promoters which not only differ in terms of strength but that can possibly act independently and not overly effect the expression of another competing promoter. This would not only be extremely useful for future cell line engineering strategies in biopharmaceutical production were control of multiple genes may be necessary but also for other applications and also in many other areas of biotechnology and biology.

Further research also needs to be done to investigate how promoter interference/competition actually occurs. Is it just competition for general transcription factors, other more specific transcription factors, and transcriptional machinery or is it more complex then this? Promoter interference can occur between promoters that are physically unlinked and this suggests that there is some kind of competition for shared transcriptional resources but without further research this cannot be said for certain and other possible explanations cannot be ruled out. The design of synthetic promoters as well as the use of decoy DNAs (Brown *et al.*, 2013) will be useful for gaining a further understanding of how this interference/competition occurs.

Another area of research which future developments in would help us to be better able to investigate and understand this phenomenon would be how vector DNA is processed and regulated once it reaches the nucleus. Although a large amount of research has been done on gene expression from chromosomes little has been done on the nuclear localisation and expression of DNA vectors which are not integrated into the genome (Mearini *et al.*, 2004). Understanding this would be extremely useful for many applications of transient gene expression.

Chapter 5 – Transcription Factor Binding Site Analysis of Three Commonly used Constitutive Promoters for Biopharmaceutical Production in CHO Cells

5.1 – Abstract

Previously we have shown how the three promoters, SV40 early, CMV major immediate-early and the human EF1α promoters interact when co-expressed and the difference in promoter interference/competition observed. The level of promoter interference was reflected in the strengths of the promoters.

The strengths of the promoters will reflect the transcriptional activators (transcription factors (TFs)) which can bind each promoter and one hypothesis for a potential cause of promoter interference is the competition for shared TFs.

The promoter sequences were analysed using the programs MatInspector and ModelInspector. Both programs showed the SV40 to have less potential transcription factor binding sites (TFBSs) and also potentially less variety in the families of TFs which it could bind when compared to the CMV and EF1α promoters. The SV40 promoter also shared a greater proportion of these TFBSs with both the CMV and EF1α promoters. Results from the ModelInspector search of the CMV promoter showed it to have less potentially functional TFBSs in common with the other two promoters. TFBSs identified by ModelInspector imply more potential functionality then those identified by MatInspector. These differences might be a reflection of both their transcriptional strengths and their ability to interfere with expression from another promoter.

Potentially functional TFBSs binding sites from the SP1 family were found to be overrepresented in all three promoters and were thought to be one potential source of promoter interference/competition. Over-expression of the TF SP1 from a vector

utilizing the CMV promoter rescued expression from SV40, CMV and EF1α-SEAP vectors and from SV40 and CMV-GFP vectors compared to co-expression of CMV-GFP and CMV-SEAP respectively.

The presence of high numbers of potential TFBSs in each promoter, although a large proportion will not be functional, highlighted that promoter interference is unlikely to occur due to competition for single TF or even just TFs alone and also the potential complexity in the regulation of expression from each promoter. Trying to discretely control expression of multiple genes using these promoters is unlikely to be viable option without modification of their sequences which could result in unexpected effects. A 'bottom up' approach for the creation of synthetic promoters is likely to be a better strategy for in attempts to control expression of multiple genes.

5.2 - Introduction

In the last chapter it was seen that the SV40 early, the CMV major immediate-early and the human EF1α promoters not only differed in expression, but also how they differentially affected the expression of a competing promoter when present in the same system. The negative effect of a competing promoter is termed promoter interference/competition. The CMV promoter had the greatest negative effects when competing with itself, the EF1α and the SV40 promoter. A GFP vector utilizing a CMV promoter was shown to reduce the expression of SEAP from a competing vector more when compared to an EF1α and SV40 promoter. The magnitude of this reduction was dependent on the promoter present in the co-transfected SEAP vector. SEAP expression from the SV40-SEAP vector was reduced the most, followed by the EF1α-SEAP vector and the least from the CMV-SEAP vector. The presence of a GFP vector containing an EF1α promoter had similar effects on SEAP expression from the SV40 and EF1α-SEAP vectors when compared to CMV-GFP, but did not affect SEAP expression from CMV-SEAP to the same extent. The SV40-GFP vector had the least competitive effect and only reduced SEAP expression from the SV40-SEAP vector to any significant extent. This showed that in terms of promoter interference/competition the CMV promoter was the most interfering while the SV40 promoter was the least. This pattern of competition matched the strengths of the promoters used and suggests that the strength of the different promoters used plays a role.

From work in chapter 4 it is not possible to say exactly why the different promoters differed in both strength and in terms of interference/competition. Was it competition for general transcription factors (TF) and transcriptional machinery (RNAPII etc) or was it competition for more specific TFs for which the promoter's share common binding sites? The second hypothesis is investigated in this chapter.

In this chapter the Genomatix Software Suite (Genomatix Software, Munich, Germany) was used to analyse the promoter sequences of these three commonly used constitutive promoters. Although the individual promoters have been studied to some extent in terms of some selected specific transcription factor binding sites (TFBSs) a more detailed analysis and a comparison between the three promoters has not be

carried out previously. Comparison of the results for the promoters is also done to see which common factors they might potentially share. The presence of common TFBSs amongst the three promoters should give clues as to which, if any, TFs the promoters compete with each other for. A TF is also over-expressed which can potentially bind all three promoters to help alleviate promoter interference/competition and show that competition for a possibly limited supply of a shared factor is one of the causes of promoter interference/competition seen in the last chapter.

This might also provide evidence or guidance for how to rationally design promoters which can produce different levels of expression but which do not significantly interact, interfere or compete with each other. In future cell engineering strategies involving the expression of more than one gene being able to somewhat independently and reliably control the expression of multiple genes will be advantage for their successful development.

5.2.1 – The Genomatix Software Suite

The Genomatix Software Suite is a collection of online programs for the analysis of genomic data as well as gene expression and gene regulation. For the purposes of this chapter, a set of sequence analysis programs which analysed the promoters used in the previous chapter for potential TFBSs, TFBS modules, over-represented TFBSs and also common TFBSs shared between the three different promoters studied. The next sections we briefly describe the programs that were used but it should be pointed out that although these programs can be quite powerful, the TFBSs they identify are only potential TFBSs and they do not prove any functionality. The program ModelInspector is better at identifying functional modules but still has its limitations. However, the results of such searches can be used and are used to help guide further experiments.

5.2.1.1 – MatInspector

MatInspector was the main analysis program that was used for identifying potential TFBSs and it also forms the basis for the analysis carried out by the other programs that will be described shortly. MatInspector uses a large library of matrix descriptions

for TFBSs to identify matches in DNA sequences (Cartharius *et al.*, 2005; Quandt *et al.*, 1995). It identifies TFBSs using nucleotide or position weight matrices (NWM or PWM) rather than simple IUPAC consensus strings. They have the advantage in that for each single position in a PWM the entire nucleotide distribution can be considered. They also allow for quantification of a similarity score between the potential TFBS and the PWM. Alternative programs also use PWM for identifying potential TFBSs in sequences, but most of these programs have greater limitations then MatInspector. Some are based on old matrix libraries which are no longer updated such as Signal Scan (Prestridge, 1996, 1991). Other programs such as MATCH (Kel *et al.*, 2003) perform searches similar to MatInspector. However, MATCH has two drawbacks. The first being the freely available version only contains 398 matrices due to using the publically available TRANSFAC matrix library. A commercial version using the updated TRANSFAC library is available and is more than six years ahead of the public one and contains >2000 entries, but this is not freely available to academic users. Although the public version of MATCH has a much bigger library than programs such as Signal Scan and ConSite (Sandelin *et al.*, 2004; Prestridge, 1996, 1991) it is still much smaller than the matrix library for MatInspector, 1381 weight matrices (Matrix Family Library Version 9.0 (August 2012)), which is also freely available for public/academic use. The second drawback of the public version of MATCH is that it does not group matrices into matrix families like MatInspector. Although this is not a significantly major problem, it does mean that MATCH can produce redundant matches when the same part of a sequence is identified for having overlapping matches to more than one similar or related matrix. MatInspector overcomes this problem by grouping similar individual matrices into so called matrix families and then when more than one overlapping match from the same matrix family occurs in a sequence only the highest scoring match is returned. This greatly reduces the number of redundant matches returned and gives a more condensed and comprehensive output.

Another advantage of the MatInspector program is its use of optimized matrix thresholds (Cartharius *et al.*, 2005). Rather than having the same matrix similarity threshold for all matrices, which would greatly affect the number of matches returned depending on the matrix length and sequence conservation, each individual matrix has

its own threshold which helps reduce the number of false positives. The optimized threshold of a weight matrix for MatInspector was defined as the matrix similarity score which resulted in only a maximum of three matches per 10,000 bp of non-regulatory test sequences (Cartharius *et al.*, 2005).

As mentioned earlier MatInspector forms a major part of all the following programs we used and each will be briefly described.

5.2.1.2 – Overrepresented TFBSs

This program uses MatInspector to search sequences to identify TFBSs which are overrepresented compared to a defined background. Overrepresentation of a TFBS can sometimes be a sign that it is functionally important. It generates statistics for matrix families including overrepresentation values and Z-scores. This is achieved by comparing the occurrence of the matches in a given sequence to the occurrences seen in a selected background. This can either be the occurrence of matches in the whole genome sequence of a selected species, the occurrence of matches in promoter sequences of a selected species which were annotated by Genomatix or to user a defined background.

The overrepresentation value is calculated from the number of matches found in a sequence compared to the number seen in an equally sized sample from the selected background. Put simply, the overrepresentation value is basically found versus expected.

The Z-score is another way of showing overrepresentation compared to a selected background. It is the number of standard deviations that a given data point is from the population mean. In this case the number of matches found compared to the mean and standard deviation of a given background. This program uses continuity correction to calculate the Z-score using the following formula:-

$$Z = \frac{(x - E - 0.5)}{SD}$$

Where x equals the number of matches found in a sequence, E equals the mean number of expected matches and SD is the standard deviation of E. This formula was described previously in a paper identifying overrepresented TFBS in co-expressed genes (Ho Sui *et al.*, 2005). The Z-score has the advantage over the overrepresentation value given that it can be used to calculate statistical significance. A Z-score value equal or greater than 1.96 shows a TFBS is overrepresented with a corresponding *p* value which is equal or less than 0.05. A Z-score equal to or less than -1.96 shows a TFBS is underrepresented with a *p* value equal or less than 0.05. As with all of these programs it must be remembered that a *p* value represents statistical significance and not necessarily biological significance.

5.2.1.3 – ModelInspector

ModelInspector is used to search sequences for so called pre-defined modules (Klingenhoff *et al.*, 1999; Frech *et al.*, 1997). A model consists of two or more TFBSs which have been shown in the literature to be important in gene regulation from a given promoter/enhancer. Information for a certain model consists of not only the appropriate TFBSs but also the distance these sites are expected to be from each other in terms of bp's and also the orientation of the site (on either the + or – strand).

Like the TFBS matrices the modules also have their own library which is currently version 5.6 and because ModelInspector uses matrices for identifying individual TFBSs this model library is linked with version 9.0 of the matrix family library.

Model matches are only returned if the individual TFBSs present satisfy the requirements of the model. This includes having a sufficient core and matrix similarity score (note that the similarity score required for a model may be different than the score for the optimized matrix threshold), being within the correct distance of each other and also on the correct DNA strands.

Unlike MatInspector and the other programs mentioned in this section ModelInspector is better at giving potential functional information about a given sequence. The other programs just give potential individual TFBSs whose presence alone can mean very little since it is known that TF often work co-operatively to modulate gene expression.

ModelInspector not only identifies possible TFBSs that could interact together but because it is also based on examples taken from the literature were these interactions have been shown to occur and have an effect on gene expression a greater functionality can be inferred. Again caution should be taken with results as they still have major limitations since not all modules can be expected to function the same in not only different species but within different tissues and cells of the same species. However, the results from ModelInspector are still much more functionally informative then the results from MatInspector.

5.2.1.4 – Common TF sites

This program uses MatInspector to simultaneously search multiple sequences for matches for TFBSs which are found to be common to all or a certain percentage of the selected sequences. Results are presented in both a graphical and a summary table form. The summary table lists the name of the TFBS matrix families identified, the number of times it was found in total, the numbers times it was found in each sequence and the number of the sequences it was found in.

It also gives a *p* value for each matrix family identified. This *p* values represents the probability of getting an equal or greater number of sequences with matches when compared a random sample which is the same size as the input sequences. These *p* values were based on already pre-defined promoter matches. These promoter matches are for known promoter sequences which were extracted from Genomatix's ElDorado (http://www.genomatix.de/online_help/help_eldorado/introduction.html) database. *P* values are only correct if the average length of the sequences analysed is around 600 bp and optimized matrix thresholds are used.

5.3 – Aims and Objectives

In this chapter the programs created by Genomatix and described in the previous section are used to analyse the sequences of the three promoters studied in the last chapter (SV40, CMV and EF1α) for potential TFBSs. The results for each promoter will be used to give an idea of the TFs that are potentially important for their regulation. They will also be compared with each other to see if they share TFBSs and this should give clues to why these promoters can negatively affect the expression of a competing promoter with one hypothesis for the promoter interference/competition being there is a competition for shared TFs.

It was also tested whether a TF whose binding site is present in all three promoters can rescue SEAP/GFP expression from three SEAP/GFP vectors even in the presence of a competing CMV promoter which was shown to most negatively affect/decrease expression from another promoter.

The aims and objectives of this chapter are as follows:-

1. Use MatchInspector to analyse the potential TFBSs present in the SV40, CMV and EF1α promoters
    a. Do the total numbers of TFBSs identified in each promoter differ from a random sample of 30 sequences who match each promoter in terms of both size and ACGT content?
    b. Analyse the results for each promoter in more depth including the results for MatInspector and also the identification of overrepresented TFBSs
2. Use ModelInspector to identify potentially functional TFBS modules present in the promoters
    a. This kind of search is more functionally informative then MatInspector
3. Compare the potential TFBSs present in each promoter from the MatInspector and ModelInspector searches to see if they share common sites for TFs as this could be a source of promoter interference/competition

4. Express a TF which each promoter contains a binding site for and which is potentially functional to see the effect on SEAP/GFP expression when a competing promoter is present
    a. This TF will be SP1, more information and why this was selected will become clearer later in this chapter
    b. A vector containing the CDS for SP1 and utilizing a CMV promoter for expression will be co-expressed with the three SEAP and GFP vectors at a one-to-one ratio in terms of vector copy number similar to the experiment in chapter 4, section 4.5.4
    c. SEAP proteins and mRNA expression will both be measured as well as GFP expression

By comparing the potential TFBSs found in each of the SV40, CMV and EF1α promoters it will give a better understanding of the regulation of expression from each and also how/why these promoters interact and compete with each other. This information may also help guide the design of novel synthetic promoters which can function somewhat independently of each other even in the same system.

5.4 – Materials and Methods

5.4.1 – Using the Genomatix Software Suite

All four programs mentioned earlier form part of an extensive collection of online programs for genomic data and gene expression and gene regulation analysis. The homepage can be found at the following web address http://www.genomatix.de/solutions/genomatix-software-suite.html. For the different programs MatInspector, Overrepresented TFBS, ModelInspector and Common TFs, all three promoter sequences are presented to the programs in a FASTA format with the first line denoting the specific promoter (see CD appendix for promoter sequences in FASTA format). Other common formats are also accepted.

On the first page of the MatInspector program you add your sequence. The transcription factor binding sites (weight matrices) option was selected under library selection (this is the default option) before proceeding to the next page. On the next page under matrix search parameters the latest version of the matrix library (Matrix Library 9.0), the vertebrate and general core promoter elements and the use all matrices from selected groups option (you can also select subsets to customise your search) were selected. Under the matrix families parameter, you can select to search using either matrix families or individual matrices. Searches of the promoter sequences were carried out using both options. Core similarity was left at the default 0.75 and for matrix similarity optimized thresholds were used. Further options are available to customise the output given from the search.

For the Overrepresented TFBS program sequences were inputted on the first page along with selecting the following parameters – overrepresentation of single TF (using MatBase), matches to matrix families and both genomic and promoter background. For comparison against a background an organism also needs to be selected and while this should be the organism your sequences came from or are being expressed in the lack of a single sequence origin and the fact *Cricetulus griseus* (Chinese hamster) was not an option *Mus musculus* was selected.

ModelInspector, the first page was simply for inputting your sequences. On the second page the following parameters were used. The latest module library was

selected (Module Library 5.6) as well as vertebrate modules and use all selected modules from this group (again you can also customise this by selecting subsets). Under search parameters, the options were left as the defaults while in the next section the output parameters could be adjusted to customise the final output of your result.

When using the Common TFs program sequences were inputted on the first page as well as selecting transcription factor binding sites (weight matrices) under library selection. The next page was very similar to the second page for MatInspector. The same options were chosen with this time only searches for matrix families being carried out. An extra parameter was also available and this was to select for the percentage of sequences a match had to appear in to be returned in the results.

For all programs the full results or just summaries could be downloaded in different formats. Files were downloaded as Excel or PDF files and all the raw search results can be found in the attached CD appendix in this format.

5.4.2 – Over-expression of the TF SP1 and its effects on promoter interference/competition

This set of experiments were carried out in the exact same manner as those in chapter 4, section 4.5.4. The cell line used, culture method, transfection protocol were all the same as those used in chapter 4 and described in section 4.4. SEAP protein and mRNA as well as GFP expression were analysed as already described in chapter 3 & 4. The SEAP, GFP and negative control DNA vectors used have already been described in chapter 4, section 4.4.2.

The CMV-SP1 vector (SC101137) was purchased from OriGene (Rockville, MD, US).

5.5 – Results and Discussion

5.5.1 – Identification of potential TFBS in the SV40, CMV and EF1α promoters by MatInspector

The program MatInspector was used to analyse the SV40 early, CMV major-immediate early and the human EF1α promoters for the presence of potential TFBSs. All three promoters were analysed using the search parameter matrix families (identifies the best match from a matrix family at a given position in a sequence) and individual matrices (gives all possible matches for a given position in a sequence). Although it is recommended that the matrix family option is used to help prevent the presence of redundant matches, it was thought it would be more informative to first look at the results of searches using both options.

The reasoning being that although the removal of redundant results is helpful at giving a result which is more condensed, presentable and potentially more accurate in terms of the actual potential numbers of TFBSs, the extra information would actually help us study the results better by understanding which potential binding sites had a greater selection in terms of potential TFs to bind to them. Most TFBSs are capable of binding more than a single TF, but by looking at the results for individual matrices this would be shown to some extent by the program and not only just an assumption made by the user. Although it should be mentioned that just because a site for a TF with a similar binding sequence at the same location was not identified using the individual matrices option does not mean a related TF cannot bind there. Some TFs can be quite promiscuous and will sometimes bind to similar sequences even if that particular sequence does not quite meet the optimized matrix threshold for a given TF but the strength of the binding or interaction might differ from a TF whose binding site is present and more conserved. Remember that the individual TF matrices and optimized thresholds have not been defined for all possible biological interactions.

After running the promoter sequences of SV40, CMV and EF1α through MatInspector the raw numbers given by the program in terms of the number of potential TFBSs were studied. Table 5.1 shows the results for each promoter and also the numbers of TFBSs identified using both matrix families and individual matrices. The SV40 promoter had

the least number of potential TFBSs when searching using the matrix family or the individual matrices option, 163 and 335, respectively. The difference in number seen when analysing sequences using either the matrix families or individual matrices options demonstrates how selecting the matrix families reduces the number of TFBSs returned by removing so called redundant matches. For the CMV and EF1α promoters, EF1α had more potential TFBSs (274) when using matrix families compared to the CMV (225). When using individual matrices the CMV promoter had a greater number of hits, 457, than the EF1α promoter, 422.

Studying this and considering what was shown in the previous chapter, it could be true that the number of potential TFBSs identified in a promoter might be a reflection of their relative strengths. The SV40 promoter was the weakest promoter in terms of the strength of expression and from this search, was also shown to have the least number of potential TFBSs. The SV40 promoter had 62 and 111 fewer sites then both CMV and EF1α, respectively, when the matrix families option was selected and 122 and 87 fewer sites when individual matrices was used. For both the CMV and EF1α promoters the numbers of potential sites identified were different but there were less differences than seen for the SV40 promoter. For matrix families, the CMV had 49 fewer sites then the EF1α promoter. Given this and what was said at the start of this paragraph, it could be expected that the EF1α promoter would have stronger expression then the CMV. In chapter 4 this was shown to not be the case. In terms of individual matrices the opposite was seen with the CMV promoter having 35 more sites then EF1α. This might potentially mean that although the EF1α may possibly have more binding sites for a greater number of TFs from different families the TF families that could potentially bind the CMV have a greater number of individual but related TFs within each or in certain families which could bind the CMV. Alternatively some of the TFBS identified for the EF1α promoter might either not be functional or their TF might not be expressed to a significant degree in CHO-S cells. However, the same could be said of any of the promoters and a more in depth look at the TFBSs identified in each promoter is done later in this chapter.

The number of TFBSs identified was compared to the length of the promoters. Interestingly even though the EF1α promoter is 734 (2.6-fold) and 600 (2.0-fold) bp

longer then the SV40 and CMV promoters, it did not have two times or greater numbers of potential TFBSs. This suggests that there are differences in the sequence/structures between viral and endogenous mammalian promoters. Indeed in this case there is. As mentioned in the last chapter when describing the promoters (chapter 4, section 4.2.2) a large part of the EF1α promoter which has been shown to be essential for high expression is the first intron (Wakabayashiito & Nagata, 1994). The first intron is 943 bp and makes up most of the EF1α sequence (79% of 1189 bp). Although it is essential for strong expression it be might expected that an intron contains less TFBS than a core promoter or enhancer might. The SV40 and CMV promoters are mostly made up of a core promoter and enhancer sequences. It might also be expected that viral promoters contain higher numbers of TFBSs in a given length of sequence due to their much smaller genome and, like some mammalian promoters *in vivo*, they can also be used for controlling multiple genes. There is also a need for some viruses to be capable of infecting and thriving in multiple related species and in different cell types which may differ in their TF profiles. On this last point on viral promoters it should also be pointed out that the EF1α gene is expressed in all cell types, except neurons and skeletal and cardiac muscle (Soares & Abbott, 2013), and therefore its promoter, like the viral ones, is capable of functioning in many different cell types with different TF profiles. Alternatively, this could also mean that all three of these promoters rely heavily on ubiquitously expressed TFs.

To look whether these differences between the number of TFBSs found in each promoter and length could be due to their different origin or structures the number of TFBSs for each promoter were compared to the average number of TFBSs found by MatInspector in 30 randomly generated DNA sequences which were matched in both length and ACGT content for each promoter (http://users-birc.au.dk/biopv/php/fabox/random_sequence_generator.php). Figure 5.1 shows for both the SV40 and CMV promoters that in comparison to 30 random DNA sequences they each had more predicted TFBSs sites. On average the SV40 and CMV had ~64 and ~77 more TFBSs for matrix families (figure 5.1 – A.) and ~170 and ~174 more TFBSs for individual matrices (figure 5.1 – B.) when compared to their respective 30 random sequences. Comparison of the EF1α promoter showed that it had roughly the same

number of predicted TFBSs for both matrix families and individual matrices when compared to random sequences which match it in length and ACGT content.

It is interesting that the viral promoters have a greater number of potential TFBS then a representative random sample of DNA sequences while the mammalian EF1α promoter did not show any difference. This would suggest differences in the structure between these two viral and this mammalian promoter. One obvious difference between the viral promoters compared to EF1α is the fact that these promoters are composed of repeated sequences to a greater or lesser extent. The SV40 promoter is almost entirely made up of two copies of three 21 bp repeats and a single copy of two 72 bp repeats (Byrne *et al.*, 1983). While the enhancer of the CMV contains multiple copies (between three and five) of four different repeats (17, 18, 19 and 21 bp repeats) (Fickenscher *et al.*, 1989). This repeated structure could be one reason why the two viral promoters had greater numbers of TFBS then their comparable random sequences while the EF1α was more comparable to its random set of sequences. This is interesting in terms of the design of synthetic promoters because it shows that you can achieve get high and comparable levels of expression from smaller promoters which contain repeated sequences (Schlabach *et al.*, 2010). Although this is interesting the fact that the CMV and EF1α promoters produce similar levels of expression shows there is multiple ways of achieving high levels of recombinant gene expression but the use of smaller promoters may be an advantage when constructing large DNA vectors which contain multiple recombinant genes.

Looking at the results of the MatInspector searches for each promoter in more detail the SV40 promoter had 163 TFBSs identified which was composed of 50 different matrix families (table 5.1). For individual matrices of the 335 TFBSs found this was made up of 129 separate matrices. For the CMV promoter out of the 225 potential TFBSs found this was made up of 77 different matrix families and out of the 457 found searching using individual matrices there were 244 different individual matrices (table 5.1). 274 TFBSs were identified for the EF1α promoter and this was composed of 112 different matrix families (table 5.1). For individual matrices out of the 422 TFBSs identified there were 276 separate individual matrices. This showed that in terms of the number of potential TFs that the EF1α promoter may have a larger selection.

Although, as was shown in figure 5.1, these larger numbers might be a reflection of the EF1α promoters longer sequence rather than actually on functional TFBSs. For any of these promoters, all of the different matrix families or individual matrices identified are only potential TFBSs and it is extremely unlikely that they are all functional. For both the SV40 and CMV promoters the numbers of different matrix families and individual matrices were of a higher proportion of the total numbers of potential TFBSs than for the EF1α promoter (table 5.1). This suggested that the CMV and SV40 promoters rely on some matrix families or individual matrices more than others and also potentially more than the EF1α promoter.

Table 5.1 – Number of potential TFBSs identified by MatInspector for the SV40, CMV and EF1α promoters

	Number of potential TFBSs	
Promoter	Matrix families	Individual matrices
SV40	163 (50 diff matrix families)	335 (129 diff individual matrices)
CMV	225 (77 diff matrix families)	457 (244 diff individual matrices)
EF1α	274 (122 diff matrix families)	422 (276 diff individual matrices)

The numbers in the table represent the number of potential TFBSs identified by MatInspector for each promoter. The numbers in the brackets represent the number of different matrix families or individual matrices found in each search. This was done searching the sequences selecting both the matrix family and individual matrices options in the search parameters of the program. See text in this section for their differences or section 5.2.1.1.

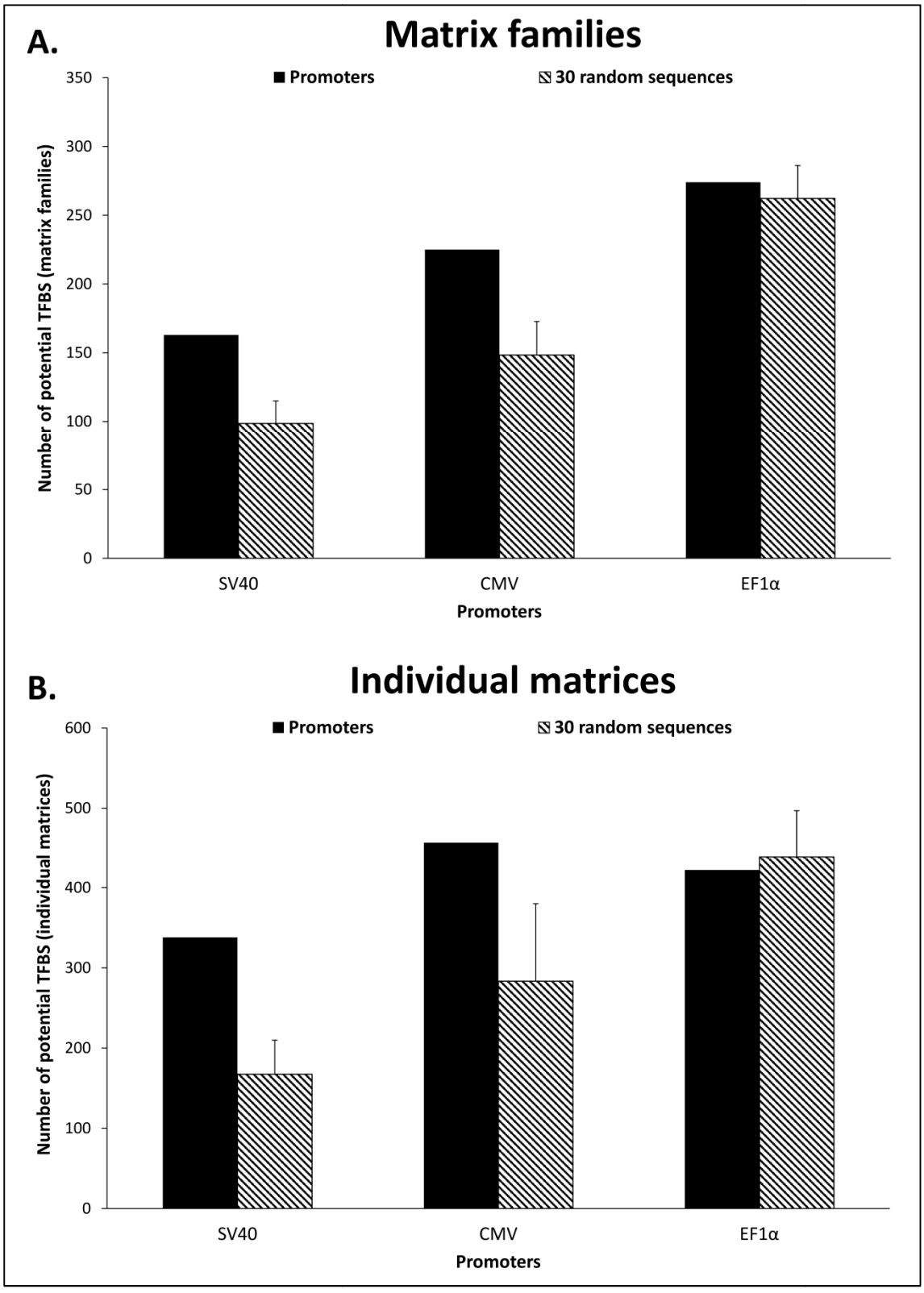

Figure 5.1 – Number of potential TFBSs in the SV40, CMV and EF1α promoters when compared to 30 randomly generated DNA sequences. The number of potential TFBSs identified by MatInspector for SV40, CMV and EF1α promoters were each compared to the average number of TFBS identified by searches of 30 randomly generated DNA sequences which matched each promoter in terms of both length and ACGT content. **A.** Shows the results when matrix families were selected in the search parameters. **B.** Shows the results when individual matrices were selected in the search parameters. Error bars represent + 1 S.D.

Table 5.2 shows the top 20 hits in terms of number of sites found for both matrix families and individual matrices found by MatInspector for the SV40 promoter. The most prevalent matrix family identified was V$KLFS with 13 potential TFBSs. This matrix family is composed of KLF (Kruppel-like factors) TFs. The KLFs are a set of 17 (KLF1-KLF17) zinc finger DNA binding proteins which have a variety of diverse functions from differentiation, proliferation, apoptosis, growth and responding to external stimuli (McConnell & Yang, 2010). Unsurprisingly the next most prevalent matrix family was the V$SP1F which contains SP (specificity protein) TFs. These proteins share homology with the KLFs and are in fact considered as a closely related family, the SP1/KLF family (McConnell & Yang, 2010; Waby *et al.*, 2008). Nine SP factors have been identified so far (SP1-SP9) (Suske *et al.*, 2005) and like the KLFs have a wide range of roles from cell cycle regulation, growth, apoptosis and angiogenesis etc (Archer, 2011). Due to their homology many members of the SP1/KLF family bind GC rich elements and therefore also share similar binding sites (Archer, 2011; McConnell & Yang, 2010).

The third most prevalent matrix families are V$OCT1 and V$HBOX with nine potential sites each. These families include binding sites for OCT1 (octamer binding protein 1) and related OCT TFs and homeobox TFs such as EN2, GBX2 and VAX1. The next most prevalent families with 8 TFBS each are V$HOMF, V$HOXF and V$ZF02. The V$HOMF and V$HOXF are matrix families for further homeobox TFs while the V$ZF02 matrix family contains binding sites for a set of proteins which contain zinc finger DNA binding domains. As was mentioned earlier in total there are 50 different kinds of matrix families present in the SV40 promoter (table 5.1). Without analysing the promoter in a more functional manner by the removal of specific TFBSs it is impossible to say with any certainty which sites are most important for expression from this promoter just using this search.

Table 5.2 also shows the top 20 individual matrices found in the SV40 promoter. This shows the redundancy seen when searching sequences using this option with the top hits being V$SP1.02 and V$GC.01 with 11 and both having very similar binding sites and were part of the V$SP1F matrix family. This is the same number of sites as seen when analysing the promoter using the matrix families' option. Although this tells us

that all the V$SP1F sites bind similar factors this is not unsurprising. However, one interesting result that cannot be seen from looking at the matrix family results summary table is that of the 13 V$KLFS sites identified seven of these are capable of binding KLF7. This is interesting because it is known that KLF7 is a ubiquitously expressed KLF protein (Matsumoto *et al.*, 1998) and therefore is likely to be expressed in CHO cells. It might also suggest that the SV40 promoter might rely on the ubiquitously expressed proteins such as SP1 and KLF7 for its expression. The majority of the rest of the individual matrices are homeobox related factors whose roles might be less important in our context due to their expression and function being in mainly developing organisms. However, it is known that some homeobox proteins are involved in cancer and in particular ovarian cancers (Basu & Roy, 2013; Christensen *et al.*, 2008; Crijns *et al.*, 2007) and the cell lines used in this thesis are CHO cells which are essentially immortal cancer cell lines. Although they may have roles post development these are less well defined and it may be sensible to consider TFBSs for TFs which are known to be ubiquitously expressed more worthy of further attention before these homeobox factors. Other individual matrices that should be mentioned are the 4 V$NFKAPPAB65.02, V$AP1F, V$E2F1_DP1.01 and V$E2F4 sites identified. Both the NFKB and AP1 proteins have been well studied and shown to be important in the expression of many genes (Fujioka *et al.*, 2004) while some E2F proteins are known to be ubiquitously expressed and are regulators of the cell cycle (Trimarchi & Lees, 2002). All may play roles in expression from the SV40 promoter.

Table 5.2 – Top 20 MatInspector results for the SV40 early promoter searched using both matrix families and individual matrices

Matrix families		Individual matrices	
Matrix Family	Match Total	Matrix	Match Total
V$KLFS	13	V$SP1.02	11
V$SP1F	11	V$GC.01	11
V$OCT1	9	V$SP1.03	9
V$HBOX	9	V$KLF7.02	7
V$HOMF	8	V$ISX.01	6
V$HOXF	8	V$NOBOX.02	6
V$ZF02	8	V$LBX2.01	6
V$AP1R	7	V$SHOX2.01	6
V$CART	7	V$ALX4.01	6
V$STEM	6	V$MSX1.01	5
V$AP1F	4	V$SP1.01	5
V$RREB	4	V$AP1.01	4
V$NFAT	4	V$NFKAPPAB65.02	4
V$GCMF	4	V$GBX1.01	4
V$E2FF	4	V$E2F1_DP1.01	4
V$LHXF	4	V$E2F4.01	4
V$ETSF	3	V$MSX2.01	4
V$DLXF	3	V$HLXB9.01	4
V$BCDF	3	V$MSX.01	4
V$CEBP	2	V$KLF6.01	4

The table shows the numbers of potential TFBS identified in the SV40 promoter sequence by MatInspector. The left hand side of the table shows the results when searching for matrix families and the left hand side shows the results for individual matrices. Only the top 20 hits are shown for each and the full results can be found in the appendix.

The most prominent matrix family identified by MatInspector for the CMV promoter was by far V$CREB with 23 potential sites, 15 more than any other matrix family (table 5.3). The V$CREB matrix family contain binding sites for many related proteins from ATF proteins and also CREB proteins which have similar binding sites and also wide ranging roles in processes such as cell growth, survival and apoptosis (Persengiev & Green, 2003; Hai & Hartman, 2001). Although it is impossible to say with any certainty, but this large number of sites for this particular family suggests they will have some role to play in the expression from the CMV promoter to some extent. However, it should also be mentioned that the consensus binding sequence for ATF/CREB is a palindromic sequence (5' – TGACGTCA – 3') (Hai & Hartman, 2001) and therefore for the majority of sites identified, will be found in pairs with a site on both the + and - strand. The next most prevalent matrix families are V$E4FF, V$HOMF and V$HOXF with 9 sites each. Both V$HOMF and V$HOXF were also found in similar numbers in the SV40 promoter. The V$E4FF matrix family contains binding sites for the E4F1 protein which is ubiquitously expressed and has been shown to be important for adenovirus E1a inducible early gene expression (Fernandes & Rooney, 1997; Lee & Green, 1987; Raychaudhuri *et al.*, 1987). The fact that adenovirus E1a protein cleaves E4F1 to produce a transcriptionally active form suggests it potentially is not important for CMV promoter expression if the E4F1 is not actively cleaved in the chosen cell line and thus acting as a transcriptional repressor (Fernandes & Rooney, 1997; Lee & Green, 1987; Raychaudhuri *et al.*, 1987). Removal of the E4F1 sites from the CMV promter could potentially improve its expression. Again as with the SV40 promoter just looking at the matrix families present it is impossible to tell which TFBSs are actually functional and important for expression.

When looking at table 5.3 and the individual matrices the top 20 are dominated by both ATF and CREB sites. Given the number of V$CREB sites identified this is not surprising and many of these are redundant matches which overlap. This does show the large number of related proteins which could potentially bind this promoter. One of the most prevalent V$CREB individual matrices is V$ATF6.02 with nine matches. In chapter 3 we showed how ATF6(50) and ERSE could be used to control transcription/expression. However, upon closer inspection none of the V$ATF6.02

sites made up a complete ERSE (no adjacent NF-Y site) and therefore even if these ATF6 sites are functional in the CMV promoter they are likely to function differently to what was shown in chapter 3. Different individual NF-κB matrices also feature quite prominently and even though again many of these were redundant matches it shows how many different forms of NF-κB might be capable of binding the CMV promoter. As well as individual matrices for ATF/CREB and NF-κB factors another prevalent matrix was the V$E4F.01 which has nine potential sites. This is the same as found for matrix families and is not surprising due to the fact the V$E4FF family is only composed of this single individual matrix for the E4F1 TF. If the numbers of potential TFBSs present for a particular group of factors reflect an involvement in the regulation of a promoter then from this basic search we would expect ATF/CREB, NF-κB and E4F factors to be involved in regulation of the CMV promoter. This assumption might be logical and will possibly be true to a certain extent but caution should be taken. This will be the case for some promoters but not for others and it also is entirely possible that TFBSs found only in small numbers could be just as important if not more for the regulation of a promoter.

Table 5.3 – Top 20 MatInspector results for the CMV major immediate-early promoter searched using both matrix families and individual matrices

Matrix families		Individual matrices	
Matrix Family	Match Total	Matrix	Match Total
V$CREB	23	V$CREB1.01	10
V$E4FF	9	V$ATF.01	10
V$HOMF	9	V$CREB.02	10
V$HOXF	9	V$ATF.02	10
V$MYBL	8	V$CREB2.01	9
V$BRNF	7	V$ATF2.01	9
V$SORY	6	V$ATF6.02	9
V$NF1F	6	V$CREB.03	9
V$E2FF	6	V$E4F.01	9
V$RXRF	5	V$CREB2CJUN.01	8
V$SP1F	5	V$JUNDM2.01	8
V$KLFS	4	V$ATF1.01	8
V$HNF6	4	V$NFKAPPAB65.02	7
V$NFKB	4	V$NFKAPPAB.02	5
V$NKX6	4	V$NFKAPPAB.01	4
V$RORA	4	V$HIVEP1.01	4
V$AP1R	4	V$CREL.01	4
V$GLIF	4	V$NFKAPPAB65.01	4
V$SRFF	4	V$SRF.01	4
V$NFAT	4	V$HOXC8.01	4

The table shows the numbers of potential TFBS identified in the CMV promoter sequence by MatInspector. The left hand side of the table shows the results when searching for matrix families and the left hand side shows the results for individual matrices. Only the top 20 hits are shown for each and the full results can be found in the appendix.

Table 5.4 shows the top 20 MatInspector hits for the EF1α promoter. The most identified matrix family was the V$E2FF with 12 sites. As mentioned earlier this is composed of E2F proteins and related factors, many of which are ubiquitously expressed and involved in cell cycle regulation and synthesis of DNA in mammalian cells (Gaubatz *et al.*, 2000; Nevins, 1998). The next most prevalent were V$EGRF and V$ZF02, with 11 potential sites. The V$EGRF matrix family is composed of binding sites for EGR (early growth response) proteins which are involved in immune responses (Gomez-Martin *et al.*, 2010). While, as mentioned earlier, the V$ZF02 family contains binding sites for a set of related zinc finger DNA binding domain proteins. The next most prevalent matrix families are V$KLFS with 9 and V$NKXH, V$SP1F and V$PAX5 with 7 each. V$KLFS and V$SP1F were both mentioned earlier when talking about the SV40 promoter. Both V$NKXH and V$PAX5 matrix families contain the binding sites for different types of homeodomain proteins (Stanfel *et al.*, 2005; Pilz *et al.*, 1993; Adams *et al.*, 1992).

When looking at the individual matrices in table 5.4, the EF1α promoter unlike the SV40 and CMV promoters, has fewer individual matrices identified for its top hits. MatInspector identified 5 possible V$GRHL1.01, V$SP1.01, V$NM23.01 and V$ZBTB7.03 sites for each individual matrix with all four being from different matrix families. In comparison the top hits for both the SV40 and CMV had 11 and 10 respectively. The individual matrices for EF1α unlike the other promoters were not as dominated by 2-3 matrix families. This could possibly be due to the EF1α promoter not being made up of repeated DNA sequences and the larger variety of potential TFBSs present. This also makes it more difficult to make any assumptions about the TFs which might possibly be important in regulation of the EF1α promoter.

Table 5.4 – Top 20 MatInspector results for the human EF1α promoter searched using both matrix families and individual matrices

Matrix families		Individual matrices	
Matrix Family	Match Total	Matrix	Match Total
V$E2FF	12	V$GRHL1.01	5
V$EGRF	11	V$SP1.01	5
V$ZF02	11	V$NM23.01	5
V$KLFS	9	V$ZBTB7.03	5
V$NKXH	7	V$AHRARNT.03	4
V$SP1F	7	V$ZF5.01	4
V$PAX5	7	V$SP1.03	4
V$ETSF	6	V$GRHL3.01	4
V$HESF	6	V$NGFIC.01	4
V$HOMF	5	V$EGR2.01	4
V$NDPK	5	V$MAZR.01	4
V$GRHL	5	V$KKLF.01	4
V$MAZF	5	V$GC.01	4
V$STAT	5	V$PAX5.02	3
V$ZF5F	4	V$E2F1_DP1.01	3
V$AHRR	4	V$MYCMAX.03	3
V$CREB	4	V$HELT.01	3
V$HEAT	4	V$KLF7.01	3
V$EBOX	4	V$ZFX.01	3
V$ZFXY	3	V$ZFP57.01	3

The table shows the numbers of potential TFBS identified in the EF1α promoter sequence by MatInspector. The left hand side of the table shows the results when searching for matrix families and the left hand side shows the results for individual matrices. Only the top 20 hits are shown for each and the full results can be found in the appendix.

In summary MatInspector identified both a higher number of matrix families and more TFBSs in the both the CMV and EF1α promoters compared to the SV40 promoter. This reflected the differences in strength between the SV40 promoter and the other two promoters. Although the CMV had less matrix families and potential TFBSs than the EF1α promoter the matrix families found in the CMV had more redundant matches in terms of individual matrices. So although the EF1α could possibly bind TFs from a greater number of matrix families, the CMV may bind a greater number of individual TFs. It was also shown that both the SV40 and CMV promoters had more TFBSs when compared to representative sequences while the EF1α promoter had a similar number. This was probably due to not only the greater length of the EF1α promoter but also that the SV40 and CMV promoters contain repeated sequences.

If the results from MatInspector, in terms of the number of both matrix family and individual matrices, where solely used as guide for experimentally studying which TFs are important in the regulation of each promoter it might suggest that factors from the SP1 and KLF families as well as homeobox TFs for the SV40 promoter. ATF/CREB TFs, the E4F1 TF, NF-κB TFs and homeobox TFs for the CMV promoter and factors from the E2F and EGR families for the EF1α. The individual matrices for the EF1α unlike the SV40 and CMV promoter came from a greater number of matrix families which made it harder to make assumptions about the TFs possibly important in its regulation.

Although studying the promoters this way gives a large amount of information in terms of potential TFBSs, it is limited in so much that it is impossible to really deduce which sites or types of sites are actually important for transcription from these promoters. One thing to consider is that although these sites have been identified by MatInspector many of these sites could have just occurred through chance and play no role in regulation of these promoters. Looking at the top hits might suggest that these sites may be important but due to the different patterns of bp's in their sequences some sites may appear more often than others in DNA sequences generally. In the next section each promoter was analysed for TFBSs which are found to be overrepresented in each promoter sequence when compared to both genomic and promoter backgrounds.

5.5.2 – Overrepresented TFBS in the SV40, CMV and EF1α promoters

To see which TFBSs appeared more often and statistically significantly in these three promoter sequences than would be expected the SV40, CMV and EF1α promoters were analysed for the presence of overrepresented TFBSs and compared to both genomic and promoter backgrounds from *Mus musculus*. The hypothesis being that TFBSs that appeared in numbers statistically significantly greater than the average of a representative sample might potentially play a role in the regulation of a promoter.

Table 5.5 shows the results for the SV40 promoter compared to a genomic background. Similar to the MatInspector results it was shown that both the related V$SP1F and V$KLFS matrix families have the highest Z-scores and are the top results, being the most statistically significantly overrepresented matrix families. This is also the case when compared to a promoter background (table 5.6). This further suggests some role for these families of factors in the regulation of the SV40 promoter. In total 19 matrix families were found to be overrepresented when compared to both backgrounds and of these 19, 16 appeared in both sets of search results (22 matrix families in total). In total when compared to a genomic background, of the 19 matrix families identified, this composed of a 107 TFBSs and 114 TFBSs when compared to a promoter background. When compared to the raw results from MatInspector (table 5.1) the matrix families identified make up less than half (38%) of the total matrix families (50) while the number of TFBSs make up between 66-70% of the total number of potential TFBSs (163).

Table 5.5 – Overrepresented TFBSs in the SV40 early promoter compared to a genomic background

Matrix family	Matches	Expected	SD	Over representation	Z-Score	*p* values
V$SP1F	11	0.64	0.80	17.09	12.29	0.0000
V$KLFS	13	1.56	1.25	8.32	8.76	0.0000
V$ZF02	8	1.00	1.00	7.98	6.50	0.0000
V$RREB	4	0.32	0.57	12.45	5.61	0.0000
V$ZFXY	2	0.09	0.30	21.94	4.67	0.0000
V$AP1R	7	1.45	1.20	4.83	4.20	0.0000
V$AP1F	4	0.52	0.72	7.67	4.12	0.0000
V$STEM	6	1.25	1.12	4.80	3.80	0.0001
V$E2FF	4	0.66	0.81	6.09	3.51	0.0004
V$HBOX	9	2.85	1.68	3.16	3.36	0.0008
V$SAL1	2	0.18	0.42	11.3	3.15	0.0016
V$GCMF	4	0.78	0.88	5.14	3.09	0.0020
V$ZICF	2	0.19	0.43	10.72	3.04	0.0024
V$OCT1	9	3.69	1.91	2.44	2.51	0.0121
V$PAX9	1	0.04	0.20	24.27	2.26	0.0238
V$CART	7	2.91	1.70	2.41	2.11	0.0349
V$NOLF	2	0.32	0.57	6.21	2.08	0.0375
V$NFAT	4	1.24	1.11	3.24	2.04	0.0414
V$HOXF	8	3.66	1.91	2.19	2.02	0.0434

The table shows matrix families which were found to be overrepresented when compared to a *Mus musculus* genomic background. Expected represents the mean number of matches in an equally sized sample of the selected background. SD is the standard deviation of the expected mean value. Over representation is the fold factor of matches compared with the expected mean value. The Z-score is a measure of how far away the number of actual matches is from the expected mean value in units of the expected SD. The Z-score can be used to calculate *p* values. Z-scores ≥ 2 and ≤ -2 are considered statistically significant and roughly equal to a *p* value of 0.05.

Table 5.6 – Overrepresented TFBSs in the SV40 early promoter when compared to a promoter background

Matrix family	Matches	Expected	SD	Over representation	Z-Score	*p* values
V$SP1F	11	1.54	1.24	7.14	7.23	0.0000
V$KLFS	13	2.32	1.52	5.60	6.70	0.0000
V$STEM	6	0.87	0.93	6.93	4.98	0.0000
V$HBOX	9	1.86	1.36	4.84	4.88	0.0000
V$RREB	4	0.41	0.64	9.69	4.81	0.0000
V$AP1F	4	0.45	0.67	8.82	4.53	0.0000
V$SAL1	2	0.11	0.33	18.86	4.28	0.0000
V$AP1R	7	1.46	1.21	4.79	4.17	0.0000
V$OCT1	9	2.35	1.53	3.83	4.02	0.0001
V$ZF02	8	2.09	1.44	3.83	3.75	0.0002
V$CART	7	1.86	1.36	3.77	3.42	0.0006
V$HOXF	8	2.45	1.56	3.26	3.23	0.0012
V$ZFXY	2	0.18	0.43	10.92	3.08	0.0021
V$GCMF	4	0.89	0.94	4.47	2.76	0.0058
V$NFAT	4	0.99	0.99	4.04	2.53	0.0114
V$HOMF	8	3.42	1.84	2.34	2.21	0.0271
V$ZICF	2	0.30	0.55	6.69	2.20	0.0278
V$BCDF	3	0.74	0.86	4.06	2.05	0.0404
V$DLXF	3	0.77	0.88	3.91	1.98	0.0477

The table shows matrix families which were found to be overrepresented when compared to a *Mus musculus* promoter background. Expected represents the mean number of matches in an equally sized sample of the selected background. SD is the standard deviation of the expected mean value. Over representation is the fold factor of matches compared with the expected mean value. The Z-score is a measure of how far away the number of actual matches is from the expected mean value in units of the expected SD. The Z-score can be used to calculate *p* values. Z-scores ≥ 2 and ≤ -2 are considered statistically significant and roughly equal to a *p* value of 0.05.

The same search for the CMV promoter unsurprisingly showed the V$CREB matrix family as statistically overrepresented when compared to both a genomic background (table 5.7) and a promoter background (table 5.8). However, the most significantly overrepresented matrix family is the V$E4FF and this is also true for both background comparisons. In terms of the total numbers of matrix families identified as overrepresented when compared to the genomic background there was 13. This increased to 18 when compared to a promoter background. 12 of these matrix families were found in both background comparisons (19 matrix families in total). This difference in the number of matrix families identified between backgrounds also meant that the number of TFBSs rose from 89 to 115 when looking at the genomic and promoter background comparisons respectively. Comparing these numbers to the MatInspector results (table 5.1) the total number of matrix families identified compared to a genomic and promoter background is around 17% and 23% respectively when compared to the total for MatInspector (77). In terms of the numbers of individual TFBS those found to overrepresented accounted for 40% and 51% of the total identified by MatInspector (225) for both a genomic and a promoter background.

Table 5.7 – Overrepresented TFBSs in the CMV major immediate-early promoter compared to a genomic background

Matrix family	Matches	Expected	SD	Over representation	Z-Score	*p* values
V$E4FF	9	0.26	0.51	34.21	16.06	0.0000
V$CREB	23	1.89	1.37	12.17	15.02	0.0000
V$NF1F	6	0.55	0.74	10.98	6.70	0.0000
V$E2FF	6	0.77	0.87	7.84	5.41	0.0000
V$MYBL	8	1.32	1.15	6.04	5.37	0.0000
V$DEAF	2	0.08	0.29	23.79	4.88	0.0000
V$SP1F	5	0.75	0.87	6.66	4.33	0.0000
V$NFKB	4	0.58	0.76	6.87	3.83	0.0001
V$GLIF	4	0.60	0.77	6.68	3.75	0.0002
V$RORA	4	0.86	0.92	4.67	2.86	0.0042
V$HNF6	4	1.21	1.10	3.31	2.09	0.0366
V$HOXF	9	4.26	2.06	2.11	2.06	0.0394
V$RXRF	5	1.80	1.34	2.78	2.02	0.0434

The table shows matrix families which were found to be overrepresented when compared to a *Mus musculus* genomic background. Expected represents the mean number of matches in an equally sized sample of the selected background. SD is the standard deviation of the expected mean value. Over representation is the fold factor of matches compared with the expected mean value. The Z-score is a measure of how far away the number of actual matches is from the expected mean value in units of the expected SD. The Z-score can be used to calculate *p* values. Z-scores ≥ 2 and ≤ -2 are considered statistically significant and roughly equal to a *p* value of 0.05.

Table 5.8 – Overrepresented TFBSs in the CMV major immediate-early promoter compared to a promoter background

Matrix family	Matches	Expected	SD	Over representation	Z-Score	*p* values
V$E4FF	9	0.30	0.54	30.43	15.09	0.0000
V$CREB	23	2.00	1.41	11.52	14.54	0.0000
V$NF1F	6	0.61	0.78	9.82	6.26	0.0000
V$MYBL	8	1.47	1.21	5.45	4.98	0.0000
V$HOXF	9	2.86	1.69	3.15	3.34	0.0008
V$HNF6	4	0.73	0.85	5.48	3.25	0.0012
V$NFKB	4	0.76	0.87	5.28	3.15	0.0016
V$RORA	4	0.76	0.87	5.25	3.14	0.0017
V$DEAF	2	0.23	0.48	8.62	2.63	0.0085
V$GLIF	4	0.95	0.97	4.21	2.62	0.0088
V$E2FF	6	2.00	1.41	2.99	2.47	0.0135
V$NKX6	4	1.09	1.05	3.65	2.30	0.0214
V$HOMF	9	3.99	1.99	2.26	2.27	0.0232
V$BRNF	7	2.80	1.67	2.50	2.22	0.0264
V$SNAP	3	0.68	0.83	4.40	2.20	0.0278
V$SRFF	4	1.15	1.07	3.48	2.19	0.0285
V$NFAT	4	1.15	1.07	3.47	2.19	0.0285
V$SP1F	5	1.80	1.34	2.78	2.02	0.0434

The table shows matrix families which were found to be overrepresented when compared to a *Mus musculus* promoter background. Expected represents the mean number of matches in an equally sized sample of the selected background. SD is the standard deviation of the expected mean value. Over representation is the fold factor of matches compared with the expected mean value. The Z-score is a measure of how far away the number of actual matches is from the expected mean value in units of the expected SD. The Z-score can be used to calculate *p* values. Z-scores ≥ 2 and ≤ -2 are considered statistically significant and roughly equal to a *p* value of 0.05.

The most overrepresented matrix family in the EF1α promoter was V$EGRF when compared to a genomic background (table 5.9) and V$E2FF when compared with a promoter background (table 5.10), although both of these matrix families are overrepresented compared with each background. In terms of the total numbers identified there was more overrepresented matrix families (19) and individual TFBS (101) when compared to a genomic background then there was when compared to a promoter background (11 and 70). All 11 overrepresented matrix families shown to be overrepresented compared with a promoter background were also overrepresented compared to a genomic background. This change between backgrounds was quite different when compared to the results for both the SV40 and CMV promoters. When the results of the search for overrepresented TFBS in the EF1α promoter is compared to the results of the MatInspector search (table 5.1) the total number of matrix families identified compared to both a genomic and a promoter background are 17% and 10% of the total number of matrix families identified by MatInspector. For the total number of TFBSs when compared to a genomic and a promoter background they are 37% and 26% respectively of the total number of potential TFBS identified by MatInspector.

Table 5.9 – Overrepresented TFBSs in the human EF1α promoter compared to a genomic background

Matrix family	Matches	Expected	SD	Over representation	Z-Score	*p* values
V$E2FF	12	1.55	1.24	7.76	8.01	0.0000
V$ZF5F	4	0.24	0.49	16.38	6.59	0.0000
V$EGRF	11	1.78	1.33	6.17	6.53	0.0000
V$ZF57	3	0.19	0.43	15.88	5.32	0.0000
V$ZF02	11	2.36	1.53	4.66	5.30	0.0000
V$ZFXY	3	0.21	0.46	13.98	4.93	0.0000
V$AHRR	4	0.46	0.68	8.72	4.49	0.0000
V$HESF	6	1.09	1.04	5.51	4.23	0.0000
V$SP1F	7	1.52	1.23	4.62	4.05	0.0001
V$NDPK	5	0.87	0.93	5.72	3.88	0.0001
V$MAZF	5	0.92	0.96	5.45	3.74	0.0002
V$PAX5	7	1.75	1.32	4.00	3.60	0.0003
V$WHNF	2	0.16	0.40	12.55	3.36	0.0008
V$GRHL	5	1.42	1.19	3.52	2.58	0.0099
V$KLFS	9	3.68	1.91	2.45	2.52	0.0117
V$MEF3	3	0.60	0.77	5.03	2.47	0.0135
V$CDEF	1	0.04	0.21	22.72	2.17	0.0300
V$HDBP	1	0.05	0.21	22.22	2.14	0.0324
V$MTF1	2	0.32	0.56	6.33	2.11	0.0349

The table shows matrix families which were found to be overrepresented when compared to a *Mus musculus* genomic background. Expected represents the mean number of matches in an equally sized sample of the selected background. SD is the standard deviation of the expected mean value. Over representation is the fold factor of matches compared with the expected mean value. The Z-score is a measure of how far away the number of actual matches is from the expected mean value in units of the expected SD. The Z-score can be used to calculate *p* values. Z-scores ≥ 2 and ≤ -2 are considered statistically significant and roughly equal to a *p* value of 0.05.

Table 5.10 – Overrepresented TFBSs in the human EF1α promoter compared to a promoter background

Matrix family	Matches	Expected	SD	Over representation	Z-Score	*p* values
V$E2FF	12	4.05	2.01	2.97	3.71	0.000207
V$EGRF	11	4.03	2.00	2.73	3.23	0.001238
V$ZFXY	3	0.43	0.66	6.95	3.15	0.001633
V$ZF57	3	0.48	0.69	6.24	2.91	0.003614
V$PAX5	7	2.23	1.49	3.14	2.86	0.004236
V$HESF	6	1.75	1.32	3.42	2.83	0.004655
V$ZF02	11	4.92	2.21	2.23	2.52	0.011735
V$GRHL	5	1.46	1.21	3.42	2.51	0.012073
V$AHRR	4	1.11	1.05	3.62	2.28	0.022608
V$MEF3	3	0.66	0.81	4.57	2.28	0.022608
V$NDPK	5	1.68	1.30	2.98	2.18	0.029257

The table shows matrix families which were found to be overrepresented when compared to a *Mus musculus* promoter background. Expected represents the mean number of matches in an equally sized sample of the selected background. SD is the standard deviation of the expected mean value. Over representation is the fold factor of matches compared with the expected mean value. The Z-score is a measure of how far away the number of actual matches is from the expected mean value in units of the expected SD. The Z-score can be used to calculate *p* values. Z-scores ≥ 2 and ≤ -2 are considered statistically significant and roughly equal to a *p* value of 0.05.

Although as with the MatInspector search there are limitations to the assumptions that can be made with these results it is interesting the differences seen between the two viral promoters and the mammalian EF1α promoter. Both the total numbers of overrepresented potential TFBSs in the SV40 and CMV promoters make up a larger proportion (66-70% and 40-51%) of the total identified by MatInspector compared to the EF1α promoter (37-26%). This could be simply due to the EF1α sequence being two times the length or greater than the viral promoters in length and the repeated nature of the viral promoter sequences. The increased length would increase the expected numbers of each matrix family in both backgrounds and therefore the required number of TFBSs for a matrix family to be statistically overrepresented. Whereas with repeated shorter sequences (viral promoters) you would expect less expected numbers in a background sequence but would possibly get more than expected if a potential TFBS was found in part of a repeated viral promoter sequence. However the more interesting difference is the fact that there was a decrease in the numbers of overrepresented matrix families identified between genomic and promoter backgrounds in the EF1α promoter but an increase for the CMV promoter. This cannot be simply due to the size of the promoters and has to be due to the different kinds of TFBSs present in each promoter, the origin of the promoters and also origin of the backgrounds used for comparison. The TFBSs identified in the EF1α promoter were less likely to occur in genomic background but more likely to occur in promoters and hence the decrease in overrepresented matrix families, while the TFBSs in the CMV promoter are more likely to occur in a genomic background and less likely to occur in a promoter background. Since the EF1α promoter is the human homolog and the backgrounds used for comparison came from *M. musculus* it is not that surprising that a mammalian promoter has less overrepresented TFBSs when compared to a mammalian promoter background than a genomic background. At least between these promoters there seems to be a distinct difference between those of a viral and that of a mammalian origin as we have also seen when analysing the sequences using MatInspector.

In summary and in comparison to the results from MatInspector TF families which may play a role in regulation of the SV40 promoter include both the SP1 and KLF families

and homeobox TFs again. TFBSs for AP1 and RREB factors both appeared near the top of both tables (5.5 & 5.6) for the SV40 promoter. AP1 (Activating protein 1) is heteromeric protein complex which is composed of JUN, FOS and ATF dimmers and shown to be important in the regulation of differentiation, proliferation and apoptosis (Hess *et al.*, 2004). The protein RREB-1 (Ras-responsive element binding protein 1) binds the RREB site and has been shown to be involved in many forms of cancer (Costello *et al.*, 2012; Zou *et al.*, 2011; Oxford *et al.*, 2007; Thiagalingam *et al.*, 1996).

The ATF/CREB family and E4F1 TFBSs as well as having the highest number of identified sites they were also the most overrepresented. NF-κB TFBSs were also shown to be over-represented along with NF1 (Nuclear factor 1) TFBSs which were not mentioned previously. NF1 factor binding sites were the third most statistically overrepresented TFBSs in the CMV promoter for both background comparisons and NF1 TFs have been shown to act as transcriptional activators (Pjanic *et al.*, 2011). It should be noted though that the NF1 TFBS can be palindromic (Devries *et al.*, 1987). E2F factor TFBSs were also shown to be overrepresented in the CMV promoter compared to both backgrounds.

Like the SV40 and CMV promoter the top two MatInspector hits for the EF1α promoter, E2F and EGR family TFBSs, were also the also overrepresented. Some of the most statistically overrepresented TFBSs in both backgrounds involved different zinc-finger families of TFs which are less well defined in terms of specific families and functions compared to the other families of TFs mentioned so far.

This search might help reduce the number of potential candidates of TFs which are potentially important in the regulation of each of these promoters compared to the MatInspector results, but only if the hypothesis that overrepresentation is a sign that a TFBS is important in a promoter's regulation. This will not be true for all promoters but it could possibly be truer for promoters which are composed of repeated sequences. It has been shown that clusters of TFBSs for the same TF, termed homotypic clusters of TFBSs (HCTs), are important in the regulation of some genes and this was first shown in invertebrates and they have also been identified in the promoters and enhancers of human genes (Gotea *et al.*, 2010; Lifanov *et al.*, 2003).

One study showed that increases in the concentration of the TF NF-κB lead to increases in expression of genes which contained multiple NF-κB TFBSs in their promoters (Giorgetti *et al.*, 2010). NF-κB TFBSs were found to be overrepresented in the CMV promoter. Although we do not have proof that this occurs in the CMV it is possible that some of the overrepresented TFBSs and even some that were not overrepresented may function this way not only in the CMV promoter but in the other promoters as well.

Further experimentation would be needed to show whether the presence of multiple TFBSs for the same factor does indeed play a part in the regulation of these promoters. The simplest way of doing this is either through the over-expression or chemical induction of specific TFs. As well as looking for overrepresented TFBSs to identify potential TFBSs which are important for the regulation of the promoters studied another program in the Genomatix Software Suite, ModelInspector, allows you to search for so called TFBS modules which have known functionality and this was carried and the results presented in the next section.

5.5.3 – ModelInspector analysis of the SV40, CMV and EF1α promoters

The previous searches have identified potential TFBSs and those that are overrepresented in the promoters studied. Although this has given some insight into which TFs might be involved in each promoter the program ModelInspector gives more evidence on the functionality of some of the TFBSs present. It searches sequences for modules which are composed of two or more TFBSs which are within a certain distance and orientation of each other and have been shown in the literature to be important for the expression of a specific gene or expression from a certain promoter. Although lab work would still need to be carried out to confirm their functionality, this is more informative than searching solely for individual TFBSs due to the fact that most TFs do not normally function as single factors.

Table 5.11 shows the results of the ModelInspector search of the SV40 promoter. 33 model matches in total from ten different modules were identified in this sequence. Six of these different modules involved SP1 or related factors. The most prevalent

model involved different variants of SP1F_SP1F and 11 were found in total. Not all of these were identical and differed slightly in the SP1 or related factors involved while also in the distance between the TFBSs. Given the high numbers of V$SP1F sites identified by MatInspector this is perhaps not surprising that these sites are also involved in model matches as well. The next most prevalent model was ETSF_SP1F with 5 matches, followed by KLFS_SP1F and SP1F_ETSF with 4 and then E2FF_SP1F with 2. The top five model matches all involved SP1 or related factors. Along with the previous searches this adds further evidence that SP1 and its related factors are important for expression from the SV40 promoter. Other factors which might be important in SV40 expression which have not been mentioned include CEBP, MYB, AP1 and NFKB factors. In total eight different matrix families are involved in the modules identified in the SV40 promoter. To what extent each TF has on expression from the SV40 promoter it is unknown but it is likely that it will not solely rely on a single TF or group of related TFs such as SP1 and its related factors.

Looking at the matrix families involved in the model matches we can see that V$SP1F had the greatest number of TFBSs with 11 followed by V$KLFS and V$ETSF with six each, both V$NFKB and V$E2FF had two identified sites while the rest had one. There were 30 TFBSs identified in total by ModelInspector in the SV40 promoter from eight different matrix families.

Table 5.11 – Results for ModelInspector search of the SV40 early promoter

Model matches		Matrix families involved	
Model	Matches	Matrix family	Number
SP1F_SP1F	11	V$SP1F	11
ETSF_SP1F	5	V$ETSF	6
KLFS_SP1F	4	V$KLFS	6
SP1F_ETSF	4	V$E2FF	2
E2FF_SP1F	2	V$NFKB	2
ETSF_NFKB	2	V$AP1F	1
SP1F_KLFS	2	V$CEBP	1
AP1F_NFKB	1	V$MYBL	1
KLFS_KLFS_NFKB	1	Total	30
MYBL_CEBP	1		
Total	33		

The table shows the results for the SV40 early promoter when analysed using the program ModelInspector. The left hand side of the table shows the modules identified along with the number model matches in the promoter. The right hand side of the table shows the individual matrix families involved along with the numbers identified in the sequence. Some of the potential TFBSs within the promoter sequence appeared in more than one model match but this was counted as a single TFBS.

The results of the ModelInspector search of the CMV promoter are presented in table 5.12. 29 matches were identified within the CMV promoter which composed of 22 different modules. The top seven hits all had two matches each while the remaining matches only had a single hit. As was shown with MatInspector V$CREB matrices were again identified in high numbers with five out of the top six model matches and six in total containing V$CREB matrices. Nine modules involved at least one V$SP1F site, 6 involved V$NFKB, and 4 involved V$NF1F. Looking at the matrix families involved we can see that ten potential V$CREB TFBSs were identified followed by 6 each of V$NF1F, V$NFKB and V$SP1F. In total 17 different matrix families were involved in the model matches identified and 48 different potential TFBS.

Table 5.12 – Results for ModelInspector search of the human CMV major immediate-early promoter

Model matches		Matrix families involved	
Model	Matches	Matrix family	Number
CREB_NFKB	2	V$CREB	10
CREB_SP1F	2	V$NF1F	6
CREB_YY1F	2	V$NFKB	6
NF1F_EBOX	2	V$SP1F	6
NFAT_CREB	2	V$YY1F	4
NFKB_CREB	2	V$NFAT	3
SP1F_YY1F	2	V$EBOX	2
CEBP_NFKB	1	V$ETSF	2
CEBP_SP1F	1	V$AP1F	1
CREB_NFAT_NFAT	1	V$CAAT	1
EREF_SF1F	1	V$CEBP	1
ETSF_NFKB	1	V$EREF	1
ETSF_SP1F	1	V$FKHD	1
ETSF_SRFF	1	V$HNF1	1
FKHD_NF1F	1	V$SF1F	1
NF1F_NF1F_HNF1	1	V$SRFF	1
NFKB_AP1F	1	V$KLFS	1
SP1F_CAAT	1	Total	48
SP1F_KLFS	1		
SP1F_NF1F	1		
SP1F_NFKB	1		
YY1F_SRFF	1		
Total	29		

The table shows the results for the human CMV major immediate-early promoter when analysed using the program ModelInspector. The left hand side of the table shows the modules identified along with the number model matches in the promoter. The right hand side of the table shows the individual matrix families involved along with the numbers identified in the sequence. Some of the potential TFBSs within the promoter sequence appeared in more than one model match but this was counted as a single TFBS.

Table 5.13 shows the model matches identified for the EF1α promoter. 22 matches were identified and this involved 18 different modules. The highest recurring model was ETSF_KLFS with 3 matches followed by ETSF_NFKB and KLFS_SP1F with two. The remaining modules had a single match. V$ETSF was found more than any other matrix family in terms of the number of modules it was part of (7) followed by V$SP1F (6). In terms of the matrix families involved there were 18 different families and a total of 35 individual TFBS. V$ETSF was the most prevalent family with 7 sites followed by V$KLFS and V$SP1F with 5 and V$HAML, V$HIFF and V$NFKB with 2. The remaining matrix families had a single site.

Table 5.13 – Results for ModelInspector search of the human EF1α promoter

Model matches		Matrix families involved	
Model	Matches	Matrix family	Number
ETSF_KLFS	3	V$ETSF	7
ETSF_NFKB	2	V$KLFS	5
KLFS_SP1F	2	V$SP1F	5
EBOX_EREF	1	V$HIFF	2
ETSF_ETSF	1	V$HAML	2
ETSF_HAML	1	V$NFKB	2
ETSF_HIFF	1	V$AP1F	1
ETSF_SP1F	1	V$AP1R	1
HAML_AP1F	1	V$AP2F	1
HIFF_ETSF	1	V$EBOX	1
NBRE_AP1F	1	V$EREF	1
NFAT_AP1R	1	V$NBRE	1
NFKB_NKXH	1	V$NFAT	1
P53F_SP1F	1	V$NKXH	1
SORY_PAX6	1	V$P53F	1
SP1F_AP2F	1	V$PAX6	1
SP1F_KLFS	1	V$SORY	1
SP1F_YY1F	1	V$YY1F	1
Total	22	Total	35

The table shows the results for the human EF1α promoter when analysed using the program ModelInspector. The left hand side of the table shows the modules identified along with the number model matches in the promoter. The right hand side of the table shows the individual matrix families involved along with the numbers identified in the sequence. Some of the potential TFBSs within the promoter sequence appeared in more than one model match but this was counted as a single TFBS.

Comparing the results of the ModelInspector searches between the three different promoters we can see that the SV40 promoter is quite different from the other two at least in terms of the numbers of matrix families involved. Although the SV40 promoter had the highest number of model matches (33) when compared to the CMV (29) and EF1α (22) promoters it had the lowest number of different matrix families involved (10). There seemed a bias towards V$SP1F sites with it being the most identified matrix family, SP1F_SP1F modules being the most prevalent as well as SP1F being involved with six of the ten different modules found (60%). In terms of model matches none of the other two promoters had this bias towards a single factor. The CMV promoter did have 11 V$CREB sites but in terms of model matches it was always involved with another matrix family and although involved in six different modules this was out of 22 different modules in total (27%). V$ETSF was the most prevalent matrix

family in the EF1α promoter with seven TFBS and was involved in seven different modules but this was out of the 18 in total (39%). These results would suggest that the SV40 promoter may be highly reliant on SP1 factors for its expression and fits in with previous work which has shown SP1 to be important factor for transcription from the SV40 promoter (Saffer *et al.*, 1990; Dynan & Tjian, 1983). Other factors are probably involved in the expression from the SV40 promoter and are either required or aid SP1.

Looking at the ModelInspector results for the CMV and EF1α promoters although they both had less model matches in total the number of different modules identified were greater, 22 and 18 respectively, then the SV40 (10). Unlike the SV40 promoter the CMV and EF1α did not seem to rely so heavily on a signal TF or TF family and both had a much greater variety in terms of both modules and also the matrix families involved. The CMV had 17 different matrix families and 48 individual TFBS, the EF1α had 18 and 35 while the SV40 had 8 and 30. It is possible that it is this extra variety of possible TFs that make the CMV and EF1α promoters stronger than the SV40 promoter. A lack or limited supply of SP1 may limit expression from the SV40 promoter were as the CMV and EF1α promoters maybe less reliant on a single TF family. Alternatively the TFs most important for expression from the CMV and EF1α promoters might be stronger transcriptional activators.

In summary the results shown by the ModelInspector searches for each promoter provide stronger evidence then the previous searches for potential TFBSs which are important in the regulation of these promoters due to the modules identified being shown to be functional in the corresponding literature that was used to construct them. The ModelInspector results should be used as a guide as to which TFs and TFBSs should be studied in each promoter to gain further understanding of their regulation.

As well as trying to identify TFBSs which may be potentially important for the regulation of these promoters we also want to understand why these promoters, especially the CMV, interfere or compete and have negative effects on the expression from a co-expressed promoter. One hypothesis is that there is competition for shared TFs. In the next section we look at TFBSs which are shared/common between the three promoters.

5.5.4 – Common TFBS in the SV40, CMV and EF1α promoters

One of the hypotheses of why these three promoters interfere with the expression of one another is that there is competition for shared TFs. One simple way of investigating this is to compare the three promoters for the potential TFBSs they contain, as identified by MatInspector, and share between themselves. As shown earlier the SV40 promoter has TFBSs from 50 different matrix families while the CMV and EF1α have 77 and 122 respectively (table 5.1). However, how many of these matrix families are present in all three promoters and how much do the numbers of different potential TFBSs identified by MatInspector for each promoter from the different matrix families' overlap? As well as looking at which were shared between all three, we also looked at which were shared between just two of the promoters. Whether the matrix family was overrepresented was also taken into account. Venn diagrams were also produced to show how the numbers and proportion of both the different matrix families and the different TFBSs identified in each of promoters were shared between each promoter. This was also done for the ModelInspector results.

In total for the MatInspector searches of all three promoters TFBSs from 137 different matrix families were identified. Table 5.14 shows the matrix families which are present in all three promoters and they are in order of the total number of matches. 28 different matrix families were shown to be present in all three promoters (table 5.14 & figure 5.2). This is 56% of the matrix families identified for the SV40 promoter (50 matrix families), 36% for the CMV (77 matrix families) and 25% for the EF1α promoter (112 matrix families). 14 of the matrix families for the SV40 were also shown in the previous section to be overrepresented compared to either a genomic or promoter background or both (table 5.14). A further 14 matrix families were also found either just in the CMV promoter (7) (table 5.15 and figure 5.2) or the EF1α promoter (7) (table 5.16 & figure 5.2). Four of the seven matrix families shared just between the SV40 and CMV promoters were found to be overrepresented in the SV40 in either or both background comparisons, while for those shared just between the SV40 and EF1α this was two out of seven. In total 20 out of 22 overrepresented matrix families in the SV40 were found in either one or both of the CMV and EF1α promoters.

Of the 50 matrix families found in the SV40 promoter, 70% were found in either the CMV (35) or EF1α (35) promoter with 84% (42) in either or both (figure 5.2).

As mentioned earlier for the CMV promoter, the 28 different matrix families shared by all three promoters (table 5.14 & figure 5.2) makes up ~36% of the total matrix families (77 matrix families) identified by MatInspector and nine of these matrix families were also shown to be overrepresented. Seven more matrix families are also shared only with the SV40 promoter, one of which is overrepresented (table 5.15). The CMV promoter therefore shares 35 matrix families in total with the SV40 which is 45% of the matrix families identified by MatInspector (77). There are 60 matrix families shared between the CMV and EF1α promoters which are 78% of the total number CMV matrix families (77) and five of these matrix families are overrepresented in the CMV promoter. 15 out of the 19 overrepresented matrix families in the CMV promoter were shared we either the SV40 or EF1α promoter or both. Of the 77 matrix families found in total in the CMV promoter, 45% (35) were found in the SV40, 78% (60) in the EF1α promoter with 87% (67) in either or both (figure 5.2).

The EF1α promoter (112 matrix families) shares 25% of its matrix families (28) with both the SV40 and CMV and 7 of these matrix families are overrepresented (table 5.14 & figure 5.2). In total it shares 35 different matrix families with the SV40 promoter, which is around 31% of the total number of matrix families (112) and 10 of these are overrepresented in the EF1α promoter (table 5.14, 5.16 & figure 5.2). With the CMV promoter the EF1a shares 60 matrix families which are 54% of the total numbers of matrix families found in the EF1α promoter (112) and eight are overrepresented in the EF1α promoter. 11 out of the 19 overrepresented matrix families in the EF1α promoter were shared with either the SV40 or CMV promoter or both. In total, of the 112 different matrix families identified in the EF1α promoter, 31% (35) are shared with the SV40 promoter, 54% (60) with the CMV and 60% (67) with both.

The number of shared potential TFBSs from the different matrix families identified in each promoter was also compared. In figure 5.3 we can see the overlap between the different promoters. 52 TFBSs were found in all three of the promoters and this was 32% of those identified by MatInspector in the SV40 promoter (163 TFBSs), 23% of

those in the CMV (225 TFBSs) and 19% of those in EF1α promoter (274 TFBSs). The SV40 promoter shared 81 TFBSs (50% of the total number of TFBS in the SV40 promoter) with the CMV, 78 TFBSs (48%) with the EF1α promoter and 107 (66%) of its TFBSs with both. In turn the CMV shared 81 TFBSs (36% of the total number of TFBS in the CMV promoter) with the SV40, 106 TFBSs with the EF1α promoter (47%) and 135 (60%) TFBSs with both promoters. The EF1α promoter shared 78 TFBSs with SV40 (28% of the total number of TFBS in the EF1α promoter), 106 TFBSs with the CMV promoter (39%) and 132 TFBSs (48%) with both the SV40 and CMV promoters.

In summary, the SV40 promoter shares a greater proportion of both its matrix families and the potential TFBSs identified by MatInspector with the CMV and EF1α promoters then both these promoters do with the SV40 promoter. This is mainly because of the lower numbers found in the SV40 compared with the other two promoters but the fact the SV40 shares around 70% similarity in terms of its matrix families with either the CMV and EF1α, and 50% and 48% similarity respectively in terms of the TFBSs potentially present in the SV40 promoter, shows that there is significant overlap between both the matrix families and TFBSs present and this could be a possible cause for the promoter interference/competition in the previous chapter and why the SV40 promoter was the most affected by co-expression with another competing promoter.

In terms of those matrix families that were overrepresented in the SV40 promoter (table 5.5 & 5.6) 20 of the 22 matrix families were found either in the CMV or EF1α or both promoters. The most overrepresented matrix family was the V$SP1F family (11 TFBSs) and this was also overrepresented for both the CMV (5 TFBSs) and EF1α (7 TFBSs) promoters (table 5.14). The V$E2FF matrix family was overrepresented in all three promoters (table 5.5 – 5.10) with 4 TFBSs in the SV40, 6 in the CMV and 12 TFBSs in the EF1α promoters. Only these two matrix families were overrepresented to one or both background comparisons in all three promoters. The V$KLFS family was also overrepresented in the SV40 promoter (13) and again it shared sites with both CMV (4 TFBSs) and the EF1α (9 TFBSs) but V$KLSF was only overrepresented in the EF1α promoter.

Five matrix families were overrepresented in both the SV40 and CMV promoters and as well as those mentioned in the last paragraph (V$SP1F and V$E2FF) they include the V$HOMF (8 & 9 TFBSs respectively), V$HOXF (8 & 9) and V$NFAT (4 in each) families. These first two are families of homeobox TFs and have been mentioned earlier in this chapter. The V$NFAT family contains TFBSs for NFAT (nuclear factor of activated T-cells) factors which have been shown to be involved in immune responses and has also a role in the development and metastasis of cancer (Pan *et al.*, 2013). The SV40 also shares 5 overrepresented matrix families with the EF1α promoter and as well as V$SP1F and V$E2FF these include V$KLFS (13 and 9 TFBSs respectively), V$ZF02 (8 and 11) and V$ZFXY (2 and 3). The V$ZFXY family contains TFBSs for the ZFY family of TFs which are implicated in mammalian sex determination with ZFX also playing a role in stem cell self-renewal and in some forms of cancer (Jiang *et al.*, 2013; Tan *et al.*, 2013; Zhou *et al.*, 2011; Galan-Caridad *et al.*, 2007; Poloumienko, 2004). TFs which bind the matrices of the V$ZF02 family have less well defined roles between them but one TF, ZF9, has been shown to work with the SP1 TF in the regulation of some genes (Yasuda *et al.*, 2002; Kim *et al.*, 1998).

In summary, 8 out of 22 of the SV40 promoters' matrix families were also found to be overrepresented in either the CMV or EF1α promoter or both. The V$SP1F and V$E2FF matrix families are the only two matrix families which were shown to be overrepresented in both the CMV and EF1α promoters. This could be one potential reason why the CMV and EF1α promoters interfered with expression more from the SV40 promoter then they did with each other in chapter 4.

Table 5.14 – Common TFBS – matrix families which are present in all three promoters

		Promoter		
Matrix Family	Match Total	SV40	CMV	EF1α
V$CREB	28	1	23[g, p]	4
V$KLFS	26	13[g, p]	4	9[g]
V$SP1F	23	11[g, p]	5[g, p]	7[g]
V$HOMF	22	8[p]	9[p]	5
V$E2FF	22	4[g]	6[p]	12[g, p]
V$ZF02	21	8[g, p]	2	11[g]
V$HOXF	19	8[g, p]	9[g, p]	2
V$HBOX	15	9[g, p]	3	3
V$EGRF	13	1	1	11[g, p]
V$AP1R	13	7[g, p]	4	2
V$ETSF	12	3	3	6
V$PAX5	11	2	2	7[g, p]
V$MYBL	11	1	8[g, p]	2
V$NFKB	9	2	4[g, p]	3
V$NFAT	9	4[g, p]	4[p]	1
V$CART	9	7[g, p]	1	1
V$LHXF	9	4	3	2
V$AP1F	8	4[g, p]	2	2
V$GRHL	7	1	1	5[g, p]
V$RORA	7	2	4[g, p]	1
V$GCMF	7	4[g, p]	2	1
V$RREB	6	4[g, p]	1	1
V$RUSH	6	2	1	3
V$NOLF	5	2[g]	1	2
V$NR2F	5	1	1	3
V$HAND	5	2	1	2
V$PAX6	5	1	2	2
V$ABDB	5	1	1	3
Total TFBS	338	117	108	113

The table shows matrix families which are found in all three of the promoters analysed along with the total number found and the number of TFBSs identified in each promoter by MatInspector. The letters next to the number denote – g = overrepresented compared to a genomic background and p = overrepresented compared to a promoter background and this taken from the searches carried out in the previous section for overrepresented TFs (5.5.2).

Table 5.15 – Common TFBS – matrix families which are present in two of the promoters – SV40 and CMV

		Promoter	
Matrix Family	Match Total	SV40	CMV
V$OCT1	12	$9^{g, p}$	3
V$STEM	7	$6^{g, p}$	1
V$DLXF	5	3^{p}	2
V$SNAP	5	2	3^{p}
V$BRN5	5	1	4
V$BCDF	4	3^{p}	1
V$ATBF	2	1	1
Total	40	25	15

The table shows matrix families which are found in two of the three promoters analysed, SV40 early and human CMV major immediate-early, along with the total number found and the number of TFBSs identified in each promoter by MatInspector. The letters next to the number denote – g = overrepresented compared to a genomic background and p = overrepresented compared to a promoter background and this taken from the searches carried out in the previous section for overrepresented TFs (5.5.2).

Table 5.16 – Common TFBS – matrix families which are present in two of the promoters – SV40 and EF1α

		Promoter	
Matrix Family	Match Total	SV40	EF1α
V$NDPK	7	2	$5^{g, p}$
V$ZFXY	5	$2^{g, p}$	$3^{g, p}$
V$CEBP	5	4	1
V$OAZF	3	1	2^{g}
V$ZICF	3	$2^{g, p}$	1
V$AP2F	3	2	1
V$BRAC	2	1	1
Total	28	14	14

The table shows matrix families which are found in two of the three promoters analysed, SV40 early and human EF1α, along with the total number found and the number of TFBSs identified in each promoter by MatInspector. The letters next to the number denote – g = overrepresented compared to a genomic background and p = overrepresented compared to a promoter background and this taken from the searches carried out in the previous section for overrepresented TFs (5.5.2).

Table 5.17 – Common TFBS – matrix families which are present in two of the promoters – CMV and EF1α

		Promoter	
Matrix Family	Match Total	CMV	EF1α
V$BRNF	9	7[p]	2
V$NKXH	9	2	7
V$STAT	8	3	5
V$RXRF	8	5[g]	3
V$SORY	8	6	2
V$NF1F	7	6[g, p]	1
V$GLIF	7	4[g, p]	3
V$PARF	7	4	3
V$YY1F	6	3	3
V$AHRR	5	1	4[g]
V$NKX6	5	4[p]	1
V$SRFF	5	4	1
V$HEAT	5	1	4
V$PDX1	4	3	1
V$CTCF	4	1	3
V$GREF	4	2	2
V$XBBF	4	3	1
V$IRXF	3	1	2
V$SAL2	3	1	2
V$RBPF	3	1	2
V$EREF	3	2	1
V$CLOX	3	2	1
V$CAAT	3	1	2
V$LEFF	3	2	1
V$FKHD	3	2	1
V$PAX1	2	1	1
V$GZF1	2	1	1
V$NBRE	2	1	1
V$ZF10	2	1	1
V$PAX3	2	1	1
V$PLAG	2	1	1
V$DMRT	2	1	1
Total	143	78	65

The table shows matrix families which are found in two of the three promoters analysed, human CMV major immediate-early and human EF1α, along with the total number found and the of TFBSs identified in each promoter by MatInspector. The letters next to the number denote – g = overrepresented compared to a genomic background and p = overrepresented compared to a promoter background and this taken from the searches carried out in the previous section for overrepresented TFs (5.5.2).

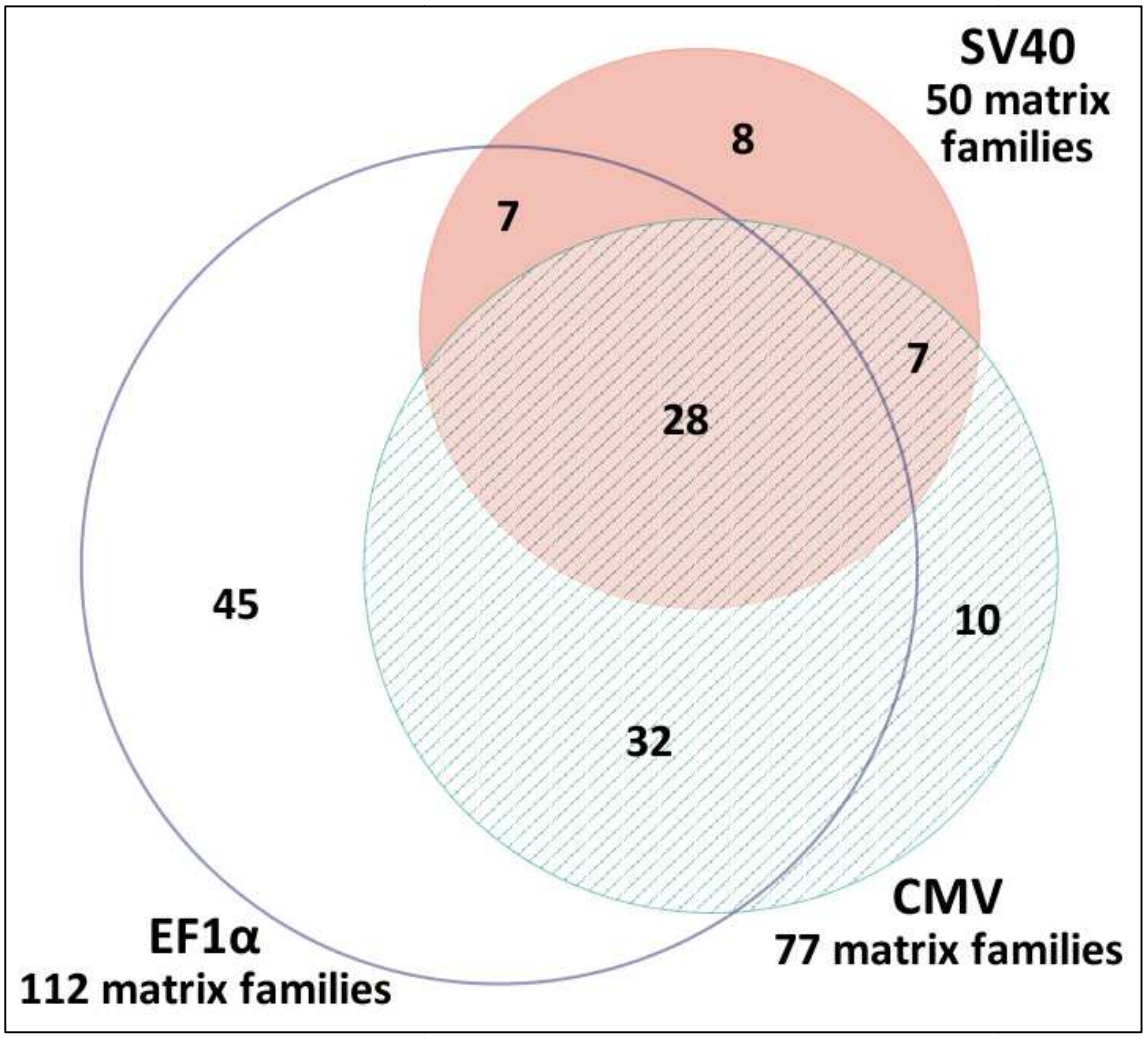

Figure 5.2 – Venn diagram showing the numbers and proportion of matrix families shared between the SV40, CMV and EF1α promoters identified by MatInspector. Lists of the different matrix families identified in each promoter by MatInspector were compared and a proportional Venn diagram was constructed to show how the matrix families were shared between the different promoters. The numbers below the names of each promoter show the total number of matrix families present.

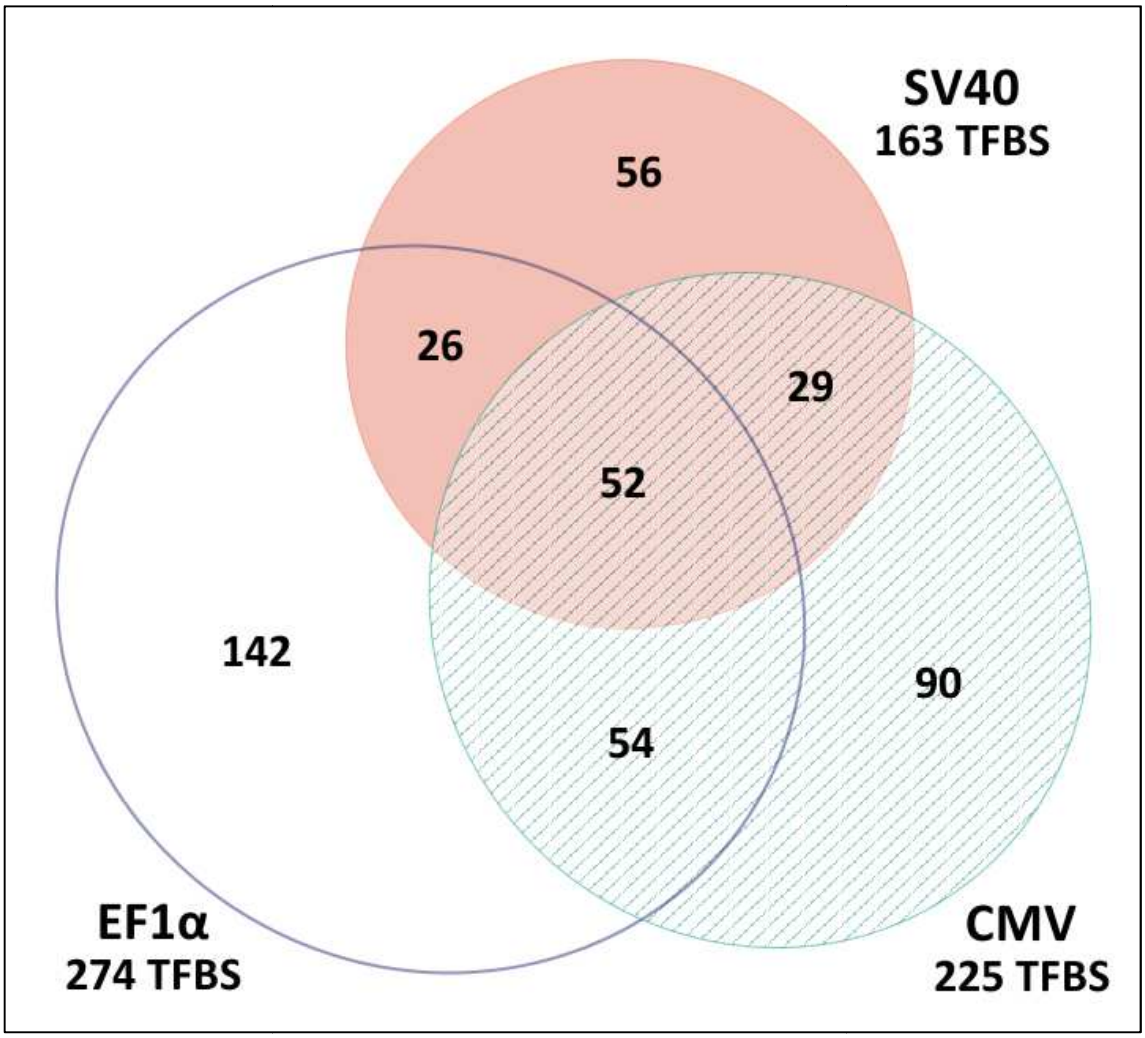

Figure 5.3 – Venn diagram showing the numbers and proportion of potential TFBSs shared between the SV40, CMV and EF1α promoters identified by MatInspector. Lists of the different potential TFBSs identified in each promoter by MatInspector were compared and a proportional Venn diagram was constructed to show how the TFBSs were shared between the different promoters. The numbers below the names of each promoter show the total number of potential TFBSs present.

The results from the ModelInspector searches (section 5.5.3) were also compared for both the common matrix families and potential TFBSs that were found in the modules of each promoter (table 5.18, figure 5.4. & 5.5). In total five different matrix families were shared between all three promoters and this was 63% of the total found in the SV40 promoter (8 matrix families), 29% for the CMV (17) and 28% for the EF1α promoter (18) (table 5.18 & figure 5.4). The SV40 promoter shared one more with the CMV promoter alone making the total shared with the CMV 6 (75%). The SV40 promoter did not share any matrix families with the EF1α promoter alone. In total the SV40 shared 75% (6) of its matrix families (8) with either the CMV or EF1α or both promoters. The CMV promoter shared 5 matrix families (29%) with both the SV40 and EF1α promoters, 6 with the SV40 (35%) and 9 with the EF1α promoter (53%). The CMV shared in total 10 of its 17 matrix families (59%) with either the SV40 or EF1α promoter or both. As mentioned earlier the EF1α promoter shares 5 of its 18 (28%) identified matrix families with both the SV40 and CMV promoters and 9 in total with the CMV promoter (50%). It shares none alone with the SV40 promoter. Therefore 50% (9) of the EF1α promoters' matrix families (18) can be found in either or both of the SV40 or CMV promoters.

In terms of the potential TFBSs identified by ModelInspector in each promoter 10 TFBSs were shared between all three promoters (figure 5.5) and this was 33% of those identified in the SV40 promoter (30 TFBSs), 21% of the CMV (48 TFBSs) and 29% of those found in the EF1α promoter (35 TFBSs). The SV40 shared 12 TFBSs with the CMV (40%), 19 with the EF1α promoter (63%) and 21 (70%) with either or both the CMV and EF1α promoters. Of the 48 TFBSs in the CMV promoter it shared 12 with the SV40 promoter (25%), 14 with the EF1α (29%) and 16 TFBSs (33%) with either or both of the other two promoters. 35 TFBSs were identified in the EF1α promoter by ModelInspector and 19 of these were shared with the SV40 (54%), 14 with the CMV (40%) and 23 TFBSs (66%) with either the SV40 or CMV or both promoters.

Of the different matrix families identified by ModelInspector only one, V$SP1F, was shown to be overrepresented compared to either a genomic or promoter background or both in all three of the promoters (table 5.18). The SV40 promoter had the most V$SP1F TFBSs with 11 which is slightly more than one third of the TFBSs identified by

ModelInspector in this promoter (30). The CMV had 5 and the EF1α had 6 but these made up less of their total numbers of TFBSs identified by ModelInspector. The V$KLFS family was overrepresented in the SV40 promoter with 6 sites but was not overrepresented in either of the other two promoters. The CMV had 2 V$KLFS TFBSs while the EF1α had 5. The V$ETSF family was present in all three promoters, the SV40 and the EF1α promoters had 6 and 7 TFBSs respectively while the CMV had 2. All three promoters had one V$AP1F site while V$NFKB TFBSs were also present in all three but with the SV40 and EF1α promoter having 2 each and the CMV having 6. These 6 V$NFKB TFBSs were considered to be overrepresented in the CMV promoter in comparison to both backgrounds.

Table 5.18 – Shared matrix families from those identified using ModelInspector

	Promoter		
Matrix family	SV40	CMV	EF1α
V$SP1F	11[g, p]	6[g, p]	5[g]
V$KLFS	6[g, p]	1	5
V$ETSF	6	2	7
V$NFKB	2	6[g, p]	2
V$AP1F	1	1	1
V$CEBP	1	1	-
V$YY1F	-	4	1
V$NFAT	-	3	1
V$EBOX	-	2	1
V$EREF	-	1	1

The table shows the matrix families shared between the three promoters when analysed by ModelInspector and the number of TFBSs identified in each promoter. The letters next to the number denote – g = overrepresented compared to a genomic background and p = overrepresented compared to a promoter background.

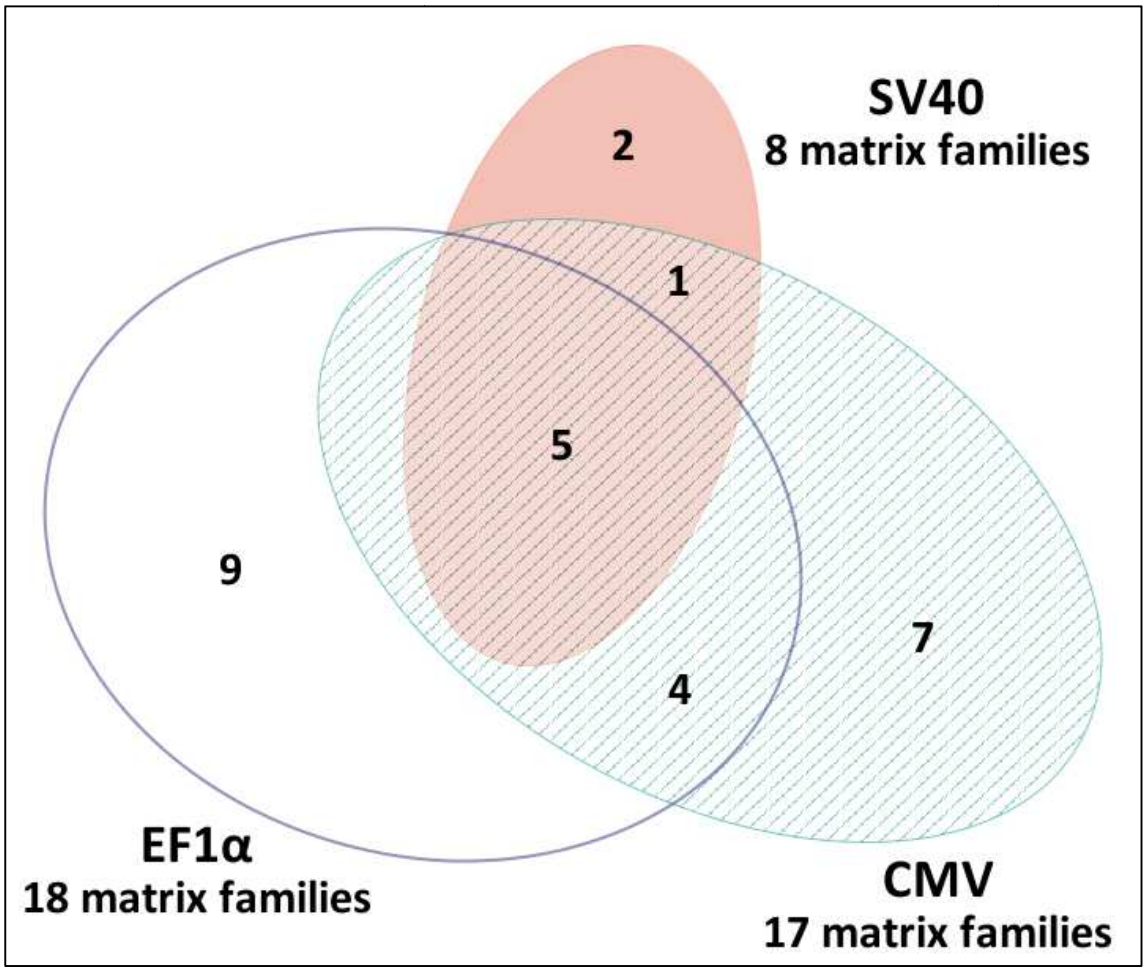

Figure 5.4 – Venn diagram showing the numbers and proportion of matrix families shared between the SV40, CMV and EF1α promoters identified ModelInspector. Lists of the different matrix families identified in each promoter by ModelInspector were compared and a proportional Venn diagram was constructed to show how the matrix families were shared between the different promoters. The numbers below the names of each promoter show the total number of matrix families present.

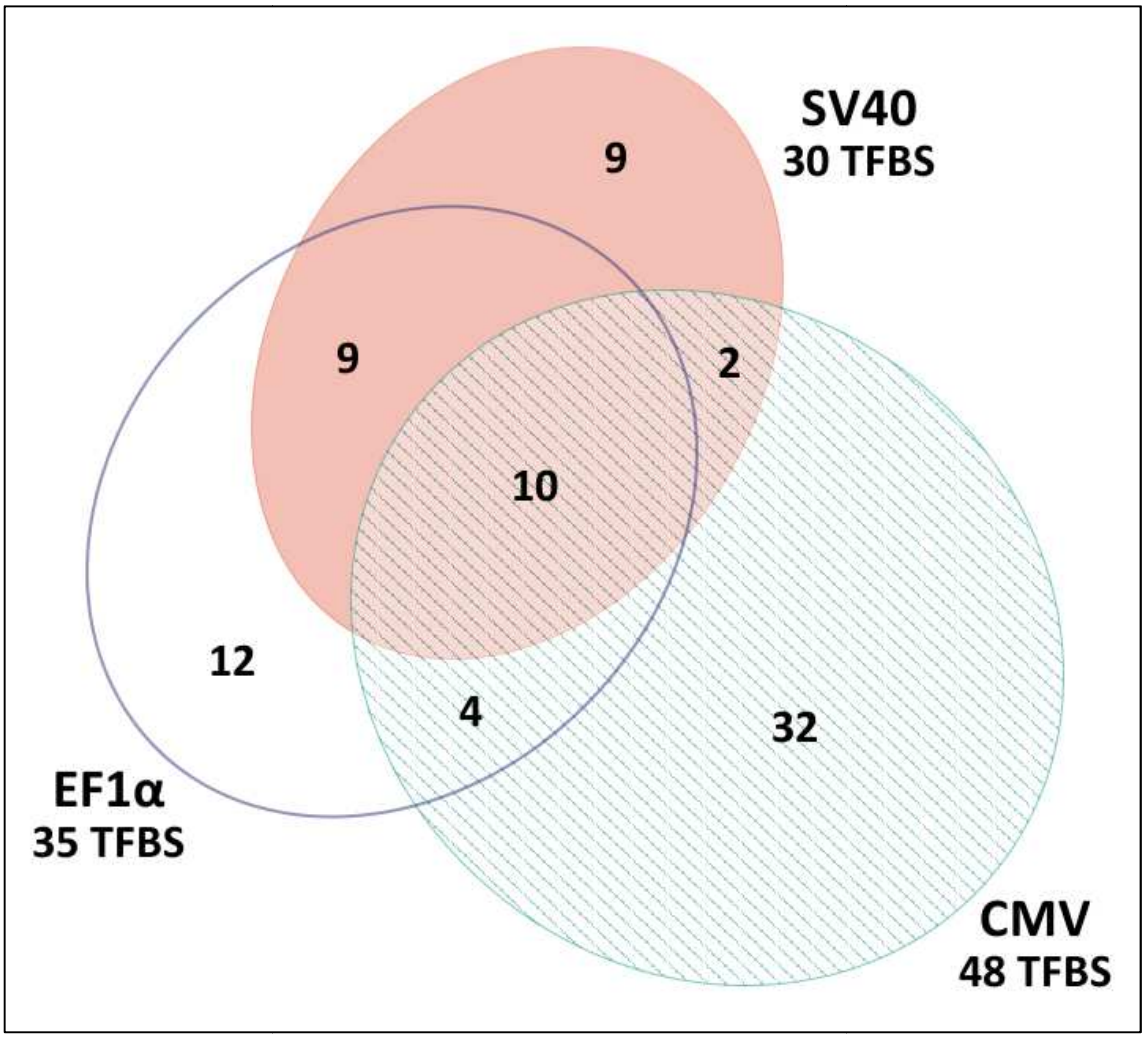

Figure 5.5 – Venn diagram showing the numbers and proportion of potential TFBSs shared between the SV40, CMV and EF1α promoters identified by ModelInspector. Lists of the different potential TFBSs identified in each promoter by ModelInspector were compared and a proportional Venn diagram was constructed to show how the TFBSs were shared between the different promoters. The numbers below the names of each promoter show the total number of potential TFBSs present.

This section has shown which matrix families and potential TFBSs that the SV40, CMV and EF1α promoters have in common from both the MatInspector and ModelInspector searches. They have both shown that the matrix families and TFBSs identified in the SV40 promoter, which are also common to either the CMV or EF1α promoter, make up a greater proportion of the total identified in SV40 than the proportion in either the CMV or EF1α respectively. It was also shown that the SV40 promoter shared more of its matrix families which were overrepresented (8 out of 22) with the other two promoters and also that a larger number of these overrepresented matrix families were also overrepresented in either the CMV (5) or EF1α (5) or both promoters. There were less overrepresented matrix families (2) which were found to be overrepresented and common in between the CMV and EF1α promoters. Taken this and what was shown in the previous sections that the SV40 promoter potentially has less TFBSs which are also from a smaller number of matrix families and that V$SP1F TFBSs make up a large proportion of the of the modules and related TFBSs identified in the SV40 promoter it would seem likely that this promoter would be more affected by the presence of a competing promoter especially one which had a greater variety of TFBSs but which also shared common factors with the SV40 such as SP1 TFBSs.

Although not all the matrix families and their TFBSs identified will be functional in the SV40, or in the other promoters, the fact that the SV40 promoter shares such a large percentage of its matrix families with either of the other two promoters or both suggests there would likely be competition for TFs which are important for expression from the SV40 promoters. Even if the TFBSs in the CMV and EF1α are not functional or important in their own expression they could still bind TFs, probably only loosely, which are needed for the SV40, removing them from the available pool of TFs and limiting their availability. This could be seen as transcriptional squelching which is believed to be caused by competition for a limited transcriptional regulator (Huliak *et al.*, 2012; Cahill *et al.*, 1994b; Prywes & Zhu, 1992) which in this case could be a TF/s. It seems highly probable that the presence of shared TFBSs contributes to the promoter interference/competition that occurs between the SV40 and CMV or EF1α promoters and the fact as well as being the weakest promoter that most of the matrix families

present in the SV40 are also present in the other two promoters could be one reason why the SV40 is more affected by competition from another promoter.

Throughout this chapter TFBSs from the V$SP1F family have been continually identified in all three of the promoter sequences but especially in the SV40 promoter. Not only were TFBSs from the V$SP1F family the second most prevalent sites identified in the SV40 (table 5.2), with V$SP1.02 being the most prevalent individual matrix identified (table 5.2) they were also the most overrepresented compared to both a genomic (table 5.5) and promoter (table 5.6) background. V$SP1F sites were also involved in the majority of modules and also the most prevalent matrix family identified in the SV40 by ModelInspector (table 5.11). This would suggest that SP1 family of TFs play a role in regulation of the SV40 promoter. The fact that V$SP1F family TFBSs are also found in the other two promoters, found in TFBS modules and also overrepresented, V$SP1F being the only matrix that was overrepresented when the results of the ModelInspector results were analysed, suggests that it also plays some role in the regulation of the CMV and EF1α promoters even if it's to a lesser extent. This made SP1 TFs a likely candidate for there to be competition over when more than one promoter is present. In the next section a recombinant form of SP1 is expressed to see the effects on all the promoters studied as well as whether its presence reduces competition between promoters and rescues expression.

5.5.5 – Promoter interference/competition: Over-expression of the TF SP1 from a CMV driven promoter and its affect on a competing promoter driving SEAP or GFP expression

In chapter 4, section 4.5.4, all combinations of SEAP and GFP vectors utilizing the SV40, CMV and EF1α promoters were co-transfected at a one-to-one ratio in terms of DNA vector copy number. SEAP protein and mRNA expression was quantified to show the effect of a competing GFP vector and the differential promoter interference/competition effects of the different promoters used. This was repeated again but GFP expression as well as SEAP was measured and a vector expressing the TF SP1, driven by a CMV promoter, was also transfected. This was done due to the results of analyses done in the previous sections in this chapter of the three promoters which showed that the matrix family V$SP1F was present and overrepresented in all three promoters while also being the most prevalent and overrepresented in terms of TFBS modules identified. The idea was that if SP1 and its related TFs are involved in expression from all three promoters it might be a limiting factor when more than one vector is expressed and are in competition for transcriptional resources. This may be especially true for the SV40 due to its potentially heavy reliance on SP1 factors as identified by the analysis of potential TFBSs and also the fact it was the most affected by co-expression of a competing vector/promoter (chapter 4). Transfection of an SP1 vector therefore should lead to an increase in available SP1 protein making it less of a limiting factor and increase or rescue the expression of SEAP or GFP when compared to the results when a CMV-GFP or -SEAP vector is co-transfected. The CMV promoter was chosen initially to drive expression of the SP1 TF because the CMV promoter had been shown previously to affect the expression of a competing promoter more than the SV40 and EF1α promoters.

5.5.5.1 – SV40 promoter

From figure 5.6 – A. we can see that similar to the results in chapter 4 (figure 4.9) that the production of SEAP protein and SEAP mRNA was reduced from the SV40-SEAP vector when a competing GFP vector was co-transfected. The CMV-GFP vector had the biggest negative effect on both SEAP protein and mRNA while the SV40-GFP vector had the least. Co-transfection of CMV-GFP reduced SEAP protein expression to around 54% and SEAP mRNA to about 37% of the control. The reduction in SEAP mRNA was not as great as seen in chapter 4 and one reason for this could be that less DNA was used in these transfections due to the larger size of the recombinant SP1 gene and therefore its vector compared to the other vectors used. When the CMV-SP1 vector was co-expressed with the SV40-SEAP vector the levels of SEAP protein and mRNA expression were higher than for the CMV-GFP and around 79% and 82% of the control. This suggests that over-expression of SP1 can transactivate the SV40-SEAP vector and help increase expression compared to the CMV-GFP vector co-transfection.

As well as analysing SEAP expression, GFP expression was also measured. As well as a SV40-GFP control co-transfection of SV40-GFP and CMV-SP1 was also carried out to see if it had similar effects on GFP as it did on SEAP expression. Only GFP protein expression was measured but from the results in figure 5.6 – B. we can see that GFP expression was more affected then SEAP expression was by the presence of a competing vector/promoter. Co-expression with all three SEAP vectors leads to significant decreases in GFP expression. GFP expression was ~28%, ~17% and ~12% of the SV40-GFP control for the SV40, EF1α and CMV-SEAP vectors respectively. However, when SV40-GFP was co-transfected at a one-to-one ratio with CMV-SP1 GFP expression increased significantly compared to CMV-SEAP and was around 76% of the control. This would further suggest over-expression of SP1 is capable of transactivating the SV40 promoter and helps alleviate some of the competition between the SV40 and CMV promoter.

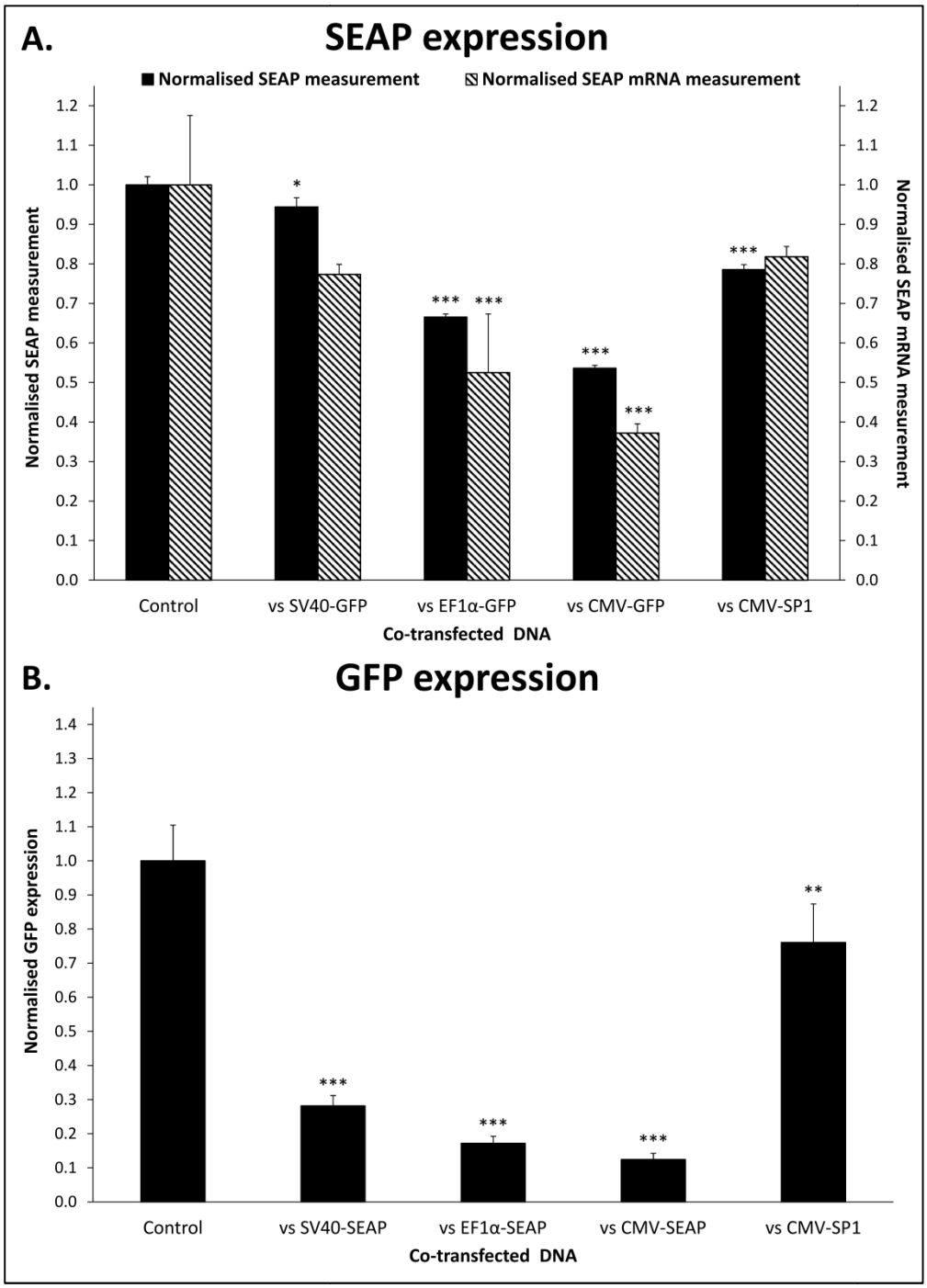

Figure 5.6 – Graph showing the effect on SEAP protein and mRNA expression from the SV40-SEAP vector and GFP expression from the SV40-GFP vector when co-transfected at a 1:1 ratio (vector copy number) with either three different GFP vectors or SEAP vectors and a vector for the expression of the transcription factor SP1 driven by CMV. A. The effects of co-transfection of SV40, EF1α and CMV-GFP and CMV-SP1 with the SV40-SEAP vector at a 1:1 ratio in CHO-S cells. **B.** The effects of co-transfection of SV40, EF1α and CMV-SEAP and CMV-SP1 with the SV40-GFP vector at a 1:1 ratio in CHO-S cells. The 1:1 ratio is in terms of DNA vector copy number rather than total amount of DNA in µg's. 1.06 x 10^{11} copies of the SEAP, GFP and SP1 DNA vectors were used. The total amount of transfected DNA was still, however, kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP expression while SEAP mRNA expression was also quantified after recovering the transfected cells. GFP expression was measured while cells were still in the media and culture plate. All results were normalised to the control sample. Mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, error bars represent + 1 S.D.

5.5.5.2 – EF1α promoter

As with the SV40-SEAP vector the EF1α-SEAP vector was co-transfected with the three different GFP vectors as well as the CMV-SP1 vector. Figure 5.7 – A. shows the results and again like the SV40-SEAP vector the results were similar to the corresponding experiment in chapter 4 in terms of the effects on SEAP mRNA expression. Co-transfection of the CMV-GFP vector again had the biggest negative effect on both SEAP protein and mRNA expression and the SV40-GFP the least. The SV40-GFP had no significant effect on either SEAP protein or mRNA while the EF1α-GFP vector caused a small decrease in SEAP protein (~95% of the control) but a larger decrease in SEAP mRNA (~60% of the control). A reduction to ~84% and ~38% of the control for SEAP protein and mRNA was seen when the CMV-GFP vector was co-transfected. As with the SV40-SEAP vector although the general trend seen was the same as chapter 4 the levels of reduction were again not exactly the same. This will be discussed later although one reason has been given already in the previous sub-section – 5.5.5.1. Co-expression of CMV-SP1 with EF1α-SEAP again, as with the SV40-SEAP vector, increased SEAP expression above that seen for CMV-GFP. SEAP protein expression slightly increased to around 91% of its control while SEAP mRNA increased to about 75%. This suggest that SP1 can play a role in the expression from an EF1α promoter.

As with the SV40-GFP vector the same was done with the EF1α-GFP vector to see the effects of SP1 on GFP as well as SEAP expression (figure 5.7 – B.). GFP expression was reduced more significantly than SEAP when GFP expression was under the control of the EF1α promoter but the same pattern that has been observed in chapter 4 and previously in this chapter was seen once again. Expression from the GFP-EF1α vector was least affected by the SV40-SEAP vector (no decrease), reduced to around 74% of the control when co-transfected with EF1α-SEAP and to around 62% when the CMV-SEAP vector was present. When the CMV-SP1 vector was co-transfected with the EF1α-GFP vector only a small increase compared to co-transfection of CMV-GFP was seen, from ~62% to ~67 of the control. This was much smaller than was seen for the SV40-GFP vector.

One explanation of why the increase in EF1α-GFP expression was negligible when co-expressed with CMV-SP1 compared to the SV40-GFP vector could be the possible

effect of the SP1 TF on cell growth. Over-expression of SP1 in one study showed it to inhibit cell-cycle progression (Deniaud *et al.*, 2009). The mRNA produced from the EF1α promoter can be called a TOP mRNA due to the presence of a 5' TOP at the cap site of the mRNA (Shibui-Nihei *et al.*, 2003). TOP mRNA is more actively translated in growing cells (Hamilton *et al.*, 2006; Meyuhas, 2000; Biberman & Meyuhas, 1997). If SP1 over-expression did inhibit cell-cycle progression then SP1 might still increase the expression of mRNA from the EF1α promoter, as was seen with the increases in SEAP mRNA, but the TOP mRNA it produces might not be as actively translated causing no increase in SEAP or GFP protein expression.

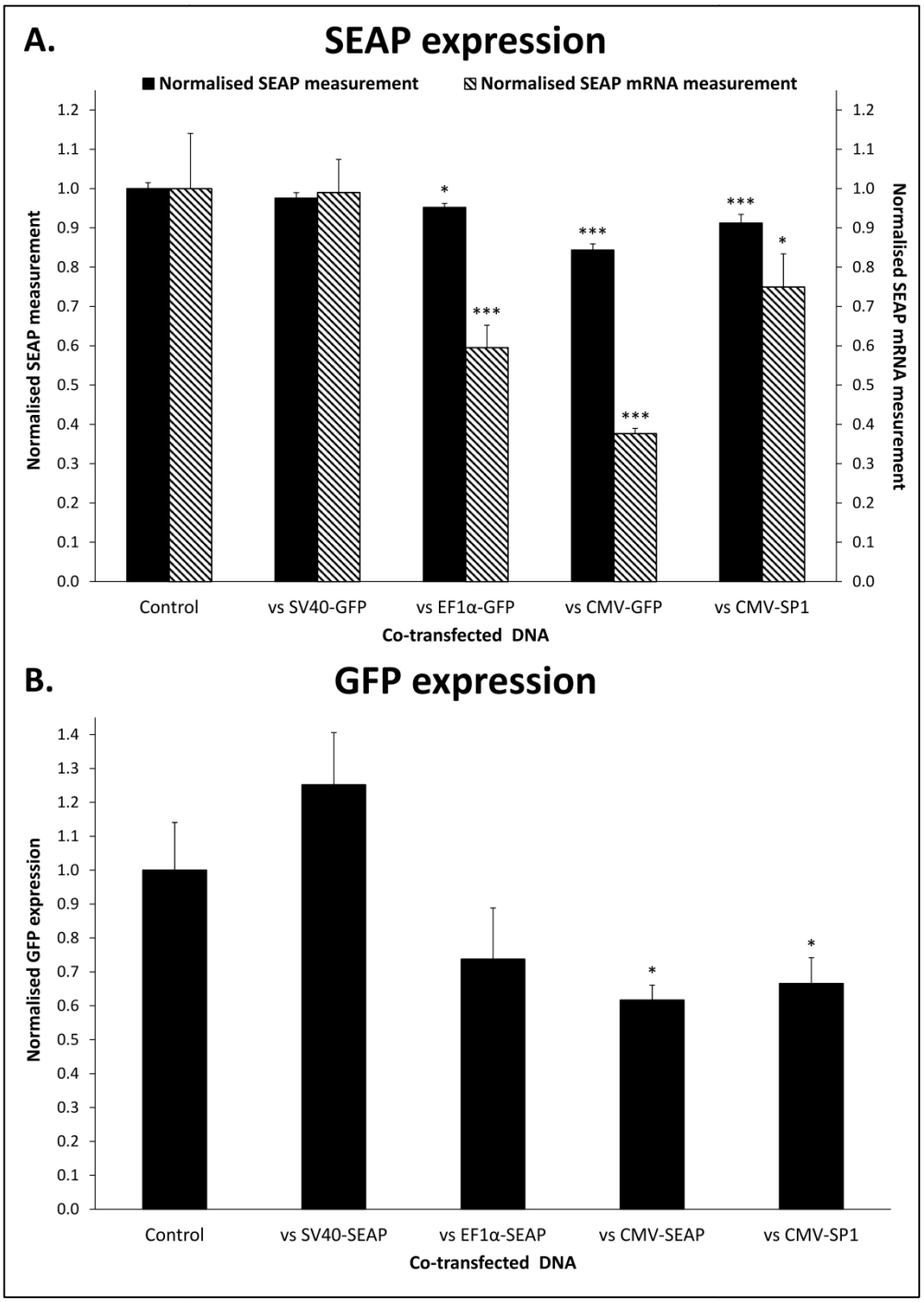

Figure 5.7 – Graph showing the effect on SEAP protein and mRNA expression from the EF1α-SEAP vector and GFP expression from the EF1α-GFP vector when co-transfected at a 1:1 ratio (vector copy number) with either three different GFP vectors or SEAP vectors and a vector for the expression of the transcription factor SP1 driven by CMV. A. The effects of co-transfection of SV40, EF1α and CMV-GFP and CMV-SP1 with the EF1α-SEAP vector at a 1:1 ratio in CHO-S cells. **B.** The effects of co-transfection of SV40, EF1α and CMV-SEAP and CMV-SP1 with the EF1α-GFP vector at a 1:1 ratio in CHO-S cells. The 1:1 ratio is in terms of DNA vector copy number rather than total amount of DNA in µg's. 1.06×10^{11} copies of the SEAP, GFP and SP1 DNA vectors were used. The total amount of transfected DNA was still, however, kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP expression while SEAP mRNA expression was also quantified after recovering the transfected cells. GFP expression was measured while cells were still in the media and culture plate. All results were normalised to the control sample. Mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, error bars represent + 1 S.D.

5.5.5.3 – CMV promoter

As in the previous two sub-sections the experiment was repeated but this time with CMV-SEAP. Figure 5.8 – A. shows the results are similar to chapter 4 and with the same pattern as all the previous experiments. As in chapter 4 the SV40-GFP vector had no negative effect on SEAP protein and mRNA expression from the CMV-SEAP vector. When co-transfected with EF1α-GFP, SEAP protein and mRNA expression were reduced to ~89% and ~88% of their controls, respectively. Co-transfection of the CMV-GFP vector once again had the biggest negative effect with SEAP protein and mRNA being decreased to ~87% and ~52% of their controls. SEAP mRNA expression from the CMV-SEAP vector was increased when co-transfected with the CMV-SP1 vector and compared to the results for co-expression of CMV-GFP, although there was no significant difference in SEAP protein expression. mRNA expression was similar to that seen in the control. As with the other SEAP vectors the increases in SEAP mRNA suggest that the CMV-SEAP vector and its promoter could be directly affected by over-expression of SP1.

The effects on GFP expression from the CMV-GFP vector were also measured when co-transfected with CMV-SP1 and the three SEAP vectors (figure 5.8 – B.). The results for the SEAP vectors were quite similar to those seen on the EF1α-GFP vector. SV40-SEAP had no negative effect while EF1α and CMV-SEAP reduced GFP expression to ~72% and ~64% of the CMV-GFP control. However, when co-transfected with the CMV-SP1 vector the result was more similar to that seen for SV40-GFP vector (figure 5.6 – B.) rather than for EF1α-GFP (figure 5.7 – B.). GFP expression increased compared to CMV-SEAP co-transfection, from ~64% of the control to ~109% of the control. This slight increase was not statistically significant compared to the CMV-GFP control. This result is further evidence that the CMV promoter can respond to the SP1 over-expression.

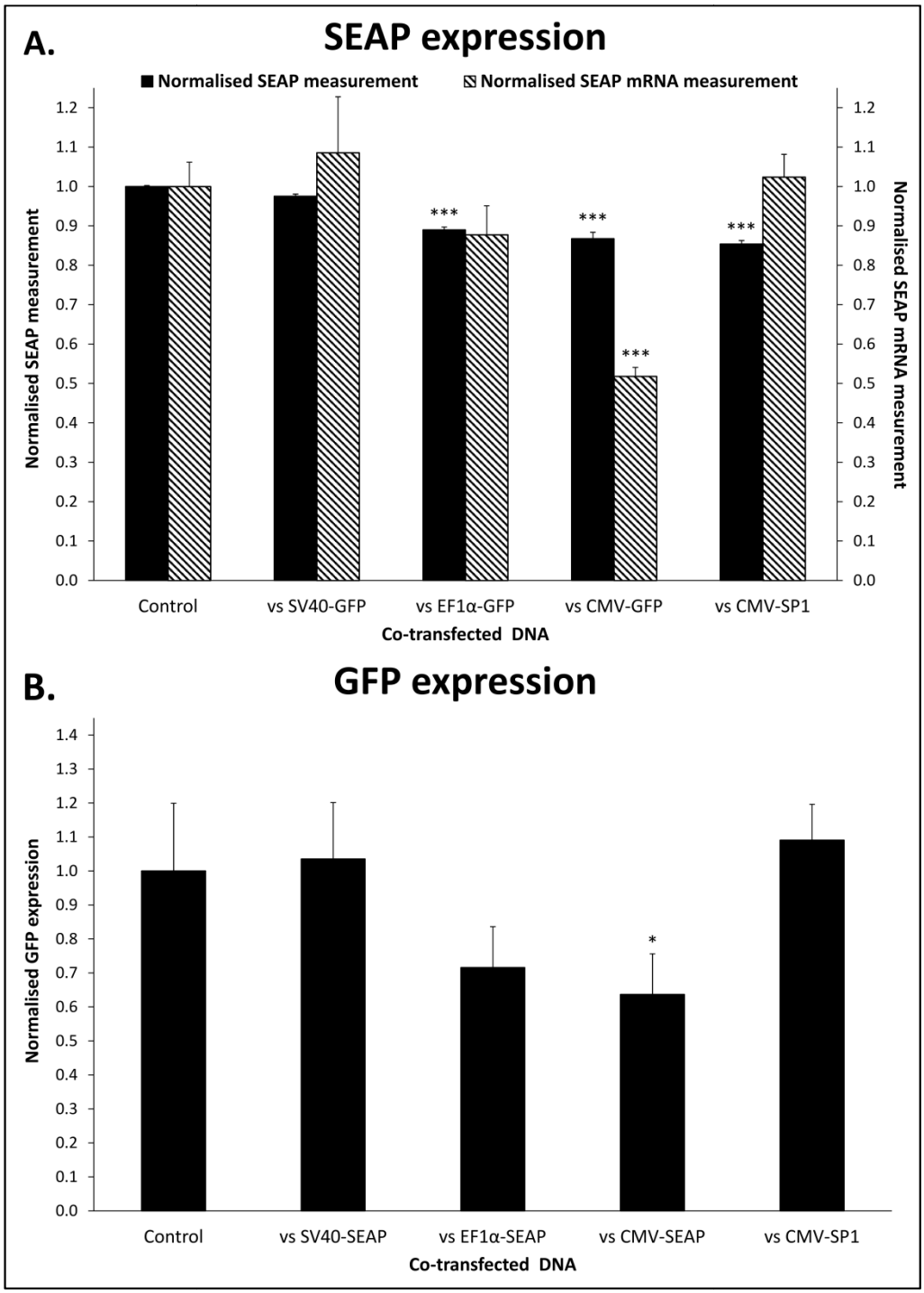

Figure 5.8 – Graph showing the effect on SEAP protein and mRNA expression from the CMV-SEAP vector and GFP expression from the CMV-GFP vector when co-transfected at a 1:1 ratio (vector copy number) with either three different GFP vectors or SEAP vectors and a vector for the expression of the transcription factor SP1 driven by CMV. A. The effects of co-transfection of SV40, EF1α and CMV-GFP and CMV-SP1 with the CMV-SEAP vector at a 1:1 ratio in CHO-S cells. **B.** The effects of co-transfection of SV40, EF1α and CMV-SEAP and CMV-SP1 with the CMV-GFP vector at a 1:1 ratio in CHO-S cells. The 1:1 ratio is in terms of DNA vector copy number rather than total amount of DNA in µg's. 1.06×10^{11} copies of the SEAP, GFP and SP1 DNA vectors were used. The total amount of transfected DNA was still, however, kept consistent in all transfections by using the empty –ve control vector. Cell media was collected 48 hours post-transfection and analysed for SEAP expression while SEAP mRNA expression was also quantified after recovering the transfected cells. GFP expression was measured while cells were still in the media and culture plate. All results were normalised to the control sample. Mean values significantly different (Dunnett's test) from their control are indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). N = 3, error bars represent + 1 S.D.

5.5.5.4 – Comparisons and conclusions

The results from this set of experiments had the same pattern as the similar experiments carried out in chapter 4 in so much that SV40-SEAP vector was again the most negatively affected in terms of both SEAP protein and mRNA expression while the CMV-SEAP was the least affected when co-expressed with a competing GFP vector (table 5.19). The CMV-GFP vector was shown to be the GFP vector which most negatively affected SEAP expression and the SV40-GFP had the least affect. The EF1α-GFP had a much greater negative influence on SEAP expression then SV40-GFP but not to the same extent as the CMV-GFP vector.

Although there were similar patterns seen between these experiments and the experiments in chapter 4 there were subtle differences. Mainly that decreases in SEAP protein expression on the whole were greater in this chapter than in the previous. For SEAP mRNA expression the opposite was true. This could be due to a few possible reasons but the main difference between the experiments in this chapter and the last is the number of copies of each vector which were used. In the last chapter 1.66×10^6 copies of each vector were transfected but in this chapter this was reduced to 1.06×10^6 copies, a 1.57-fold reduction. The reason for this was that the recombinant SP1 gene which was transfected was much larger then both the SEAP and GFP genes. The change in the number of copies of both SEAP and GFP transfected is bound to have some affect on the promoter interference/competition between competing vectors and on SEAP protein and GFP expression. A reduction in DNA vector copies present will likely put less pressure on the cells transcriptional machinery to some extent which will affect promoter competition and mRNA production. In turn this will also affect protein expression but this affect will not necessarily mirror the change in mRNA expression. In summary, the levels of promoter interference/competition will change depending on the amount of DNA or copies of DNA vectors transfected and changes in mRNA and protein will not necessarily correlate with each other and this had been shown previously in this thesis especially in terms of SEAP mRNA and protein expression.

As well as measuring SEAP expression when co-transfecting the SEAP and GFP vectors, GFP expression was measured and compared to a GFP control for each promoter used

(table 5.19). GFP expression was reduced when in the presence of a competing SEAP vector compared to its control. The reductions in GFP expression, in terms of fold change from their control, were greater then what was seen for SEAP expression but once again there were familiar patterns in the results. The GFP vector containing the SV40 promoter saw the biggest negative fold changes in GFP expression when co-transfected with the SEAP vectors. These changes followed the familiar pattern of the SV40 promoter containing SEAP vector having the least affect (-3.55-fold change), the EF1α-SEAP vector having a greater effect (-5.83-fold change) then the SV40 vector and the SEAP vector utilizing the CMV promoter having the greatest negative effect (-8.04-fold change). A similar pattern was seen for both the EF1α and CMV-GFP vectors but the reductions in GFP expression were not as great as seen for the SV40-GFP vector. The SV40-SEAP vector had no negative effect on GFP expression from either the EF1α or CMV-GFP vectors while the EF1α-SEAP vector reduced GFP expression 1.36 and 1.40-fold and the CMV-SEAP reduced expression 1.62 and 1.57-fold, respectively for these two GFP vectors. As with the SEAP results we can see that the presence of a competing CMV promoter has the biggest negative influence while a competing SV40 promoter has the least.

The most interesting result and the main purpose of this set of experiments however, was the effect of co-transfection of an SP1 expression vector. The SP1 vector utilized the CMV promoter and this was chosen because we had already seen in chapter 4 that the CMV promoter had the greatest negative effects on a competing promoter. We hypothesized from the earlier work in this chapter that the SP1 TF might be involved in expression or at least capable of binding all three of the promoters used in this study. SP1 has already been shown to bind and activate the SV40 promoter (Saffer *et al.*, 1990; Dynan & Tjian, 1983).

When CMV-SP1 was co-expressed with the SEAP vectors we saw an increase in SEAP mRNA when compared to the results of co-transfection of the CMV-GFP vector for all SEAP vectors (table 5.19). For all three SEAP vectors the amount of SEAP mRNA expressed roughly doubled compared to the amount seen when co-expressed with CMV-GFP. This could be evidence of transactivation of these promoters by SP1. For the SV40-SEAP this also involved an increase in SEAP protein expression but this

increase was still below the level of SEAP protein expression seen for its control. SEAP protein expression increased slightly when the EF1α-SEAP vector was co-transfected with CMV-SP1 compared to CMV-GFP but there was no increase seen in SEAP protein expression from the CMV-SEAP vector.

CMV-SP1 had a greater effect on GFP expression especially for the SV40 and CMV-GFP vectors (table 5.19). GFP expression when co-expressed with CMV-SP1 increased compared to co-transfection of CMV-SEAP. For the EF1α-GFP vector this increase was negligible (1.08-fold increase), while for the CMV-GFP vector there was 1.71-fold increase in GFP expression which even slightly surpassed the level of the CMV-GFP control. However, by far the biggest increase in GFP expression was seen when CMV-SP1 was co-transfected with SV40-GFP. GFP expression was 6.11-fold higher than when CMV-SEAP was co-expressed.

Taking into account both of the SEAP and GFP results when CMV-SP1 was co-transfected it would seem that SP1 does play some role in the expression of these promoters to some extent or at least in the competition between promoters. It has already been shown that SP1 is an important factor especially for the SV40 promoter (Saffer *et al.*, 1990; Dynan & Tjian, 1983) and while it also known to be involved in CMV promoter expression it is slightly less important than for the SV40 but only in terms of it being one of multiple TFs, such NF-κB and ATF/CREB factors, which can also influence CMV transcription (Bakovic *et al.*, 2000; Luu & Flores, 1997; Yurochko *et al.*, 1997; Olive *et al.*, 1990). SP1 have also been shown previously to bind to the EF1α promoter and possibly play some role in its regulation (Wakabayashiito & Nagata, 1994). The results from our experiments, especially the increases in SEAP mRNA seen, add further evidence that SP1 plays a role in transcription from all these promoters. It also points to competition for this TF as possibly being one of the causes of promoter interference/competition, i.e. – the competition for a shared TF or TFs.

Further work is needed to confirm though that the increases seen in SEAP and GFP when CMV-SP1 was co-expressed compared to a CMV GFP or SEAP vector was the result of SP1 over-expression rather than just the absence of GFP or SEAP. This should first involve confirmation that the recombinant SP1 is actually expressed in our cells. It

then may involve the co-transfection of different amounts of decoy DNAs which could compete not only for endogenous SP1 protein but also any additional recombinant SP1 protein. Although the fact that the CMV-SP1 vector itself also contains SP1 sites could further complicate the results and their interpretation. It would also be informative to express SP1 from vectors using the SV40 and EF1α promoter to see how the results differ from the CMV-SP1 vector.

Although we have seen that SP1 is likely to be a TF that plays some role in the regulation of all three of these promoters and therefore also competition between them it will not be the only important factor. Other factors will also be involved in expression from these promoters and it is also likely that both shared and unshared factors will have an influence on the levels of not only expression but also promoter interference/competition. The greater promoter strengths of the CMV and EF1α promoters are likely to rely on both their increased variety of matrix families and TFBSs and also the possibility that some of these factors or combinations of these factors may be stronger transcriptional inducers then SP1. Future work for further investigation of the interactions between competing promoters will be presented in the next section.

Table 5.19 – Fold change in SEAP protein and mRNA expression and GFP expression when the different SEAP and GFP vectors are co-transfected at a one-to-one ratio plus the effect of co-transfection of an SP1 expressing vector

	SEAP expression					
	SV40-SEAP		EF1α-SEAP		CMV-SEAP	
GFP vector/SP1	Protein	mRNA	Protein	mRNA	Protein	mRNA
SV40-GFP	-1.06	-1.29	-1.02	-1.01	-1.03	+1.09
EF1α-GFP	-1.50	-1.90	-1.05	-1.68	-1.12	-1.14
CMV-GFP	-1.87	-2.69	-1.19	-2.66	-1.15	-1.93
CMV-SP1	-1.27	-1.22	-1.09	-1.33	-1.17	+1.02

	GFP expression		
SEAP vector/SP1	SV40-GFP	EF1α-GFP	CMV-GFP
SV40-SEAP	-3.55	+1.25	+1.04
EF1α-SEAP	-5.83	-1.36	-1.40
CMV-SEAP	-8.04	-1.62	-1.57
CMV-SP1	-1.31	-1.50	+1.09

The numbers in the top part of the table represent the fold change in SEAP protein and mRNA expression when compared to the control for the respective SEAP vector. The numbers in the bottom part of the table represent the fold change in GFP expression when compared to the control for the respective GFP vector. SEAP, GFP and SP1 vectors were co-transfected at a one-to-one ratio in terms of DNA vector copy number (1.06×10^{11} copies of each vector). Fold changes come from the data in figures 5.6 (SV40-SEAP & -GFP), 5.7 (EF1α-SEAP & -GFP) and 5.8 (CMV-SEAP & -GFP).

5.6 – Further Discussion, Final Conclusions and Future Directions

5.6.1 – Further discussion

In this chapter, a promoter analysis of three commonly used promoters for biopharmaceutical production was carried out to help answer two main questions. Which TFs and TFBSs are potentially important for the function and regulation of the SV40, CMV and EF1α promoters and how they may affect interference/competition between these promoters?

Using a variety of programs from the Genomatix Software Suite we identified potential TFBS in each of the three promoters. The initial searches using MatInspector gives a large number of predicted TFBSs. Many of these will not be functional. However it did show that the weakest promoter the SV40 had less predicted TFBSs and that these also came from fewer matrix families. In terms of both the potential TFBSs and their matrix families the SV40 shared a greater proportion of both with the CMV and EF1α promoters. Of the matrix families found to be overrepresented in the SV40 promoter a greater proportion of these were also overrepresented in either or both the other two promoters then was the case for either the CMV or EF1α promoters.

ModelInspector searches of the promoter sequences gave more information on TFBSs that may be functionally important in these promoters. There were major differences between the SV40 and both the CMV and EF1α promoters. Although the SV40 had a higher number of model matches these matches were composed of a smaller number of different modules and different TFBSs. The search seemed to show that the SV40 promoter might have a bias to SP1 TFs as they were involved in the majority of modules identified and had the highest number of TFBSs. Both the other two promoters had a greater number of different modules and different TFBSs. The SV40 promoter again was shown to share a greater proportion of its TFBSs and the respective matrix families with both the CMV and EF1α promoters then both these latter two promoter did. The V$SP1F family was only matrix family to be overrepresented in all three promoters from the results of the ModelInspector search.

The results of these searches suggest that the SV40 may be a weaker promoter due to containing less potential TFBSs and that it is also the most negatively by promoter interference because it has less TFs from different matrix families but while also sharing a large proportion of these sites with both the CMV and EF1α promoters. The greater variety of potential TFBSs in both the CMV and EF1α may lead to them being both stronger promoters and also more capable at competing with another for expression.

These searches also highlighted that SP1 factors were potentially important TFs in the expression for all three promoters. This is in agreement with the published literature which has shown SP1 to play a role in transactivation of both the SV40 and CMV promoters (Wu *et al.*, 1998; Luu & Flores, 1997; Yurochko *et al.*, 1997; Saffer *et al.*, 1990; Dynan & Tjian, 1983). Over-expression of SP1 was shown to rescue expression of from SEAP and GFP vectors caused by promoter interference/competition from the CMV promoter. This showed that SP1 can possibly play a role in the expression from all the promoters studied. Indeed SP1 is important in both the basal and constitutive activation of many endogenous promoters (Infantino *et al.*, 2011; Kaczynski *et al.*, 2003). However, it is unlikely that the competition for SP1 will be the only source of promoter competition. TFs for the other TFBSs identified by ModelInspector and shared between the three promoters will also be a source of competition and might also have a greater impact. This includes factors such as KLF factors, ETS factors, AP1 factors and NF-kB factors.

KLF TFs as already mentioned earlier are highly related to the SP1 TFs and are important in many cellular processes (McConnell & Yang, 2010). They bind similar sequences and depending on the cellular context and promoter a KLF factor can act as both a transcriptional activator or repressor (Kaczynski *et al.*, 2003). KLF11 has been shown to activate the SV40 promoter (Martin *et al.*, 2000).

ETS TFs are a large family of TFs (27 member) and have roles in apoptosis, proliferation, development, differentiation, angiogenesis, tissue remodelling and cancer (Shaikhibrahim & Wernert, 2012; Meadows *et al.*, 2011; Randi *et al.*, 2009;

Hashiya *et al.*, 2004). The family member ERF can repress expression from the CMV promoter (Bain *et al.*, 2003).

AP-1 factors play roles in cell proliferation and survival (Shaulian & Karin, 2001) and has also been shown to activate the SV40 promoter (Harshman *et al.*, 1988). They have also been shown to co-activate the CMV promoter (Isern *et al.*, 2011; Lee *et al.*, 2004; Lembo *et al.*, 1994).

NF-kB factors are a very widely studied family with roles in inflammation and cancer (Hoesel & Schmid, 2013) and it has been shown to activate the SV40 (Tabakin-Fix *et al.*, 2004; Espel *et al.*, 1990; Kawakami *et al.*, 1988) and CMV promoter (He & Weber, 2004; Lee *et al.*, 2004; Prosch *et al.*, 1995).

The knowledge that these factors have been shown to already function in either the SV40 or CMV promoter or both is further evidence they could be potential sources of competition. Also the fact that these factors were identified by ModelInspector shows the power of the program to identify potentially functional TFBSs. Whether the modules identified in each promoter function in the same manner as they do in the studies they were taken from is unknown. Promoter interference/competition could be further studied by the over-expression of more of these potentially important transcription factors and each promoter's response might reflect the importance of that factor for its regulation.

As well as competition for TFs such as the ones previously mentioned it is likely that competition will occur not only occur for them but that there will also be competition for general TFs, transcriptional machinery and the initiation of transcription. Even though there are common TFBSs and TFs in all three of the promoters studied they will initiate transcription using different sets of TFs and the strength of transcriptional initiation is likely to vary. This will also be a source of promoter interference/competition.

The fact that these promoters use multiple TFBSs and also the fact that they can interfere considerably with the expression of another promoter make these promoters less than ideal candidates for expression of multiple genes at differing expression

levels. The creation of synthetic promoters which rely on only a small number of TFs and where promoters differ in the transcription factors they bind may help create promoters which not only differ in strength but also removes one source of promoter interference/competition and potentially allow different promoters to function more independently and maintain their own levels of expression even in the presence of another promoter. If this is achievable this could potentially be another method for controlling the expression of multiple genes in the same system.

5.6.2 – Final conclusions

A MatInspector search of SV40 promoter showed it to contain less numbers and variety of potential TFBSs then both the CMV and EF1α. It also shared a greater a proportion of its potential TFBSs with the CMV and EF1α promoters. Both these could be indications of why this promoter was weaker and also more negatively affected by the presence of a stronger promoter.

When searching the promoters using ModelInspector the SV40 had a greater number of matches returned then both the CMV and EF1α promoter. However these matches were composed of TFBSs from a smaller number of matrix families and also less TFBSs in total. These TFBSs were also shared to a greater proportion with the sites found in both the CMV and EF1α promoter. The CMV promoter shared less of its potential TFBSs identified using ModelInspector than both SV40 and EF1α promoters. This could be an indication of why the CMV promoter was both stronger and interfered with the expression of the SV40 and EF1α more than they did with expression from the CMV promoter.

One source of promoter interference/competition could be the competition for a limited amount of a shared TF. V$SP1F TFBSs were shown to be not only involved in TFBS modules for all three promoters they were also considered to be overrepresented. This made V$SP1F TFs one potential source of competition. Co-expression of a vector for the expression of the TF SP1 under the control of the CMV promoter increased SEAP mRNA and GFP expression from SEAP and GFP vector utilizing all three promoters when compared to co-expression of a CMV GFP or SEAP

vector. This suggests that when there is an excess of this TF promoter interference caused by a CMV promoter is reduced.

However, promoter interference is unlikely to be caused by competition for a single shared TF, such as SP1, it will involve multiple TFs and also the rate at which each promoter can initiate transcription which will be related to the level of transcriptional activation caused by the different combinations of TF binding in the different promoters. The clearance or retention of bound TFs to the promoters will also affect promoter competition as this will impact on the available pools of TFs.

Although these promoters are useful because they function in many cell lines they will not be sufficient for the controlled expression of multiple genes due to the presence of multiple different TFBSs for multiple TFs suggesting they have redundancy built into them making their expression hard to control. New synthetic promoters will be needed and one idea for a potential strategy for controlled expression of multiple genes is presented in the next section.

5.6.3 – Future directions

To further understand which TFs are important for the regulation of all three of these promoters it will be necessary to express more than just the SP1 TF. TFs from the AP-1, KLF, ETS and NF-κB families could be used as these were also involved in modules for all three promoters.

Alternatively decoy DNAs (Brown *et al.*, 2013) could be co-expressed to limit the availability of specific TFs. The level of reduction seen from a promoter when a decoy for specific TF is used would be a reflection of the importance of that TF.

The fact that all three of these promoters contain multiple TFBSs for multiple different TFs and the fact the CMV promoter can severely interfere with expression of another promoter, especially a weaker promoter such as the SV40, make using combinations of these promoters ill suited for the expression of multiple genes were discrete control of differing levels of expression may be required.

An expression system which uses engineered TALEs (transcription activator–like effectors) and synthetic promoters capable of binding them could be one potential strategy to control the expression of multiple genes. TALEs have the advantage that they have two domains, one which can be engineered to bind almost any DNA sequence and another which can be made to recruit transcriptional machinery (Perez-Pinera *et al.*, 2013; Cermak *et al.*, 2011; Miller *et al.*, 2011; Zhang *et al.*, 2011; Boch *et al.*, 2009; Moscou & Bogdanove, 2009). Multiple TALEs could be engineered to bind multiple specific sequences in a library of synthetic promoters. It may then be possible to control gene expression multiple ways such as altering the amount of TALE expressed, the number of a specific TALE binding sequences in a synthetic promoter, the use of multiple TALEs and the presence of multiple TALE binding sequences in synthetic promoters.

Chapter 6 – Thesis Discussion and Future Perspectives

6.1 – Discussion and future perspectives

In this thesis the development of a tunable protein expression system for use in mammalian cells has been presented along with a comparison of three commonly used promoters (SV40, CMV and EF1α), their effects on expression from a competing co-expressed promoter (promoter interference/competition), their potential TFBSs and how common TFBSs could be one source of this promoter interference/competition. Each results chapter had its own discussion and this section will not to cover all the same points and instead show how this research fits in with the current knowledge and the future challenges of biopharmaceutical production in mammalian cells.

There are numerous published studies presenting a variety of cell engineering attempts to improve recombinant protein production in mammalian cells. This has involved the over-expression/suppression of genes/proteins to improve cell growth and survival as well as to improve the production capacity (protein synthesis and secretion) of mammalian cells for increased recombinant protein production (Lee *et al.*, 2013; Nishimiya *et al.*, 2013; Pybus *et al.*, 2013; Dreesen & Fussenegger, 2011; Peng *et al.*, 2011; Ku *et al.*, 2010; Hwang & Lee, 2009; Kuystermans & Al-Rubeai, 2009; Peng & Fussenegger, 2009; Majors *et al.*, 2008b; Mohan *et al.*, 2007). The main goals for increasing the production of a therapeutic recombinant protein is to lower the cost of production for recombinant proteins which are already well expressed and to improve the expression of proteins whose low expression make them unviable for large scale production. Recombinant proteins that have low production titres can be termed "difficult-to-express" (DTE) proteins. DTE proteins can suffer problems in one or more of the following processes – transcription, translation, protein folding/assembly and secretion. This can result in bottlenecks in recombinant protein production and a specific protein can be over-expressed to try and remove this bottleneck. Although this can sometimes prove successful the expression of a single protein can sometimes just move the bottleneck. Therefore the expression of multiple

genes/proteins will be required to improve the production of some DTE proteins. Although the expression of multiple genes has been carried out (Lee *et al.*, 2013; Nishimiya *et al.*, 2013) many studies just involve the high level expression of two or more genes using strong constitutive promoters and the results can be both cell line and protein specific. The expression of proteins such as XBP1 and mTOR are considered good targets for over-expression due to the fact they have influence over multiple target genes but both proteins, especially XBP1, will also have cell line and product specific effects (Ku *et al.*, 2008). In both these attempted strategies there is a lack of discrete control over the expression of the proteins over-expressed and due to the cell line and protein specificity of many cell engineering strategies there is unlikely to be generic solution which can be used for all cell lines and recombinant proteins. Therefore, there will be a need for gene/protein expression systems were the expression of multiple genes can be controlled to reflect the requirements of a particular cell line or a specific product. As well as using such systems for the improved expression of DTE proteins this will also be applicable to already high producing cell lines were there might be a need to turn these cell lines into so called super producers to further lower the cost of production.

The work in this thesis has been carried out largely with the view of multi-gene expression technology for mammalian cells producing recombinant proteins in mind. In chapter 3 a tunable protein expression system for multiple genes was presented. Controlled gene expression has mainly been focused on promoters whose expression can be turned on or off given the correct stimulus (Blount *et al.*, 2012; Hurley *et al.*, 2012). Although complex synthetic gene regulatory networks have been created most of the work on such systems has been carried out in prokaryotes or in eukaryotic microbial systems. Complex synthetic gene circuits for use in mammalian cells have been created (Muller *et al.*, 2013; Auslander *et al.*, 2012) but there no published studies on their application in biopharmaceutical production. The complexity of such circuits may also mean that it is difficult for researchers who are not experts in mammalian synthetic biology to adapt them for their own applications. Many of the circuits also do not necessarily allow the simultaneous expression of multiple genes at different but chosen levels of expression. We believe the system we have developed is

a simpler method for the controlled/tunable expression of multiple genes at multiple different levels of expression.

As stated earlier the majority of cell line engineering strategies use strong constitutive promoters to express proteins. Depending on the protein and its effect on the cell, high expression may have negative rather than positive consequences such as shifting the bottleneck in protein production or severely inhibiting cell growth. Expression may be manipulated by lowering the amount of transfected vector DNA for expression of the protein, but with strong promoters such as the CMV the window for manipulating between high and low expression can be quite small and make attaining intermediate levels of expression difficult. This could be solved by using a weaker promoter such as the SV40 but if more than one protein is to be expressed and high levels of expression of one protein is required, while intermediate levels of the other is also needed problems will be encountered if a strong promoter is used for high expression and a weaker promoter used of intermediate expression. Promoter interference is likely to occur and complicate matters further which has also been shown in this thesis. The system presented in chapter 3 took advantage of promoter interference to increase the range of expression. The ERSE sequences present in the reporter protein vectors allowed these vectors to not only maintain the intermediate expression of a reporter protein when its binding factor ATF6(50) was co-expressed, in some situations (depending on the amount of ATF6(50) expressed and the number of ERSE used in the reporter vector), it also allowed it to achieve levels of expression as high as that seen from a CMV driven vector. Using both the absence and presence of differing numbers of ERSEs and also differing the amount of ATF6(50) co-expression high levels of SEAP protein expression was achieved along with intermediate (or medium) levels of GFP expression.

It is envisioned that the system could be further developed for situations where multiple bottlenecks have been discovered in the production of a recombinant protein (i.e. – a DTE proteins) and there is need for the expression of multiple genes to help resolve them. It is unlikely that the same level of expression will be required for all genes expressed and it is also unlikely that the optimal levels of gene expression will be achieved through manipulating the numbers of copies of each transfected gene

vector, especially if strong constitutive promoters such as the CMV are used. Using the system presented in chapter 3 it will make it easier to find the optimum levels of expression for each gene due to the control in expression of multiple genes given by the system. A system such as this one is likely to be restricted to use initially in transient expression experiments. However, transient expression is vitally important for producing a rapid supply of a pre-clinical recombinant protein and also for initially testing a cell line engineering strategy. Large scale transient production is also being studied to try and make it a viable means of production (Baldi *et al.*, 2007). The use of this system will not necessarily be restricted to the improvement of biopharmaceutical production but could also be applicable in other biotech areas as well.

An alternate strategy for the expression of multiple genes could be through the use of multiple synthetic promoters which do not severely interfere with the expression of another promoter. It was shown that three commonly used promoters, the SV40, CMV and the EF1α promoters all interfere with expression of a competing promoter and the strength of interference was related to the transcriptional strength of the promoter. The SV40 promoter was shown to be the weakest promoter and the most negatively affected by the presence of another promoter. All three promoters contained potential TFBSs for multiple TFs. The SV40 had the least number and variety of TFBSs and also shared a greater number of its potential TFBS with both the CMV and EF1α. Its limited choice of TFs compared to both the CMV and EF1α promoter may be one reason why this promoter was relatively weaker and because it shared TFBSs with these promoters one reason why it was most affected by competition from another promoter. The large numbers of potential TFBSs found in all three promoters, although a large proportional of these may not be functional, it did show the potential complexity of the regulation of these promoters. This will make it difficult to predictably control there expression. Strong promoters such as the CMV also have a small window of expression again making control gene expression a challenge. Once more than one promoter is added to a system, promoter interference/competition will occur and the problems of control increase. This we likely further increase with the addition of more promoters. Although both the CMV and SV40 promoters were used in the expression system present in chapter 3 the use of ATF6(50) and ERSEs were

needed to manipulate expression. SV40, CMV and EF1α when used in there standard forms will not be suited to for future engineering strategies for the controlled expression of multiple genes. There may also be difficulties when trying to re-engineer them and it is likely that a "bottom up" synthetic biology approach will be needed to create more suitable promoters for multi-gene expression systems.

The design of synthetic promoters which do not share binding sites for TFs may be one approach to create promoters which do not suffer from competition for shared TFs. However, due to the use of different TFBSs there could still be promoter competition between synthetic promoters due to the different endogenous TFs potentially having different levels of transcriptional activation and therefore differing in their abilities to recruit factors involved in the initiation of transcription. It may be possible to overcome this by controlling the expression from a synthetic promoter using a similar strategy as in chapter 3, but through use of a synthetic transcriptional activator. At the end of chapter 5 the possible use of TALEs was mentioned for the controlled transcriptional activation of synthetic promoters. A system using synthetic TFs (TALEs) and synthetic promoters may also be capable of functioning more predictably in a number of cell lines more so than synthetic promoters whose expression relies on endogenous TFs whose expression will differ significantly. The ability to customize TALEs means they can also act as a transcriptional repressor which further increases the capabilities of controlling expression.

In the future there will be an increasing need for any expression system that controls the expression of multiple genes for production of recombinant proteins in CHO cells for developments in other technologies that increase our understanding of recombinant protein production.

The greatest aid in this will be the further development of –omics approaches. These approaches will not only help in the identification of genes/proteins as targets for manipulation but also in analysing the effects of future gene engineering strategies. The publication of the CHO genome was a landmark moment for researchers in the CHO community (Xu *et al.*, 2011). Since then genomes from other CHO cell lines as well as the genome of *Cricetulus griseus* have also been sequenced and made publicly

available (Brinkrolf *et al.*, 2013; Lewis *et al.*, 2013). All this information is also made easily available at http://www.chogenome.org/ with the being continually updated (Hammond *et al.*, 2012). Publication of the CHO genome has also helped in the progression of other –omics approaches such as transcriptomics, proteomics and metabolomics. All these approaches will not only aid in the design and implementation of multi-gene engineering strategies in CHO cells but also in every area of CHO cell research.

Reference List

Acosta-Alvear, D., Zhou, Y., Blais, A., Tsikitis, M., Lents, N.H., Arias, C., Lennon, C.J., Kluger, Y., & Dynlacht, B.D., (2007)

XBP1 controls diverse cell type- and condition-specific transcriptional regulatory networks

Molecular Cell, 27, 1, 53-66

Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurerfogy, I., & Busslinger, M., (1992)

Pax-5 encodes the transcription factor Bsap and is expressed in lymphocytes-B, the developing CNS, and adult testis

Genes & Development, 6, 9, 1589-1607

Adhya, S. & Gottesman, M., (1982)

Promoter Occlusion - transcription through a promoter may inhibit its activity

Cell, 29, 3, 939-944

Alper, H., Fischer, C., Nevoigt, E., & Stephanopoulos, G., (2005)

Tuning genetic control through promoter engineering

Proceedings of the National Academy of Sciences of the United States of America, 102, 36, 12678-12683

Altamirano, C., Illanes, A., Becerra, S., Cairo, J.J., & Godia, F., (2006)

Considerations on the lactate consumption by CHO cells in the presence of galactose

Journal of Biotechnology, 125, 4, 547-556

Altamirano, C., Illanes, A., Casablancas, A., Gamez, X., Cairo, J.J., & Godia, C., (2001)

Analysis of CHO cells metabolic redistribution in a glutamate-based defined medium in continuous culture

Biotechnology Progress, 17, 6, 1032-1041

Archer, M.C., (2011)

Role of sp transcription factors in the regulation of cancer cell metabolism

Genes & Cancer, 2, 7, 712-719

Atochina, O., Mylvaganam, R., Akselband, Y., & McGrath, P., (2004)

Comparison of results using the gel microdrop cytokine secretion assay with ELISPOT and intracellular cytokine staining assay

Cytokine, 27, 4-5, 120-128

Auslander, S., Auslander, D., Muller, M., Wieland, M., & Fussenegger, M., (2012)

Programmable single-cell mammalian biocomputers

Nature, 487, 7405, 123-127

Avni, D., Biberman, Y., & Meyuhas, O., (1997)

The 5' terminal oligopyrimidine tract confers translational control on TOP mRNAs in a cell type- and sequence context-dependent manner

Nucleic Acids Research, 25, 5, 995-1001

Baik, J.Y., Ha, T.K., Kim, Y.H., & Lee, G.M., (2011)

Proteomic understanding of intracellular responses of recombinant chinese hamster ovary cells adapted to grow in serum-free suspension culture

Biotechnology Progress, 27, 6, 1680-1688

Bailey, C.G., Tait, A.S., & Sunstrom, N.A., (2002)

High-throughput clonal selection of recombinant CHO cells using a dominant selectable and amplifiable metallothionein-GFP fusion protein

Biotechnology and Bioengineering, 80, 6, 670-676

Bain, M., Mendelson, M., & Sinclair, J., (2003)

Ets-2 Repressor Factor (ERF) mediates repression of the human cytomegalovirus major immediate-early promoter in undifferentiated non-permissive cells

Journal of General Virology, 84, 41-49

Bakovic, M., Waite, K.A., & Vance, D.E., (2000)

Functional significance of Sp1, Sp2, and Sp3 transcription factors in regulation of the murine CTP:phosphocholine cytidylyltransferase alpha promoter

Journal of Lipid Research, 41, 4, 583-594

Baldi, L., Hacker, D.L., Adam, M., & Wurm, F.M., (2007)

Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives

Biotechnology Letters, 29, 5, 677-684

Banerji, J., Rusconi, S., & Schaffner, W., (1981)

Expression of a Beta-globin gene is enhanced by remote SV40 DNA-sequences

Cell, 27, 2, 299-308

Baserga, R., Resnicoff, M., D'Ambrosio, C., & Valentinis, B., (1997)

The role of the IGF-I receptor in apoptosis

Vitamins & Hormones, 53, 65-98

Basu, M. & Roy, S.S., (2013)

Wnt/beta-catenin pathway is regulated by PITX2 homeodomain protein and thus contributes to the proliferation of human ovarian adenocarcinoma cell, SKOV-3

Journal of Biological Chemistry, 288, 6, 4355-4367

Bateman, B.T., Donegan, N.P., Jarry, T.M., Palma, M., & Cheung, A.L., (2001)

Evaluation of a tetracycline-inducible promoter in Staphylococcus aureus in vitro and in vivo and its application in demonstrating the role of sigB in microcolony formation

Infection and Immunity, 69, 12, 7851-7857

Bebbington, C.R., Renner, G., Thomson, S., King, D., Abrams, D., & Yarranton, G.T., (1992)

High-level expression of a recombinant antibody from myeloma cells ssing a glutamine-synthetase gene as an amplifiable selectable marker

Bio-Technology, 10, 2, 169-175

Becker, E., Florin, L., Pfizenmaier, K., & Kaufmann, H., (2008)

An XBP-1 dependent bottle-neck in production of IgG subtype antibodies in chemically defined serum-free Chinese hamster ovary (CHO) fed-batch processes

Journal of Biotechnology, 135, 2, 217-223

Becker, E., Florin, L., Pfizenmaier, K., & Kaufmann, H., (2010)

Evaluation of a combinatorial cell engineering approach to overcome apoptotic effects in XBP-1(s) expressing cells

Journal of Biotechnology, 146, 4, 198-206

Benoist, C. & Chambon, P., (1981)

In vivo sequence requirements of the SV40 early promotor region

Nature, 290, 5804, 304-310

Benton, T., Chen, T., McEntee, M., Fox, B., King, D., Crombie, R., Thomas, T.C., & Bebbington, C., (2002)

The use of UCOE vectors in combination with a preadapted serum free, suspension cell line allows for rapid production of large quantities of protein

Cytotechnology, 38, 1-2, 43-46

Berg, P.E., Popovic, Z., & Anderson, W.F., (1984)

Promoter dependence of enhancer activity

Molecular and Cellular Biology, 4, 8, 1664-1668

Biberman, Y. & Meyuhas, O., (1997)

Substitution of just five nucleotides at and around the transcription start site of rat beta-actin promoter is sufficient to render the resulting transcript a subject for translational control

FEBS Letters, 405, 3, 333-336

Birch, J.R. & Racher, A.J., (2006)

Antibody production

Advanced Drug Delivery Reviews, 58, 5-6, 671-685

Blount, B.A., Weenink, T., Vasylechko, S., & Ellis, T., (2012)

Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology

PLoS One, 7, 3, e33279

Bobrovnikova-Marjon, E. & Diehl, J.A., (2007)

Coping with stress: ATF6 alpha takes the stage

Developmental Cell, 13, 3, 322-324

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U., (2009)

Breaking the code of DNA binding specificity of TAL-type III effectors

Science, 326, 5959, 1509-1512

Borth, N., Mattanovich, D., Kunert, R., & Katinger, H., (2005)

Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line

Biotechnology Progress, 21, 1, 106-111

Boshart, M., Weber, F., Jahn, G., Dorschhasler, K., Fleckenstein, B., & Schaffner, W., (1985)

A very strong enhancer is located upstream of an immediate early gene of human cytomegalo-virus

Cell, 41, 2, 521-530

Bowers, L.M., Lapoint, K., Anthony, L., Pluciennik, A., & Filutowicz, M., (2004)

Bacterial expression system with tightly regulated gene expression and plasmid copy number

Gene, 340, 1, 11-18

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., & Choder, M., (2011)

Promoter elements regulate cytoplasmic mRNA decay

Cell, 147, 7, 1473-1483

Brezinsky, S.C.G., Chiang, G.G., Szilvasi, A., Mohan, S., Shapiro, R.I., MacLean, A., Sisk, W., & Thill, G., (2003)

A simple method for enriching populations of transfected CHO cells for cells of higher specific productivity

Journal of Immunological Methods, 277, 1-2, 141-155

Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F., Budach, W.E., Galosy, S., Muller, D., Noll, T., Wienberg, J., Jostock, T., Leonard, M., Grillari, J., Tauch, A., Goesmann, A., Helk, B., Mott, J.E., Puhler, A., & Borth, N., (2013)

Chinese hamster genome sequenced from sorted chromosomes

Nature Biotechnology, 31, 8, 694-695

Brown, A.J., Mainwaring, D.O., Sweeney, B., & James, D.C., (2013)

Block decoys: Transcription-factor decoys designed for in vitro gene regulation studies

Analytical Biochemistry, 443, 2, 205-210

Browne, S.M. & Al-Rubeai, M., (2007)

Selection methods for high-producing mammalian cell lines

Trends in Biotechnology, 25, 9, 425-432

Bulter, T., Lee, S.G., Woirl, W.W.C., Fung, E., Connor, M.R., & Liao, J.C., (2004)

Design of artificial cell-cell communication using gene and metabolic networks

Proceedings of the National Academy of Sciences of the United States of America, 101, 8, 2299-2304

Bunch, T.A., Grinblat, Y., & Goldstein, L.S.B., (1988)

Characterization and use of the drosophila metallothionein promoter in cultured Drosophila-melanogaster cells

Nucleic Acids Research, 16, 3, 1043-1061

Byrne, B.J., Davis, M.S., Yamaguchi, J., Bergsma, D.J., & Subramanian, K.N., (1983)

Definition of the Simian Virus-40 Early Promoter Region and Demonstration of a Host Range Bias in the Enhancement Effect of the Simian Virus-40 72-Base-Pair Repeat

Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 80, 3, 721-725

Cahill, M.A., Ernst, W.H., Janknecht, R., & Nordheim, A., (1994a)

Regulatory squelching

FEBS Letters, 344, 2-3, 105-108

Cahill, M.A., Ernst, W.H., Janknecht, R., & Nordheim, A., (1994b)

Regulatory squelching

FEBS Letters, 344, 2-3, 105-108

Cain, K., Peters, S., Hailu, H., Sweeney, B., Stephens, P., Heads, J., Sarkar, K., Ventom, A., Page, C., & Dickson, A., (2013)

A CHO cell line engineered to express XBP1 and ERO1-L alpha has increased levels of transient protein expression

Biotechnology Progress, 29, 3, 697-706

Callen, B.P., Shearwin, K.E., & Egan, J.B., (2004)

Transcriptional interference between convergent promoters caused by elongation over the promoter

Molecular Cell, 14, 5, 647-656

Campbell, M., Corisdeo, S., McGee, C., & Kraichely, D., (2010)

Utilization of site-specific recombination for generating therapeutic protein producing cell lines

Molecular Biotechnology, 45, 3, 199-202

Campos-da-Paz, M., Costa, C.S., Quilici, L.S., de Carmo Simoes, I., Kyaw, C.M., Maranhao, A.Q., & Brigido, M.M., (2008)

Production of recombinant human factor VIII in different cell lines and the effect of human XBP1 co-expression

Molecular Biotechnology, 39, 2, 155-158

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., & Werner, T., (2005)

MatInspector and beyond: promoter analysis based on transcription factor binding sites

Bioinformatics, 21, 13, 2933-2942

Castro, P.M.L., Hayter, P.M., Ison, A.P., & Bull, A.T., (1992)

Application of a statistical design to the optimization of culture-medium for recombinant Interferon-Gamma production by Chinese-hamster ovary cells

Applied Microbiology and Biotechnology, 38, 1, 84-90

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., & Voytas, D.F., (2011)

Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting

Nucleic Acids Research, 39, 12, e82

Chakrabarti, A., Chen, A.W., & Varner, J.D., (2011)

A review of the mammalian unfolded protein response

Biotechnology and Bioengineering, 108, 12, 2777-2793

Chapman, B.S., Thayer, R.M., Vincent, K.A., & Haigwood, N.L., (1991)

Effect of intron-a from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian-cells

Nucleic Acids Research, 19, 14, 3979-3986

Chee Furng Wong, D., Tin Kam Wong, K., Tang Goh, L., Kiat Heng, C., & Gek Sim Yap, M., (2005)

Impact of dynamic online fed-batch strategies on metabolism, productivity and N-glycosylation quality in CHO cell cultures

Biotechnology and Bioengineering, 89, 2, 164-177

Chiang, G.G. & Sisk, W.P., (2005)

Bcl-x(L) mediates increased production of humanized monoclonal antibodies in chinese hamster ovary cells

Biotechnology and Bioengineering, 91, 7, 779-792

Choi, O.R.B. & Engel, J.D., (1988)

Developmental regulation of Beta-globin gene switching

Cell, 55, 1, 17-26

Christensen, K.L., Patrick, A.N., McCoy, E.L., & Ford, H.L., (2008)

The six family of homeobox genes in development and cancer

Advances in Cancer Research, 101, 93-126

Chung, J.Y., Lim, S.W., Hong, Y.J., Hwang, S.O., & Lee, G.M., (2004)

Effect of doxycycline-regulated calnexin and calreticulin expression on specific thrombopoietin productivity of recombinant Chinese hamster ovary cells

Biotechnology and Bioengineering, 85, 5, 539-546

Conte, C., Dastugue, B., & Vaury, C., (2002)

Promoter competition as a mechanism of transcriptional interference mediated by retrotransposons

EMBO Journal, 21, 14, 3908-3916

Cooper, G.M., (2000)

The cell: A molecular approach - regulation of transcription in eukaryotes

2nd edition, Sinauer Associates, Sunderland (MA)

Available from: http://www.ncbi.nlm.nih.gov/books/NBK9904/

Corish, P. & Tyler-Smith, C., (1999)

Attenuation of green fluorescent protein half-life in mammalian cells

Protein Engineering, 12, 12, 1035-1040

Cost, G.J., Freyvert, Y., Vafiadis, A., Santiago, Y., Miller, J.C., Rebar, E., Collingwood, T.N., Snowden, A., & Gregory, P.D., (2010)

BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells

Biotechnology and Bioengineering, 105, 2, 330-340

Costello, L.C., Zou, J., Desouki, M.M., & Franklin, R.B., (2012)

Evidence for changes in RREB-1, ZIP3, and Zinc in the early development of pancreatic adenocarcinoma

Journal of Gastrointestinal Cancer, 43, 4, 570-578

Cowling, V.H., (2010)

Regulation of mRNA cap methylation

Biochemical Journal, 425, 295-302

Crijns, A.P., de Graeff, P., Geerts, D., Ten Hoor, K.A., Hollema, H., van der Sluis, T., Hofstra, R.M., de Bock, G.H., de Jong, S., van der Zee, A.G., & de Vries, E.G., (2007)

MEIS and PBX homeobox proteins in ovarian cancer

European Journal of Cancer, 43, 17, 2495-2505

Cullen, B.R., Lomedico, P.T., & Ju, G., (1984)

Transcriptional interference in avian retroviruses - implications for the promoter insertion model of leukemogenesis

Nature, 307, 5948, 241-245

Curtin, J.A., Dane, A.P., Swanson, A., Alexander, I.E., & Ginn, S.L., (2008)

Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct

Gene Therapy, 15, 5, 384-390

Davis, R., Schooley, K., Rasmussen, B., Thomas, J., & Reddy, P., (2000)

Effect of PDI overexpression on recombinant protein secretion in CHO cells

Biotechnology Progress, 16, 5, 736-743

De Jesus, M. & Wurm, F.M., (2011)

Manufacturing recombinant proteins in kg-ton quantities using animal cells in bioreactors

European Journal of Pharmaceutics and Biopharmaceutics, 78, 2, 184-188

de Poorter, J.J., Lipinski, K.S., Nelissen, R.G.H.H., Huizinga, T.W.J., & Hoeben, R.C., (2007)

Optimization of short-term transgene expression by sodium butyrate and ubiquitous chromatin opening elements (UCOEs)

Journal of Gene Medicine, 9, 8, 639-648

Dehli, T., Solem, C., & Jensen, P.R., (2012)

Tunable promoters in synthetic and systems biology

Sub-cellular Biochemistry, 64, 181-201

Deniaud, E., Baguet, J., Chalard, R., Blanquier, B., Brinza, L., Meunier, J., Michallet, M.C., Laugraud, A., Ah-Soon, C., Wierinckx, A., Castellazzi, M., Lachuer, J., Gautier, C., Marvel, J., & Leverrier, Y., (2009)

Overexpression of transcription factor Sp1 leads to gene expression perturbations and cell cycle inhibition

PLoS One, 4, 9, e7035

Desai, P. & Person, S., (1998)

Incorporation of the green fluorescent protein into the herpes simplex virus type 1 capsid

Journal of Virology, 72, 9, 7563-7568

Devries, E., Vandriel, W., Vandenheuvel, S.J.L., & Vandervliet, P.C., (1987)

Contactpoint analysis of the Hela Nuclear Factor-I recognition site reveals symmetrical binding at one side of the DNA helix

EMBO Journal, 6, 1, 161-168

Dharshanan, S., Chong, H., Hung, C.S., Zamrod, Z., & Kamal, N., (2011)

Rapid automated selection of mammalian cell line secreting high level of humanized monoclonal antibody using Clone Pix FL system and the correlation between exterior median intensity and antibody productivity

Electronic Journal of Biotechnology, 14, 2,

Dickerson, I.M., Peden, K.W.C., & Mains, R.E., (1989)

Metallothionein-I promoter-directed expression of foreign proteins in a mouse pituitary corticotrope tumor-cell line

Molecular and Cellular Endocrinology, 64, 2, 205-212

Dinnis, D.M. & James, D.C., (2005)

Engineering mammalian cell factories for improved recombinant monoclonal antibody production: Lessons from nature?

Biotechnology and Bioengineering, 91, 2, 180-189

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K., Leonard, M., Charlebois, T., & Clynes, M., (2010)

Microarray and proteomics expression profiling identifies several candidates, including the Valosin-Containing Protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines

Biotechnology and Bioengineering, 106, 1, 42-56

Dorner, A.J. & Kaufman, R.J., (1994)

The levels of endoplasmic reticulum proteins and ATP affect folding and secretion of selective proteins

Biologicals, 22, 2, 103-112

Dreesen, I.A.J. & Fussenegger, M., (2011)

Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of Chinese hamster ovary cells

Biotechnology and Bioengineering, 108, 4, 853-866

Du, Z.M., Treiber, D., Mccoy, R.E., Miller, A.K., Han, M., He, F., Domnitz, S., Heath, C., & Reddy, P., (2013)

Non-invasive UPR monitoring system and its applications in CHO production cultures

Biotechnology and Bioengineering, 110, 8, 2184-2194

Dynan, W.S. & Tjian, R., (1983)

The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter

Cell, 35, 1, 79-87

Eddy, B.E., Young, R.D., Borman, G.S., & Berkeley, W.H., (1961)

Tumors induced in hamsters by injection of rhesus monkey kidney cell extracts

Proceedings of the Society for Experimental Biology and Medicine, 107, 1, 191-&

Endo, K., Stapleton, J.A., Hayashi, K., Saito, H., & Inoue, T., (2013)

Quantitative and simultaneous translational control of distinct mammalian mRNAs

Nucleic Acids Research, 41, 13, e135

Espel, E., Fromental, C., Reichenbach, P., & Nabholz, M., (1990)

Activity and Interleukin-1 responsiveness of SV40 enhancer motifs in a rodent immature T-cell line

EMBO Journal, 9, 3, 929-937

Eszterhas, S.K., Bouhassira, E.E., Martin, D.I.K., & 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, 2, 469-479

Fan, L., Kadura, I., Krebs, L.E., Larson, J.L., Bowden, D.M., & Frye, C.C., (2013)

Development of a highly-efficient CHO cell line generation system with engineered SV40E promoter

Journal of Biotechnology, 168, 4, 652-658

Fan, L.C., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M., & Frye, C.C., (2012)

Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells

Biotechnology and Bioengineering, 109, 4, 1007-1015

Farzadfard, F., Perli, S.D., & Lu, T.K., (2013)

Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas

ACS Synthetic Biology, 2, 10, 604-613

Fernandes, E.R. & Rooney, R.J., (1997)

The adenovirus E1A-regulated transcription factor E4F is generated from the human homolog of nuclear factor phi AP3

Molecular and Cellular Biology, 17, 4, 1890-1903

Fickenscher, H., Stamminger, T., Ruger, R., & Fleckenstein, B., (1989)

The role of a repetitive palindromic sequence element in the human cytomegalovirus major immediate early enhancer

Journal of General Virology, 70 (Pt 1), 107-123

Figueroa, B., Jr., Sauerwald, T.M., Oyler, G.A., Hardwick, J.M., & Betenbaugh, M.J., (2003)

A comparison of the properties of a Bcl-xL variant to the wild-type anti-apoptosis inhibitor in mammalian cell cultures

Metabolic Engineering, 5, 4, 230-245

Fingar, D.C. & Blenis, J., (2004)

Target of rapamycin (TOR): an integrator of nutrient and growth factor signals and coordinator of cell growth and cell cycle progression

Oncogene, 23, 18, 3151-3171

Fingar, D.C., Richardson, C.J., Tee, A.R., Cheatham, L., Tsou, C., & Blenis, J., (2004)

mTOR controls cell cycle progression through its cell growth effectors S6K1 and 4E-BP1/eukaryotic translation initiation factor 4E

Molecular and Cellular Biology, 24, 1, 200-216

Fingar, D.C., Salama, S., Tsou, C., Harlow, E., & Blenis, J., (2002)

Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E

Genes & Development, 16, 12, 1472-1487

Florin, L., Pegel, A., Becker, E., Hausser, A., Olayioye, M.A., & Kaufmann, H., (2009)

Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells

Journal of Biotechnology, 141, 1-2, 84-90

Foecking, M.K. & Hofstetter, H., (1986)

Powerful and versatile enhancer-promoter unit for mammalian expression vectors

Gene, 45, 1, 101-105

Frech, K., DanescuMayer, J., & Werner, T., (1997)

A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter

Journal of Molecular Biology, 270, 5, 674-687

Freedman, R.B., Hirst, T.R., & Tuite, M.F., (1994)

Protein disulphide isomerase: building bridges in protein folding

Trends in Biochemical Sciences, 19, 8, 331-336

Fujioka, S., Niu, J., Schmidt, C., Sclabas, G.M., Peng, B., Uwagawa, T., Li, Z., Evans, D.B., Abbruzzese, J.L., & Chiao, P.J., (2004)

NF-kappaB and AP-1 connection: mechanism of NF-kappaB-dependent regulation of AP-1 activity

Molecular and Cellular Biology, 24, 17, 7806-7819

Fussenegger, M., Fassnacht, D., Schwartz, R., Zanghi, J.A., Graf, M., Bailey, J.E., & Portner, R., (2000)

Regulated overexpression of the survival factor bcl-2 in CHO cells increases viable cell density in batch culture and decreases DNA release in extended fixed-bed cultivation

Cytotechnology, 32, 1, 45-61

Galan-Caridad, J.M., Harel, S., Arenzana, T.L., Hou, Z.E., Doetsch, F.K., Mirny, L.A., & Reizis, B., (2007)

Zfx controls the self-renewal of embryonic and hematopoietic stem cells

Cell, 129, 2, 345-357

Gaubatz, S., Lindeman, G.J., Ishida, S., Jakoi, L., Nevins, J.R., Livingston, D.M., & Rempel, R.E., (2000)

E2F4 and E2F5 play an essential role in pocket protein-mediated G1 control

Molecular Cell, 6, 3, 729-735

Gawlitzek, M., Estacio, M., Furch, T., & Kiss, R., (2009)

Identification of cell culture conditions to control N-glycosylation site-occupancy of recombinant glycoproteins expressed in CHO cells

Biotechnology and Bioengineering, 103, 6, 1164-1175

Ghazal, P. & Nelson, J.A., (1991)

Enhancement of RNA polymerase II initiation complexes by a novel DNA control domain downstream from the cap site of the cytomegalovirus major immediate-early promoter

Journal of Virology, 65, 5, 2299-2307

Giorgetti, L., Siggers, T., Tiana, G., Caprara, G., Notarbartolo, S., Corona, T., Pasparakis, M., Milani, P., Bulyk, M.L., & Natoli, G., (2010)

Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs

Molecular Cell, 37, 3, 418-428

Girod, P.A., Nguyen, D.Q., Calabrese, D., Puttini, S., Grandjean, M., Martinet, D., Regamey, A., Saugy, D., Beckmann, J.S., Bucher, P., & Mermod, N., (2007)

Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells

Nature Methods, 4, 9, 747-753

Girod, P.A., Zahn-Zabal, M., & Mermod, N., (2005)

Use of the chicken lysozyme 5' matrix attachment region to generate high producer CHO cell lines

Biotechnology and Bioengineering, 91, 1, 1-11

Goldman, L.A., Cutrone, E.C., Kotenko, S.V., Krause, C.D., & Langer, J.A., (1996)

Modifications of vectors pEF-BOS, pcDNA1 and pcDNA3 result in improved convenience and expression

Biotechniques, 21, 6, 1013-1015

Gomez-Martin, D., Diaz-Zamudio, M., Galindo-Campos, M., & Alcocer-Varela, J., (2010)

Early growth response transcription factors and the modulation of immune response: implications towards autoimmunity

Autoimmunity Reviews, 9, 6, 454-458

Gopalkrishnan, R.V., Christiansen, K.A., Goldstein, N.I., DePinho, R.A., & Fisher, P.B., (1999)

Use of the human EF-1 alpha promoter for expression can significantly increase success in establishing stable cell lines with consistent expression: a study using the tetracycline-inducible system in human cancer cells

Nucleic Acids Research, 27, 24, 4775-4782

Gossen, M. & Bujard, H., (1992)

Tight control of gene-expression in mammalian-cells by tetracycline-responsive promoters

Proceedings of the National Academy of Sciences of the United States of America, 89, 12, 5547-5551

Gotea, V., Visel, A., Westlund, J.M., Nobrega, M.A., Pennacchio, L.A., & Ovcharenko, I., (2010)

Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers

Genome Research, 20, 5, 565-577

Grainger, R.K. & James, D.C., (2013)

CHO cell line specific prediction and control of recombinant monoclonal antibody N-glycosylation

Biotechnology and Bioengineering, 110, 11, 2970-2983

Grillberger, L., Kreil, T.R., Nasr, S., & Reiter, M., (2009)

Emerging trends in plasma-free manufacturing of recombinant protein therapeutics expressed in mammalian cells

Biotechnology Journal, 4, 2, 186-201

Grushkin, D., (2012)

The new DRUG CIRCUIT

Nature Medicine, 18, 10, 1452-1454

Guo, Z.S., Wang, L.H., Eisensmith, R.C., & Woo, S.L.C., (1996)

Evaluation of promoter strength for hepatic gene expression in vivo following adenovirus-mediated gene transfer

Gene Therapy, 3, 9, 802-810

Haas, I.G. & Wabl, M., (1983)

Immunoglobulin heavy chain binding protein

Nature, 306, 5941, 387-389

Hacker, D.L., De Jesus, M., & Wurm, F.M., (2009)

25 years of recombinant proteins from reactor-grown cells - where do we go from here?

Biotechnology Advances, 27, 6, 1023-1027

Hai, T. & Hartman, M.G., (2001)

The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis

Gene, 273, 1, 1-11

Hamilton, T.L., Stoneley, M., & Bushell, M., (2006)

TOPs and their regulation

Biochemical Society Transactions, 34, 12-16

Hammond, S., Kaplarevic, M., Borth, N., Betenbaugh, M.J., & Lee, K.H., (2012)

Chinese hamster genome database: An online resource for the CHO community at www.CHOgenome.org

Biotechnology and Bioengineering, 109, 6, 1353-1356

Han, Y.K., Ha, T.K., Lee, S.J., Lee, J.S., & Lee, G.M., (2011)

Autophagy and apoptosis of recombinant Chinese hamster ovary cells during fed-batch culture: Effect of nutrient supplementation

Biotechnology and Bioengineering, 108, 9, 2182-2192

Hanania, E.G., Fieck, A., Stevens, J., Bodzin, L.J., Palsson, B.O., & Koller, M.R., (2005)

Automated in situ measurement of cell-specific antibody secretion and laser-mediated purification for rapid cloning of highly-secreting producers

Biotechnology and Bioengineering, 91, 7, 872-876

Haredy, A.M., Nishizawa, A., Honda, K., Ohya, T., Ohtake, H., & Omasa, T., (2011)

ATF4 over-expression increased IgG1 productivity in Chinese hamster ovary cells

BMC Proceedings, 5 Suppl 8, O11

Harshman, K.D., Moyerowley, W.S., & Parker, C.S., (1988)

Transcriptional activation by the SV40 Ap-1 recognition element in yeast is mediated by a factor similar to Ap-1 that is distinct from Gcn4

Cell, 53, 2, 321-330

Hasegawa, H., Wendling, J., He, F., Trilisky, E., Stevenson, R., Franey, H., Kinderman, F., Li, G., Piedmonte, D.M., Osslund, T., Shen, M., & Ketchem, R.R., (2011)

In vivo crystallization of human IgG in the endoplasmic reticulum of engineered Chinese hamster ovary (CHO) cells

Journal of Biological Chemistry, 286, 22, 19917-19931

Hashiya, N., Jo, N., Aoki, M., Matsumoto, K., Nakamura, T., Sato, Y., Ogata, N., Ogihara, T., Kaneda, Y., & Morishita, R., (2004)

In vivo evidence of angiogenesis induced by transcription factor ets-1 - Ets-1 is located upstream of angiogenesis cascade

Circulation, 109, 24, 3035-3041

Haze, K., Yoshida, H., Yanagi, H., Yura, T., & Mori, K., (1999)

Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress

Molecular Biology of the Cell, 10, 11, 3787-3799

He, B. & Weber, G.F., (2004)

Synergistic activation of the CMV promoter by NF-kappaB P50 and PKG

Biochemical and Biophysical Research Communications, 321, 1, 13-20

Hennighausen, L. & Fleckenstein, B., (1986)

Nuclear factor 1 interacts with five DNA elements in the promoter region of the human cytomegalovirus major immediate early gene

EMBO Journal, 5, 6, 1367-1371

Hess, J., Angel, P., & Schorpp-Kistner, M., (2004)

AP-1 subunits: quarrel and harmony among siblings

Journal of Cell Science, 117, 25, 5965-5973

Hetz, C., (2012)

The unfolded protein response: controlling cell fate decisions under ER stress and beyond

Nature Reviews Molecular Cell Biology, 13, 2, 89-102

Hirschman, J.E., Durbin, K.J., & Winston, F., (1988)

Genetic-evidence for promoter competition in Saccharomyces-cerevisiae

Molecular and Cellular Biology, 8, 11, 4608-4615

Ho Sui, S.J., Mortimer, J.R., Arenillas, D.J., Brumm, J., Walsh, C.J., Kennedy, B.P., & Wasserman, W.W., (2005)

oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes

Nucleic Acids Research, 33, 10, 3154-3164

Hoesel, B. & Schmid, J.A., (2013)

The complexity of NF-kappaB signaling in inflammation and cancer

Molecular Cancer, 12, 86

Huang, Y., Li, Y., Wang, Y.G., Gu, X., Wang, Y., & Shen, B.F., (2007)

An efficient and targeted gene integration system for high-level antibody expression

Journal of Immunological Methods, 322, 1-2, 28-39

Huliak, I., Sike, A., Zencir, S., & Boros, I.M., (2012)

The objectivity of reporters: Interference between physically unlinked promoters affects reporter gene expression in transient transfection experiments

DNA and Cell Biology, 31, 11, 1580-1584

Hurley, J.M., Chen, C.H., Loros, J.J., & Dunlap, J.C., (2012)

Light-inducible system for tunable protein expression in Neurospora crassa

G3: Genes, Genomes, Genetics, 2, 10, 1207-1212

Hwang, S.O., Chung, J.Y., & Lee, G.M., (2003)

Effect of doxycycline-regulated ERp57 expression on specific thrombopoietin productivity of recombinant CHO cells

Biotechnology Progress, 19, 1, 179-184

Hwang, S.O. & Lee, G.M., (2008)

Nutrient deprivation induces autophagy as well as apoptosis in Chinese hamster ovary cell culture

Biotechnology and Bioengineering, 99, 3, 678-685

Hwang, S.O. & Lee, G.M., (2009)

Effect of Akt overexpression on programmed cell death in antibody-producing Chinese hamster ovary cells

Journal of Biotechnology, 139, 1, 89-94

Infantino, V., Convertini, P., Iacobazzi, F., Pisano, I., Scarcia, P., & Iacobazzi, V., (2011)

Identification of a novel Sp1 splice variant as a strong transcriptional activator

Biochemical and Biophysical Research Communications, 412, 1, 86-91

Ingelbrecht, I., Breyne, P., Vancompernolle, K., Jacobs, A., Vanmontagu, M., & Depicker, A., (1991)

Transcriptional interference in transgenic plants

Gene, 109, 2, 239-242

Isern, E., Gustems, M., Messerle, M., Borst, E., Ghazal, P., & Angulo, A., (2011)

The activator protein 1 binding motifs within the human cytomegalovirus major immediate-early enhancer are functionally redundant and act in a cooperative manner with the NF-{kappa}B sites during acute infection

Journal of Virology, 85, 4, 1732-1746

Jaluria, P., Betenbaugh, M., Konstantopoulos, K., & Shiloach, J., (2007)

Enhancement of cell proliferation in various mammalian cell lines by gene insertion of a cyclin-dependent kinase homolog

BMC Biotechnology, 7, 71

Jardon, M.A., Sattha, B., Braasch, K., Leung, A.O., Cote, H.C., Butler, M., Gorski, S.M., & Piret, J.M., (2012)

Inhibition of glutamine-dependent autophagy increases t-PA production in CHO cell fed-batch processes

Biotechnology and Bioengineering, 109, 5, 1228-1238

Jensen, P.R. & Hammer, K., (1998)

The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters

Applied and Environmental Microbiology, 64, 1, 82-87

Jiang, R., Gao, Z.L., Sun, M., Zhang, X.Y., Wang, J.C., & Wu, H., (2013)

Zinc finger X-chromosomal protein promotes growth and tumorigenesis in human osteosarcoma cells

Pakistan Journal of Medical Science, 29, 4, 997-1002

Jobbagy, Z., Ward, J.L., Toan, S.V., Leung, G.P., & Tse, C.M., (2002)

One-step unidirectional cloning of tandem repeats of DNA fragments: an application for fusion protein production

Analytical Biochemistry, 303, 1, 104-107

Jordan, M.C., (1983)

Latent infection and the elusive cytomegalovirus

Reviews of Infectious Diseases, 5, 2, 205-215

Jung, C.H., Ro, S.H., Cao, J., Otto, N.M., & Kim, D.H., (2010)

mTOR regulation of autophagy

FEBS Letters, 584, 7, 1287-1295

Kaczynski, J., Cook, T., & Urrutia, R., (2003)

Sp1- and Kruppel-like transcription factors

Genome Biology, 4, 2, 206

Kadesch, T. & Berg, P., (1986)

Effects of the position of the Simian Virus-40 enhancer on expression of multiple transcription units in a single plasmid

Molecular and Cellular Biology, 6, 7, 2593-2601

Kameyama, Y., Kawabe, Y., Ito, A., & Kamihira, M., (2010)

An accumulative site-specific gene integration system using Cre recombinase-mediated cassette exchange

Biotechnology and Bioengineering, 105, 6, 1106-1114

Kaufman, R.J. & Sharp, P.A., (1982)

Amplification and expression of sequences cotransfected with a modular dihydrofolate-reductase complementary-DNA gene

Journal of Molecular Biology, 159, 4, 601-621

Kaufman, R.J., Wasley, L.C., Spiliotes, A.J., Gossels, S.D., Latt, S.A., Larsen, G.R., & Kay, R.M., (1985)

Coamplification and coexpression of human tissue-type plasminogen activator and murine dihydrofolate reductase sequences in Chinese hamster ovary cells

Molecular and Cellular Biology, 5, 7, 1750-1759

Kawakami, K., Scheidereit, C., & Roeder, R.G., (1988)

Identification and purification of a human immunoglobulin-enhancer-binding protein (NF-kappa B) that activates transcription from a human immunodeficiency virus type 1 promoter in vitro

Proceedings of the National Academy of Sciences of the United States of America, 85, 13, 4700-4704

Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., & Wingender, E., (2003)

MATCH (TM): a tool for searching transcription factor binding sites in DNA sequences

Nucleic Acids Research, 31, 13, 3576-3579

Kelley, B., (2009)

Industrialization of mAb production technology The bioprocessing industry at a crossroads

Mabs, 1, 5, 443-452

Khalili, K., Khoury, G., & Brady, J., (1986)

Spacing between Simian-Virus 40 early transcriptional control sequences is important for regulation of early RNA-synthesis and gene-expression

Journal of Virology, 60, 3, 935-942

Khattak, S.F., Xing, Z., Kenty, B., Koyrakh, I., & Li, Z.J., (2010)

Feed development for fed-batch CHO production process by semisteady state analysis

Biotechnology Progress, 26, 3, 797-804

Kim, D.W., Uetsuki, T., Kaziro, Y., Yamaguchi, N., & Sugano, S., (1990)

Use of the human Elongation Factor-1-Alpha promoter as a versatile and efficient expression system

Gene, 91, 2, 217-223

Kim, J., Kundu, M., Viollet, B., & Guan, K.L., (2011)

AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1

Nature Cell Biology, 13, 2, 132-141

Kim, J.D., Yoon, Y., Hwang, H.Y., Park, J.S., Yu, S., Lee, J., Baek, K., & Yoon, J., (2005)

Efficient selection of stable Chinese hamster ovary (CHO) cell lines for expression of recombinant proteins by using human interferon beta SAR element

Biotechnology Progress, 21, 3, 933-937

Kim, J.M., Kim, J.S., Park, D.H., Kang, H.S., Yoon, J., Baek, K., & Yoon, Y., (2004)

Improved recombinant gene expression in CHO cells using matrix attachment regions

Journal of Biotechnology, 107, 2, 95-105

Kim, J.Y., Kim, Y.G., & Lee, G.M., (2012)

CHO cells in biotechnology for production of recombinant proteins: current state and further potential

Applied Microbiology and Biotechnology, 93, 3, 917-930

Kim, N.S. & Lee, G.M., (2001)

Overexpression of bcl-2 inhibits sodium butyrate-induced apoptosis in Chinese hamster ovary cells resulting in enhanced humanized antibody production

Biotechnology and Bioengineering, 71, 3, 184-193

Kim, N.S. & Lee, G.M., (2002)

Inhibition of sodium butyrate-induced apoptosis in recombinant Chinese hamster ovary cells by constitutively expressing antisense RNA of caspase-3

Biotechnology and Bioengineering, 78, 2, 217-228

Kim, S.H. & Lee, G.M., (2009)

Development of serum-free medium supplemented with hydrolysates for the production of therapeutic antibodies in CHO cell cultures using design of experiments

Applied Microbiology and Biotechnology, 83, 4, 639-648

Kim, Y., Ratziu, V., Choi, S.G., Lalazar, A., Theiss, G., Dang, Q., Kim, S.J., & Friedman, S.L., (1998)

Transcriptional activation of transforming growth factor beta1 and its receptors by the Kruppel-like factor Zf9/core promoter-binding protein and Sp1. Potential mechanisms for autocrine fibrogenesis in response to injury

Journal of Biological Chemistry, 273, 50, 33750-33758

Kim, Y.G., Kim, J.Y., Mohan, C., & Lee, G.M., (2009)

Effect of Bcl-xL overexpression on apoptosis and autophagy in recombinant Chinese hamster ovary cells under nutrient-deprived condition

Biotechnology and Bioengineering, 103, 4, 757-766

Kito, M., Itami, S., Fukano, Y., Yamana, K., & Shibui, T., (2002)

Construction of engineered CHO strains for high-level production of recombinant proteins

Applied Microbiology and Biotechnology, 60, 4, 442-448

Klingenhoff, A., Frech, K., Quandt, K., & Werner, T., (1999)

Functional promoter modules can be defected by formal models independent of overall nucleoside sequence similarity

Bioinformatics, 15, 3, 180-186

Kohno, K., (2010)

Stress-sensing mechanisms in the unfolded protein response: similarities and differences between yeast and mammals

Journal of Biochemistry, 147, 1, 27-33

Koller, M.R., Hanania, E.G., Stevens, J., Eisfeld, T.M., Sasaki, G.C., Fieck, A., & Palsson, B.O., (2004)

High-throughput laser-mediated in situ cell purification with high purity and yield

Cytometry A, 61, 2, 153-161

Kriz, A., Schmid, K., Baumgartner, N., Ziegler, U., Berger, I., Ballmer-Hofer, K., & Berger, P., (2010)

A plasmid-based multigene expression system for mammalian cells

Nature Communications, 1, 120

Ku, S.C., Ng, D.T., Yap, M.G., & Chao, S.H., (2008)

Effects of overexpression of X-box binding protein 1 on recombinant protein production in Chinese hamster ovary and NS0 myeloma cells

Biotechnology and Bioengineering, 99, 1, 155-164

Ku, S.C.Y., Toh, P.C., Lee, Y.Y., Chusainow, J., Yap, M.G.S., & Chao, S.H., (2010)

Regulation of XBP-1 signaling during transient and stable recombinant protein production in CHO cells

Biotechnology Progress, 26, 2, 517-526

Kuystermans, D. & Al-Rubeai, M., (2009)

cMyc increases cell number through uncoupling of cell division from cell size in CHO cells

BMC Biotechnology, 9, 76

Lai, T., Yang, Y., & Ng, S.K., (2013)

Advances in mammalian cell line development technologies for recombinant protein production

Pharmaceuticals, 6, 5, 579-603

Laplante, M. & Sabatini, D.M., (2012)

mTOR signaling in growth control and disease

Cell, 149, 2, 274-293

Le, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., Karypis, G., & Hu, W.S., (2012)

Multivariate analysis of cell culture bioprocess data--lactate consumption as process indicator

Journal of Biotechnology, 162, 2-3, 210-223

Lee, A.H., Chu, G.C., Iwakoshi, N.N., & Glimcher, L.H., (2005)

XBP-1 is required for biogenesis of cellular secretory machinery of exocrine glands

EMBO Journal, 24, 24, 4368-4380

Lee, A.H., Iwakoshi, N.N., & Glimcher, L.H., (2003)

XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response

Molecular and Cellular Biology, 23, 21, 7448-7459

Lee, J., Lau, J., Chong, G., & Chao, S.H., (2007)

Cell-specific effects of human cytomegalovirus unique region on recombinant protein expression

Biotechnology Letters, 29, 12, 1797-1802

Lee, J.S., Ha, T.K., Park, J.H., & Lee, G.M., (2013)

Anti-cell death engineering of CHO cells: Co-overexpression of Bcl-2 for apoptosis inhibition, Beclin-1 for autophagy induction

Biotechnology and Bioengineering, 110, 8, 2195-2207

Lee, K.A. & Green, M.R., (1987)

A cellular transcription factor E4F1 interacts with an E1a-inducible enhancer and mediates constitutive enhancer function in vitro

EMBO Journal, 6, 5, 1345-1353

Lee, S.K. & Lee, G.M., (2003)

Development of apoptosis-resistant dihydrofolate reductase-deficient Chinese hamster ovary cell line

Biotechnology and Bioengineering, 82, 7, 872-876

Lee, Y., Sohn, W.J., Kim, D.S., & Kwon, H.J., (2004)

NF-kappaB- and c-Jun-dependent regulation of human cytomegalovirus immediate-early gene enhancer/promoter in response to lipopolysaccharide and bacterial CpG-oligodeoxynucleotides in macrophage cell line RAW 264.7

European Journal of Biochemistry, 271, 6, 1094-1105

Lembo, D., Angeretti, A., Foresta, P., Gribaudo, G., Gariglio, M., & Landolfo, S., (1994)

Trans-activation of the mouse cytomegalovirus immediate early gene enhancer by ras oncogenes

Journal of General Virology, 75 (Pt 7), 1685-1692

Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J., Bian, C., Xie, M., Chen, W., Li, N., Baycin-Hizal, D., Latif, H., Forster, J., Betenbaugh, M.J., Famili, I., Xu, X., Wang, J., & Palsson, B.O., (2013)

Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome

Nature Biotechnology, 31, 8, 759-765

Li, F., Vijayasankaran, N., Shen, A., Kiss, R., & Amanullah, A., (2010)

Cell culture processes for monoclonal antibody production

Mabs, 2, 5, 466-479

Li, M.Q., Baumeister, P., Roy, B., Phan, T., Foti, D., Luo, S.Z., & Lee, A.S., (2000)

ATF6 as a transcription activator of the endoplasmic reticulum stress element: Thapsigargin stress-induced changes and synergistic interactions with NF-Y and YY1

Molecular and Cellular Biology, 20, 14, 5096-5106

Lifanov, A.P., Makeev, V.J., Nazina, A.G., & Papatsenko, D.A., (2003)

Homotypic regulatory clusters in Drosophila

Genome Research, 13, 4, 579-588

Liu, H.S., Jan, M.S., Chou, C.K., Chen, P.H., & Ke, N.J., (1999)

Is green fluorescent protein toxic to the living cells?

Biochemical and Biophysical Research Communications, 260, 3, 712-717

Liu, Z., Cashion, L.M., & Twu, J.J., (1997)

A systematic comparison of relative promoter/enhancer activities in mammalian cell lines

Analytical Biochemistry, 246, 1, 150-152

Luu, P. & Flores, O., (1997)

Binding of SP1 to the immediate-early protein-responsive element of the human cytomegalovirus DNA polymerase promoter

Journal of Virology, 71, 9, 6683-6691

Ma, X.M. & Blenis, J., (2009)

Molecular mechanisms of mTOR-mediated translational control

Nature Reviews Molecular Cell Biology, 10, 5, 307-318

Majors, B.S., Arden, N., Oyler, G.A., Chiang, G.G., Pederson, N.E., & Betenbaugh, M.J., (2008a)

E2F-1 overexpression increases viable cell density in batch cultures of Chinese hamster ovary cells

Journal of Biotechnology, 138, 3-4, 103-106

Majors, B.S., Betenbaugh, M.J., Pederson, N.F., & Chiang, G.G., (2008b)

Enhancement of transient gene expression and culture viability using Chinese hamster ovary cells overexpressing Bcl-x(L)

Biotechnology and Bioengineering, 101, 3, 567-578

Manz, R., Assenmacher, M., Pfluger, E., Miltenyi, S., & Radbruch, A., (1995)

Analysis and sorting of live cells according to secreted molecules, relocated to a cell-surface affinity matrix

Proceedings of the National Academy of Sciences of the United States of America, 92, 6, 1921-1925

Mariati, Ng, Y.K., Chao, S.H., Yap, M.G.S., & Yang, Y.S., (2010)

Evaluating regulatory elements of human cytomegalovirus major immediate early gene for enhancing transgene expression levels in CHO K1 and HEK293 cells

Journal of Biotechnology, 147, 3-4, 160-163

Martin, K.M., Cooper, W.N., Metcalfe, J.C., & Kemp, P.R., (2000)

Mouse BTEB3, a new member of the basic transcription element binding protein (BTEB) family, activates expression from GC-rich minimal promoter regions

Biochemical Journal, 345 Pt 3, 529-533

Mason, M., Sweeney, B., Cain, K., Stephens, P., & Sharfstein, S.T., (2012)

Identifying bottlenecks in transient and stable production of recombinant monoclonal-antibody sequence variants in chinese hamster ovary cells

Biotechnology Progress, 28, 3, 846-855

Matsumoto, N., Laub, F., Aldabe, R., Zhang, W., Ramirez, F., Yoshida, T., & Terada, M., (1998)

Cloning the cDNA for a new human zinc finger protein defines a group of closely related Kruppel-like transcription factors

Journal of Biological Chemistry, 273, 43, 28229-28237

McConnell, B.B. & Yang, V.W., (2010)

Mammalian Kruppel-like factors in health and diseases

Physiological Reviews, 90, 4, 1337-1381

Meadows, S.M., Myers, C.T., & Krieg, P.A., (2011)

Regulation of endothelial cell development by ETS transcription factors

Seminars in Cell & Developmental Biology, 22, 9, 976-984

Mearini, G., Nielsen, P.E., & Fackelmayer, F.O., (2004)

Localization and dynamics of small circular DNA in live mammalian nuclei

Nucleic Acids Research, 32, 8, 2642-2651

Meng, Y.G., Liang, J., Wong, W.L., & Chisholm, V., (2000)

Green fluorescent protein as a second selectable marker for selection of high producing clones from transfected CHO cells

Gene, 242, 1-2, 201-207

Meunier, L., Usherwood, Y.K., Chung, K.T., & Hendershot, L.M., (2002)

A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins

Molecular Biology of the Cell, 13, 12, 4456-4469

Meyuhas, O., (2000)

Synthesis of the translational apparatus is regulated at the translational level

European Journal of Biochemistry, 267, 21, 6321-6330

Mijakovic, I., Petranovic, D., & Jensen, P.R., (2005)

Tunable promoters in systems biology

Current Opinion in Biotechnology, 16, 3, 329-335

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., Dulay, G.P., Hua, K.L., Ankoudinova, I., Cost, G.J., Urnov, F.D., Zhang, H.S., Holmes, M.C., Zhang, L., Gregory, P.D., & Rebar, E.J., (2011)

A TALE nuclease architecture for efficient genome editing

Nature Biotechnology, 29, 2, 143-148

Mirkovitch, J., Mirault, M.E., & Laemmli, U.K., (1984)

Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold

Cell, 39, 1, 223-232

Mizushima, S. & Nagata, S., (1990)

pEF-BOS, a powerful mammalian expression vector

Nucleic Acids Research, 18, 17, 5322

Mohan, C., Park, S.H., Chung, J.Y., & Lee, G.M., (2007)

Effect of doxycycline-regulated protein disulfide isomerase expression on the specific productivity of recombinant CHO cells: Thrombopoietin and antibody

Biotechnology and Bioengineering, 98, 3, 611-615

Mori, K., (2009)

Signalling pathways in the unfolded protein response: Development from yeast to mammals

Journal of Biochemistry, 146, 6, 743-750

Morris, A.E. & Schmid, J., (2000)

Effects of insulin and LongR(3) on serum-free Chinese hamster ovary cell cultures expressing two recombinant proteins

Biotechnology Progress, 16, 5, 693-697

Moscou, M.J. & Bogdanove, A.J., (2009)

A simple cipher governs DNA recognition by TAL effectors

Science, 326, 5959, 1501

Muller, K., Engesser, R., Schulz, S., Steinberg, T., Tomakidi, P., Weber, C.C., Ulm, R., Timmer, J., Zurbriggen, M.D., & Weber, W., (2013)

Multi-chromatic control of mammalian gene expression and signaling

Nucleic Acids Research, 41, 12, e124

Nakai, H., Herzog, R.W., Hagstrom, J.N., Walter, J., Kung, S.H., Yang, E.Y., Tai, S.J., Iwaki, Y., Kurtzman, G.J., Fisher, K.J., Colosi, P., Couto, L.B., & High, K.A., (1998)

Adeno-associated viral vector-mediated gene transfer of human blood coagulation factor IX into mouse liver

Blood, 91, 12, 4600-4607

Nasser, W., Rochman, M., & Muskhelishvili, G., (2002)

Transcriptional regulation of fis operon involves a module of multiple coupled promoters

EMBO Journal, 21, 4, 715-724

Nelson, J.A., Reynolds-Kohler, C., & Smith, B.A., (1987)

Negative and positive regulation by a short segment in the 5'-flanking region of the human cytomegalovirus major immediate-early gene

Molecular and Cellular Biology, 7, 11, 4125-4129

Nevins, J.R., (1998)

Toward an understanding of the functional complexity of the E2F and retinoblastoma families

Cell Growth & Differentiation, 9, 8, 585-593

Ng, S.K., Wang, D.I.C., & Yap, M.G.S., (2007)

Application of destabilizing sequences on selection marker for improved recombinant protein productivity in CHO-DG44

Metabolic Engineering, 9, 3, 304-316

Nishimiya, D., Mano, T., Miyadai, K., Yoshida, H., & Takahashi, T., (2013)

Overexpression of CHOP alone and in combination with chaperones is effective in improving antibody production in mammalian cells

Applied Microbiology and Biotechnology, 97, 6, 2531-2539

O'Callaghan, P.M. & James, D.C., (2008)

Systems biotechnology of mammalian cell factories

Briefings in Functional Genomics & Proteomics, 7, 2, 95-110

Ohya, T., Hayashi, T., Kiyama, E., Nishii, H., Miki, H., Kobayashi, K., Honda, K., Omasa, T., & Ohtake, H., (2008)

Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression

Biotechnology and Bioengineering, 100, 2, 317-324

Olive, D.M., Almulla, W., Simsek, M., Zarban, S., & Alnakib, W., (1990)

The human cytomegalovirus immediate early enhancer-promoter is responsive to activation by the adenovirus-5 13s E1a gene

Archives of Virology, 112, 1-2, 67-80

Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., & Remington, S.J., (1996)

Crystal structure of the Aequorea victoria green fluorescent protein

Science, 273, 5280, 1392-1395

Oxford, G., Smith, S.C., Hampton, G., & Theodorescu, D., (2007)

Expression profiling of Ral-depleted bladder cancer cells identifies RREB-1 as a novel transcriptional Ral effector

Oncogene, 26, 50, 7143-7152

Pan, M.G., Xiong, Y., & Chen, F., (2013)

NFAT gene family in inflammation and cancer

Current Molecular Medicine, 13, 4, 543-554

Parekh, R., Forrester, K., & Wittrup, D., (1995)

Multicopy overexpression of bovine pancreatic trypsin-inhibitor saturates the protein-folding and secretory capacity of Saccharomyces-cerevisiae

Protein Expression and Purification, 6, 4, 537-545

Peng, R.W., Abellan, E., & Fussenegger, M., (2011)

Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells

Biotechnology and Bioengineering, 108, 3, 611-620

Peng, R.W. & Fussenegger, M., (2009)

Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells

Biotechnology and Bioengineering, 102, 4, 1170-1181

Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., & Gersbach, C.A., (2013)

Synergistic and tunable human gene activation by combinations of synthetic transcription factors

Nature Methods, 10, 3, 239-242

Persengiev, S.P. & Green, M.R., (2003)

The role of ATF/CREB family members in cell growth, survival and apoptosis

Apoptosis, 8, 3, 225-228

Pfaffl, M.W., Georgieva, T.M., Georgiev, I.P., Ontsouka, E., Hageleit, M., & Blum, J.W., (2002)

Real-time RT-PCR quantification of insulin-like growth factor (IGF)-1, IGF-1 receptor, IGF-2, IGF-2 receptor, insulin receptor, growth hormone receptor, IGF-binding proteins 1, 2 and 3 in the bovine species

Domestic Animal Endocrinology, 22, 2, 91-102

Pilz, A.J., Povey, S., Gruss, P., & Abbott, C.M., (1993)

Mapping of the human homologs of the murine paired-box-containing genes

Mammalian Genome, 4, 2, 78-82

Pjanic, M., Pjanic, P., Schmid, C., Ambrosini, G., Gaussin, A., Plasari, G., Mazza, C., Bucher, P., & Mermod, N., (2011)

Nuclear factor I revealed as family of promoter binding transcription activators

BMC Genomics, 12,

Poloumienko, A., (2004)

Cloning and comparative analysis of the bovine, porcine, and equine sex chromosome genes ZFX and ZFY

Genome, 47, 1, 74-83

Prescott, E.M. & Proudfoot, N.J., (2002)

Transcriptional collision between convergent genes in budding yeast

Proceedings of the National Academy of Sciences of the United States of America, 99, 13, 8796-8801

Prestridge, D.S., (1991)

Signal Scan - a computer-program that scans DNA-sequences for eukaryotic transcriptional elements

Computer Applications in the Biosciences, 7, 2, 203-206

Prestridge, D.S., (1996)

SIGNAL SCAN 4.0: additional databases and sequence formats

Computer Applications in the Biosciences, 12, 2, 157-160

Prosch, S., Staak, K., Stein, J., Liebenthal, C., Stamminger, T., Volk, H.D., & Kruger, D.H., (1995)

Stimulation of the human cytomegalovirus IE enhancer promoter in Hl-60 cells by Tnf-Alpha is mediated via induction of Nf-Kappa-B

Virology, 208, 1, 197-206

Prywes, R. & Zhu, H., (1992)

In vitro squelching of activated transcription by serum response factor: evidence for a common coactivator used by multiple transcriptional activators

Nucleic Acids Research, 20, 3, 513-520

Purcell, O., Savery, N.J., Grierson, C.S., & di Bernardo, M., (2010)

A comparative analysis of synthetic genetic oscillators

Journal of the Royal Society Interface, 7, 52, 1503-1524

Pybus, L.P., Dean, G., West, N.R., Smith, A., Daramola, O., Field, R., Wilkinson, S.J., & James, D.C., (2013)

Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells

Biotechnology and Bioengineering,

Qin, J.Y., Zhang, L., Clift, K.L., Hulur, I., Xiang, A.P., Ren, B.Z., & Lahn, B.T., (2010)

Systematic comparison of constitutive promoters and the doxycycline-inducible promoter

PLoS One, 5, 5, e10611

Qin, X.L., Qian, J.C., Yao, G.F., Zhuang, Y.P., Zhang, S.L., & Chu, J., (2011)

GAP promoter library for fine-tuning of gene expression in Pichia pastoris

Applied and Environmental Microbiology, 77, 11, 3600-3608

Quandt, K., Frech, K., Karas, H., Wingender, E., & Werner, T., (1995)

MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data

Nucleic Acids Research, 23, 23, 4878-4884

Rahimpour, A., Vaziri, B., Moazzami, R., Nematollahi, L., Barkhordari, F., Kokabee, L., Adeli, A., & Mahboudi, F., (2013)

Engineering the cellular protein secretory pathway for enhancement of recombinant tissue plasminogen activator expression in Chinese hamster ovary cells: Effects of CERT and XBP1s genes

Journal of Microbiology and Biotechnology, 23, 8, 1116-1122

Ramezani, A., Hawley, T.S., & Hawley, R.G., (2000)

Lentiviral vectors for enhanced gene expression in human hematopoietic cells

Molecular Therapy, 2, 5, 458-469

Randi, A.M., Sperone, A., Dryden, N.H., & Birdsey, G.M., (2009)

Regulation of angiogenesis by ETS transcription factors

Biochemical Society Transactions, 37, 1248-1253

Raychaudhuri, P., Rooney, R., & Nevins, J.R., (1987)

Identification of an E1a-inducible cellular factor that interacts with regulatory sequences within the Adenovirus E4 promoter

EMBO Journal, 6, 13, 4073-4081

Richardson, G., Lin, N., Lacy, K., Davis, L., Gray, M., Cresswell, J., Gerber, M., Caple, M., & Kayser, K., (2010)

Cell Xpress™ technology facilitates high-producing Chinese hamster ovary cell line generation using glutamine synthetase gene expression system

Cells and Culture, 4, 45-48

Richter, J.D. & Sonenberg, N., (2005)

Regulation of cap-dependent translation by eIF4E inhibitory proteins

Nature, 433, 7025, 477-480

Russell, J.P., Chang, D.W., Tretiakova, A., & Padidam, M., (2006)

Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells

Biotechniques, 40, 4, 460, 462, 464

Saffer, J.D., Jackson, S.P., & Thurston, S.J., (1990)

SV40 stimulates expression of the transacting factor Sp1 at the mRNA level

Genes & Development, 4, 4, 659-666

Sandelin, A., Wasserman, W.W., & Lenhard, B., (2004)

ConSite: web-based prediction of regulatory elements using cross-species comparison

Nucleic Acids Research, 32, W249-W252

Sauerwald, T.M., Betenbaugh, M.J., & Oyler, G.A., (2002)

Inhibiting apoptosis in mammalian cell culture using the caspase inhibitor XIAP and deletion mutants

Biotechnology and Bioengineering, 77, 6, 704-716

Sauerwald, T.M., Oyler, G.A., & Betenbaugh, M.J., (2003)

Study of caspase inhibitors for limiting death in mammalian cell culture

Biotechnology and Bioengineering, 81, 3, 329-340

Sautter, K. & Enenkel, B., (2005)

Selection of high-producing CHO cells using NPT selection marker with reduced enzyme activity

Biotechnology and Bioengineering, 89, 5, 530-538

Schieke, S.M., Phillips, D., McCoy, J.P., Jr., Aponte, A.M., Shen, R.F., Balaban, R.S., & Finkel, T., (2006)

The mammalian target of rapamycin (mTOR) pathway regulates mitochondrial oxygen consumption and oxidative capacity

Journal of Biological Chemistry, 281, 37, 27643-27652

Schimke, R.T., (1984)

Gene amplification in cultured animal-cells

Cell, 37, 3, 705-713

Schlabach, M.R., Hu, J.K., Li, M., & Elledge, S.J., (2010)

Synthetic design of strong promoters

Proceedings of the National Academy of Sciences of the United States of America, 107, 6, 2538-2543

Schlaeger, E.J., Kitas, E.A., & Dorn, A., (2003)

SEAP expression in transiently transfected mammalian cells grown in serum-free suspension culture

Cytotechnology, 42, 1, 47-55

Schlatter, S., Bailey, J.E., & Fussenegger, M., (2001)

Novel surface tagging technology for selection of complex proliferation-controlled mammalian cell phenotypes

Biotechnology and Bioengineering, 75, 5, 597-606

Schlatter, S., Rimann, M., Kelm, J., & Fussenegger, M., (2002)

SAMY, a novel mammalian reporter gene derived from Bacillus stearothermophilus alpha-amylase

Gene, 282, 1-2, 19-31

Schroder, M. & Friedl, P., (1997)

Overexpression of recombinant human antithrombin III in Chinese hamster ovary cells results in malformation and decreased secretion of recombinant protein

Biotechnology and Bioengineering, 53, 6, 547-559

Schroder, M., Korner, C., & Friedl, P., (1999)

Quantitative analysis of transcription and translation in gene amplified Chinese hamster ovary cells on the basis of a kinetic model

Cytotechnology, 29, 2, 93-102

Serpieri, F., Inocencio, A., de Oliveira, J.M., Pimenta, A.A., Garbuio, A., Kalil, J., Brigido, M.M., & Moro, A.M., (2010)

Comparison of humanized IgG and FvFc anti-CD3 monoclonal antibodies expressed in CHO cells

Molecular Biotechnology, 45, 3, 218-225

Shaffer, A.L., Shapiro-Shelef, M., Iwakoshi, N.N., Lee, A.H., Qian, S.B., Zhao, H., Yu, X., Yang, L.M., Tan, B.K., Rosenwald, A., Hurt, E.M., Petroulakis, E., Sonenberg, N., Yewdell, J.W., Calame, K., Glimcher, L.H., & Staudt, L.M., (2004)

XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation

Immunity, 21, 1, 81-93

Shaikhibrahim, Z. & Wernert, N., (2012)

ETS transcription factors and prostate cancer: The role of the family prototype ETS-1

International Journal of Oncology, 40, 6, 1748-1754

Shatkin, A.J. & Manley, J.L., (2000)

The ends of the affair: Capping and polyadenylation

Nature Structural Biology, 7, 10, 838-842

Shaulian, E. & Karin, M., (2001)

AP-1 in cell proliferation and survival

Oncogene, 20, 19, 2390-2400

Shearwin, K.E., Callen, B.P., & Egan, J.B., (2005)

Transcriptional interference - a crash course

Trends in Genetics, 21, 6, 339-345

Shibui-Nihei, A., Ohmori, Y., Yoshida, K., Imai, J., Oosuga, I., Iidaka, M., Suzuki, Y., Mizushima-Sugano, J., Yoshitomo-Nakagawa, K., & Sugano, S., (2003)

The 5' terminal oligopyrimidine tract of human elongation factor 1A-1 gene functions as a transcriptional initiator and produces a variable number of Us at the transcriptional level

Gene, 311, 137-145

Shuman, S., (2002)

What messenger RNA capping tells us about eukaryotic evolution

Nature Reviews Molecular Cell Biology, 3, 8, 619-625

Simari, R.D., Yang, Z.Y., Ling, X., Stephan, D., Perkins, N.D., Nabel, G.J., & Nabel, E.G., (1998)

Requirements for enhanced transgene expression by untranslated sequences from the human cytomegalovirus immediate-early gene

Molecular Medicine, 4, 11, 700-706

Smith, J.D. & Robinson, A.S., (2002)

Overexpression of an archaeal protein in yeast: Secretion bottleneck at the ER

Biotechnology and Bioengineering, 79, 7, 713-723

Smith, M.C.M. & Thorpe, H.M., (2002)

Diversity in the serine recombinases

Molecular Microbiology, 44, 2, 299-307

Soares, D.C. & Abbott, C.M., (2013)

Highly homologous eEF1A1 and eEF1A2 exhibit differential post-translational modification with significant enrichment around localised sites of sequence variation

Biology Direct, 8, 1, 29

Solem, C. & Jensen, P.R., (2002)

Modulation of gene expression made easy

Applied and Environmental Microbiology, 68, 5, 2397-2403

Solem, C., Koebmann, B., & Jensen, P.R., (2008)

Control analysis of the role of triosephosphate isomerase in glucose metabolism in Lactococcus lactis

IET Systems Biology, 2, 2, 64-72

Song, S., Morgan, M., Ellis, T., Poirier, A., Chesnut, K., Wang, J.M., Brantly, M., Muzyczka, N., Byrne, B.J., Atkinson, M., & Flotte, T.R., (1998)

Sustained secretion of human alpha-1-antitrypsin from murine muscle transduced with adeno-associated virus vectors

Proceedings of the National Academy of Sciences of the United States of America, 95, 24, 14384-14388

Stanfel, M.N., Moses, K.A., Schwartz, R.J., & Zimmer, W.E., (2005)

Regulation of organ development by the NKX-homeodomain factors: an NKX code

Cell & Molecular Biology, Suppl 51, OL785-799

Stapleton, J.A., Endo, K., Fujita, Y., Hayashi, K., Takinoue, M., Saito, H., & Inoue, T., (2012)

Feedback control of protein expression in mammalian cells by tunable synthetic translational inhibition

ACS Synthetic Biology, 1, 3, 83-88

Stenberg, R.M., Witte, P.R., & Stinski, M.F., (1985)

Multiple spliced and unspliced transcripts from human cytomegalovirus immediate-early region 2 and evidence for a common initiation site within immediate-early region 1

Journal of Virology, 56, 3, 665-675

Stinski, M.F. & Roehr, T.J., (1985)

Activation of the major immediate early gene of human cytomegalovirus by cis-acting elements in the promoter-regulatory sequence and by virus-specific trans-acting components

Journal of Virology, 55, 2, 431-441

Stricker, J., Cookson, S., Bennett, M.R., Mather, W.H., Tsimring, L.S., & Hasty, J., (2008)

A fast, robust and tunable synthetic gene oscillator

Nature, 456, 7221, 516-519

Sudarsanam, S. & Johnson, D.E., (2010)

Functional consequences of mTOR inhibition

Current Opinion in Drug Discovery & Development, 13, 1, 31-40

Sung, Y.H., Lee, J.S., Park, S.H., Koo, J., & Lee, G.M., (2007)

Influence of co-down-regulation of caspase-3 and caspase-7 by siRNAs on sodium butyrate-induced apoptotic cell death of Chinese hamster ovary cells producing thrombopoietin

Metabolic Engineering, 9, 5-6, 452-464

Sung, Y.H., Lim, S.W., Chung, J.Y., & Lee, G.M., (2004)

Yeast hydrolysate as a low-cost additive to serum-free medium for the production of human thrombopoietin in suspension cultures of Chinese hamster ovary cells

Applied Microbiology and Biotechnology, 63, 5, 527-536

Sunstrom, N.A., Gay, R.D., Wong, D.C., Kitchen, N.A., DeBoer, L., & Gray, P.P., (2000)

Insulin-like growth factor-I and transferrin mediate growth and survival of Chinese hamster ovary cells

Biotechnology Progress, 16, 5, 698-702

Suske, G., Bruford, E., & Philipsen, S., (2005)

Mammalian SP/KLF transcription factors: bring in the family

Genomics, 85, 5, 551-556

Sweet, B.H. & Hilleman, M.R., (1960)

The vacuolating virus, SV40

Proceedings of the Society for Experimental Biology and Medicine, 105, 2, 420-427

Tabakin-Fix, Y., Huleihel, M., & Aboud, M., (2004)

Activation of simian virus 40 promoter by HTLV-I Tax protein: role of NF-kappaB and CBP

Biochemical and Biophysical Research Communications, 318, 4, 1052-1056

Tan, Z., Zhang, S., Li, M., Wu, X., Weng, H., Ding, Q., Cao, Y., Bao, R., Shu, Y., Mu, J., Ding, Q., Wu, W., Yang, J., Zhang, L., & Liu, Y., (2013)

Regulation of cell proliferation and migration in gallbladder cancer by zinc finger X-chromosomal protein

Gene, 528, 2, 261-266

Tey, B.T., Singh, R.P., Piredda, L., Piacentini, M., & Al-Rubeai, M., (2000)

Influence of Bcl-2 on cell death during the cultivation of a Chinese hamster ovary cell line expressing a chimeric antibody

Biotechnology and Bioengineering, 68, 1, 31-43

Thiagalingam, A., DeBustros, A., Borges, M., Jasti, R., Compton, D., Diamond, L., Mabry, M., Ball, D.W., Baylin, S.B., & Nelkin, B.D., (1996)

RREB-1, a novel zinc finger protein, is involved in the differentiation response to ras in human medullary thyroid carcinomas

Molecular and Cellular Biology, 16, 10, 5335-5345

Thompson, A.J. & Myatt, S.C., (1997)

Tetracycline-dependent activation of an upstream promoter reveals transcriptional interference between tandem genes within T-DNA in tomato

Plant Molecular Biology, 34, 4, 687-692

Thompson, E.M., Nagata, S., & Tsuji, F.I., (1990)

Vargula-Hilgendorfii luciferase - a secreted reporter enzyme for monitoring gene-expression in mammalian-cells

Gene, 96, 2, 257-262

Tigges, M. & Fussenegger, M., (2006)

Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells

Metabolic Engineering, 8, 3, 264-272

Tigges, M., Marquez-Lago, T.T., Stelling, J., & Fussenegger, M., (2009)

A tunable synthetic mammalian oscillator

Nature, 457, 7227, 309-312

Tornoe, J., Kusk, P., Johansen, T.E., & Jensen, P.R., (2002)

Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites

Gene, 297, 1-2, 21-32

Trcek, T., Larson, D.R., Moldon, A., Query, C.C., & Singer, R.H., (2011)

Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast

Cell, 147, 7, 1484-1497

Trimarchi, J.M. & Lees, J.A., (2002)

Sibling rivalry in the E2F family

Nature Reviews Molecular Cell Biology, 3, 1, 11-20

Ubeda, M. & Habener, J.F., (2000)

CHOP gene expression in response to endoplasmic-reticular stress requires NFY interaction with different domains of a conserved DNA-binding element

Nucleic Acids Research, 28, 24, 4987-4997

Uetsuki, T., Naito, A., Nagata, S., & Kaziro, Y., (1989)

Isolation and characterization of the human chromosomal gene for polypeptide chain elongation factor-1 alpha

Journal of Biological Chemistry, 264, 10, 5791-5798

Urlaub, G. & Chasin, L.A., (1980)

Isolation of Chinese-hamster cell mutants deficient in dihydrofolate-reductase activity

Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 77, 7, 4216-4220

Vogel, C. & Marcotte, E.M., (2012)

Insights into the regulation of protein abundance from proteomic and transcriptomic analyses

Nature Reviews Genetics, 13, 4, 227-232

Waby, J.S., Bingle, C.D., & Corfe, B.M., (2008)

Post-translational control of Sp-family transcription factors

Current Genomics, 9, 5, 301-311

Wakabayashiito, N. & Nagata, S., (1994)

Characterization of the regulatory elements in the promoter of the human Elongation Factor-1-Alpha gene

Journal of Biological Chemistry, 269, 47, 29831-29837

Walsh, G., (2006)

Biopharmaceutical benchmarks 2006

Nature Biotechnology, 24, 7, 769-776

Walsh, G., (2010)

Biopharmaceutical benchmarks 2010

Nature Biotechnology, 28, 9, 917-924

Walsh, G. & Jefferis, R., (2006)

Post-translational modifications in the context of therapeutic proteins

Nature Biotechnology, 24, 10, 1241-1252

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., & Church, G.M., (2009)

Programming cells by multiplex genome engineering and accelerated evolution

Nature, 460, 7257, 894-898

Wang, P.X., Yang, J., Ishihama, A., & Pittard, A.J., (1998)

Demonstration that the TyrR protein and RNA polymerase complex formed at the divergent P3 promoter inhibits binding of RNA polymerase to the major promoter, P1, of the aroP gene of Escherichia coli

Journal of Bacteriology, 180, 20, 5466-5472

Wayte, J., Boraston, R., Bland, H., Varley, J., & Brown, M., (1997)

pH: Effects on growth and productivity of cell lines producing monoclonal antibodies: Control in large-scale fermenters

Genetic Engineer and Biotechnologist, 17, 2-3, 125-132

Weber, W., Stelling, J., Rimann, M., Keller, B., Daoud-El Baba, M., Weber, C.C., Aubel, D., & Fussenegger, M., (2007)

A synthetic time-delay circuit in mammalian cells and mice

Proceedings of the National Academy of Sciences of the United States of America, 104, 8, 2643-2648

Westwood, A.D., Rowe, D.A., & Clarke, H.R.G., (2010)

Improved recombinant protein yield using a codon deoptimized DHFR selectable marker in a CHEF1 expression plasmid

Biotechnology Progress, 26, 6, 1558-1566

Wiberg, F.C., Rasmussen, S.K., Frandsen, T.P., Rasmussen, L.K., Tengbjerg, K., Coljee, V.W., Sharon, J., Yang, C.Y., Bregenholt, S., Nielsen, L.S., Haurum, J.S., & Tolstrup, A.B., (2006)

Production of target-specific recombinant human polyclonal antibodies in mammalian cells

Biotechnology and Bioengineering, 94, 2, 396-405

Wildeman, A.G., (1988)

Regulation of SV40 early gene expression

Biochemistry and Cell Biololgy, 66, 6, 567-577

Wong, D.C.F., Wong, K.T.K., Lee, Y.Y., Morin, P.N., Heng, C.K., & Yap, M.G.S., (2006a)

Transcriptional profiling of apoptotic pathways in batch and fed-batch CHO cell cultures

Biotechnology and Bioengineering, 94, 2, 373-382

Wong, D.C.F., Wong, K.T.K., Nissom, P.M., Heng, C.K., & Yap, M.G.S., (2006b)

Targeting early apoptotic genes in batch and fed-batch CHO cell cultures

Biotechnology and Bioengineering, 95, 3, 350-361

Wu, J., O'Neill, J., & Barbosa, M.S., (1998)

Transcription factor Sp1 mediates cell-specific trans-activation of the human cytomegalovirus DNA polymerase gene promoter by immediate-early protein IE86 in glioblastoma U373MG cells

Journal of Virology, 72, 1, 236-244

Wurm, F.M., (2004)

Production of recombinant protein therapeutics in cultivated mammalian cells

Nature Biotechnology, 22, 11, 1393-1398

Xia, W., Bringmann, P., McClary, J., Jones, P.P., Manzana, W., Zhu, Y., Wang, S.J., Liu, Y., Harvey, S., Madlansacay, M.R., McLean, K., Rosser, M.P., MacRobbie, J., Olsen, C.L., & Cobb, R.R., (2006)

High levels of protein expression using different mammalian CMV promoters in several cell lines

Protein Expression and Purification, 45, 1, 115-124

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M.R., Neff, N., Passarelli, B., Koh, W., Fan, H.C., Wang, J., Gui, Y., Lee, K.H., Betenbaugh, M.J., Quake, S.R., Famili, I., Palsson, B.O., & Wang, J., (2011)

The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line

Nature Biotechnology, 29, 8, 735-741

Xu, Z.L., Mizuguchi, H., Ishii-Watabe, A., Uchida, E., Mayumi, T., & Hayakawa, T., (2002)

Strength evaluation of transcriptional regulatory elements for transgene expression by adenovirus vector

Journal of Controlled Release, 81, 1-2, 155-163

Yang, F., Moss, L.G., & Phillips, G.N., (1996)

The molecular structure of green fluorescent protein

Nature Biotechnology, 14, 10, 1246-1251

Yasuda, K., Hirayoshi, K., Hirata, H., Kubota, H., Hosokawa, N., & Nagata, K., (2002)

The Kruppel-like factor Zf9 and proteins in the Sp1 family regulate the expression of HSP47, a collagen-specific molecular chaperone

Journal of Biological Chemistry, 277, 47, 44613-44622

Ye, J., Rawson, R.B., Komuro, R., Chen, X., Dave, U.P., Prywes, R., Brown, M.S., & Goldstein, J.L., (2000)

ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs

Molecular Cell, 6, 6, 1355-1364

Ye, J.X., Alvin, K., Latif, H., Hsu, A., Parikh, V., Whitmer, T., Tellers, M., Edmonds, M.C.D., Ly, J., Salmon, P., & Markusen, J.F., (2010)

Rapid protein production using CHO stable transfection pools

Biotechnology Progress, 26, 5, 1431-1437

Ye, Z.Q., Qiu, P., Burkholder, J.K., Turner, J., Culp, J., Roberts, T., Shahidi, N.T., & Yang, N.S., (1998)

Cytokine transgene expression and promoter usage in primary CD34+ cells using particle-mediated gene delivery

Human Gene Therapy, 9, 15, 2197-2205

Yim, S.S., An, S.J., Kang, M., Lee, J., & Jeong, K.J., (2013)

Isolation of fully synthetic promoters for high-level gene expression in Corynebacterium glutamicum

Biotechnology and Bioengineering, 110, 11, 2959-2969

Yoshida, H., Haze, K., Yanagi, H., Yura, T., & Mori, K., (1998)

Identification of the cis-acting endoplasmic reticulum stress response element responsible for transcriptional induction of mammalian glucose-regulated proteins. Involvement of basic leucine zipper transcription factors

Journal of Biological Chemistry, 273, 50, 33741-33749

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., & Mori, K., (2001a)

XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor

Cell, 107, 7, 881-891

Yoshida, H., Okada, T., Haze, K., Yanagi, H., Yura, T., Negishi, M., & Mori, K., (2001b)

Endoplasmic reticulum stress-induced formation of transcription factor complex ERSF including NF-Y (CBF) and activating transcription factors 6 alpha and 6 beta that activates the mammalian unfolded protein response

Molecular and Cellular Biology, 21, 4, 1239-1248

Yoshikawa, T., Nakanishi, F., Ogura, Y., Oi, D., Omasa, T., Katakura, Y., Kishimoto, M., & Suga, K., (2001)

Flow cytometry: An improved method for the selection of highly productive gene-amplified CHO cells using flow cytometry

Biotechnology and Bioengineering, 74, 5, 435-442

Yun, C.Y., Liu, S., Lim, S.F., Wang, T.H., Chung, B.Y.F., Teo, J.J., Chuan, K.H., Soon, A.S.C., Goh, K.S., & Song, Z.W., (2007)

Specific inhibition of caspase-8 and-9 in CHO cells enhances cell viability in batch and fed-batch cultures

Metabolic Engineering, 9, 5-6, 406-418

Yurochko, A.D., Mayo, M.W., Poma, E.E., Baldwin, A.S., Jr., & Huang, E.S., (1997)

Induction of the transcription factor Sp1 during human cytomegalovirus infection mediates upregulation of the p65 and p105/p50 NF-kappaB promoters

Journal of Virology, 71, 6, 4638-4648

Zahn-Zabal, M., Kobr, M., Girod, P.A., Imhof, M., Chatellard, P., de Jesus, M., Wurm, F., & Mermod, N., (2001)

Development of stable cell lines for production or regulated expression using matrix attachment regions

Journal of Biotechnology, 87, 1, 29-42

Zarrin, A.A., Malkin, L., Fong, I., Luk, K.D., Ghose, A., & Berinstein, N.L., (1999)

Comparison of CMV, RSV, SV40 viral and Vlambda1 cellular promoters in B and T lymphoid and non-lymphoid cell lines

Biochimica et Biophysica Acta, 1446, 1-2, 135-139

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., & Arlotta, P., (2011)

Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription

Nature Biotechnology, 29, 2, 149-153

Zhou, W.C., Rehm, J., Europa, A., & Hu, W.S., (1997)

Alteration of mammalian cell metabolism by dynamic nutrient feeding

Cytotechnology, 24, 2, 99-108

Zhou, W.C., Rehm, J., & Hu, W.S., (1995)

High viable cell concentration fed-batch cultures of hybridoma cells through online nutrient feeding

Biotechnology and Bioengineering, 46, 6, 579-587

Zhou, Y.X., Su, Z.P., Huang, Y.L., Sun, T., Chen, S.S., Wu, T.F., Chen, G.L., Xie, X.S., Li, B., & Du, Z.W., (2011)

The Zfx gene is expressed in human gliomas and is important in the proliferation and apoptosis of the human malignant glioma cell line U251

Journal of Experimental & Clinical Cancer Research, 30

Zou, J., Milon, B.C., Desouki, M.M., Costello, L.C., & Franklin, R.B., (2011)

hZIP1 zinc transporter down-regulation in prostate cancer involves the overexpression of ras responsive element binding protein-1 (RREB-1)

Prostate