Kent Academic Repository

Full text document (pdf)

Citation for published version

Bledowski, Nicole (2022) Development of Reporter Systems of Cellular Readouts of Chinese Hamster Ovary (CHO) Cells. Master of Science by Research (MScRes) thesis, University of Kent,.

DOI

Link to record in KAR

https://kar.kent.ac.uk/98518/

Document Version

UNSPECIFIED

Copyright & reuse
Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research
The version in the Kent Academic Repository may differ from the final published version.
Users are advised to check **http://kar.kent.ac.uk** for the status of the paper. **Users should always cite the published version of record.**

Enquiries
For any further enquiries regarding the licence status of this document, please contact:
researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at **http://kar.kent.ac.uk/contact.html**

KAR Kent Academic Repository

University of Kent

Development of Reporter Systems of Cellular Readouts of Chinese Hamster Ovary (CHO) Cells

Thesis submitted in fulfilment of the requirements for a postgraduate degree at the University of Kent.

University of Kent, Canterbury, England

Degree of Masters in Biochemistry

Nicole Bledowski

2020/21

Abstract

Biological stress is perceived and occurs across all cellular organisms. When cells undergo or perceive changes in homeostasis, such as temperature fluctuations, changes in pH/ pressure or oxidative stress, cellular responses are initiated that results in the production of various molecular players to relieve the stress. Different types of stresses cause the specific initiation of diverse molecular pathways and elements to counter the cellular stress. Such stress is perceived in *in vitro* cultured mammalian cells, such as Chinese hamster ovary (CHO) cells when they are grown in culture to produce secreted biotherapeutic proteins. The perception of different stresses can limit the yield of biotherapeutic protein produced and/or the quality of the protein from such cells. Here, genetic reporter systems that can report on the perception of such stresses have been designed for use in CHO cells during production of biopharmaceuticals so that strategies to prevent or harness such stresses can be developed to enhance recombinant protein production from such cell systems.

A major controlling factor in responding to and combating these cellular stresses begins at the transcriptional level, by the activation of transcription factors (TF), and their complementary DNA binding motifs that regulate gene transcription. The goal of the work presented here was to harness transcription factor binding sites upstream of promoters driving the expression of reporter genes in mammalian cells culture whereby the expression of the fluorescence reporter proteins would provide a real time readout of cellular stress perception that the targeted transcription factor is involved in controlling. Using previously reported transcription factor motif sequences, cloning them into a plasmid vector with selected promoter regions, and multiple cloning sites, reporter gene constructs were developed. A destabilized GFP reporter system was used as the readout as the rapid turnover of this gives a more accurate representation of the current stress perception rather than a historical readout. The reporter constructs were designed, assembled and successfully cloned. The specific stresses and binding sequences utilised targeted transcription factors that were activated upon ER, energy, lipid and oxidative stress. Transfections of a selection of the constructs was then undertaken into CHO-S mammalian cells to establish baseline expression from the reporter constructs as determined by flow cytometry analysis of GFP fluorescence. In particular, carbohydrate response element (ChoRE, energy stress response) and fatty acid stress response (Peroxisome Proliferator-Activated Receptor, PPAR) transcription factor binding sequence containing constructs were transfected and investigated under spent media and oxidative stress conditions. However, under these harsh environmental conditions there was low culture viability and reporter gene expression levels. A further set of transfections were undertaken with the ER stress response (ERSE) TF containing motifs, with varying constructs evaluated that contained different elements, sequence lengths and promotor regions. The cellular external conditions were not altered from the CHO-S maintenance media, and differences in reporter gene expression from these were detected. In conclusion, a toolbox of reporter gene constructs that allow the monitoring in real time of specific stresses on mammalian cells during culture have been designed and developed and some of these have had preliminary validation undertaken. Further analyses can be performed with these under stress conditions to continue to validate and develop a toolbox of stress reporter systems. The potential application of these in CHO cells during the development of cells and bioprocesses for the production of secreted biotherapeutic proteins may help define when such cells perceive different stresses that might limit recombinant protein product yields and quality, allowing the development of systems and processes that can either limit or utilise these responses to further enhance the productivity and quality of product from these cell expression systems.

Acknowledgments:

I would like to thank the following people who have supported me in an enormous way throughout this project:

Professor Mark Smales for supervising this project, reviewing my thesis and for all the support he has given me. Your support has been greatly appreciated for the past year and I wouldn't have made it this far without you.

Dr. James Budge for the amazing training and supervision throughout the laboratory work. For the academic support provided during my time in the Smales Biosciences Facility at University of Kent. All the MSc, and PhD candidates within the shared laboratories have greatly impacted my time during my studies at the university with physical and mental support!

DISCLAIMER:
All figures throughout are original unless otherwise indicated in the figure legend.

TABLE OF CONTENTS

List of abbreviations:

Amp Ampicillin

bp Base Pair

ChoRE Carbohydrate Response Element

CMV Cytomegalovirus

DNA Deoxyribonucleic Acid

ERSE Endoplasmic Reticulum Stress Element

FASN Fatty Acid Synthase

GOI Gene of Interest

Kan Kanamycin

kbp Kilo Base Pair

LB Lysogeny Broth

NF H_2O Nuclease-Free Water

PBS Phosphate-Buffered Saline

PCR Polymerase Chain Reaction

PPAR Peroxisome Proliferator-Activated Receptor

RNA Ribonucleic Acid

SDS Sodium Dodecyl Sulphate

SRE Sterol Regulatory Element

SV40 Simian Virus 40

TBS-T Tris-Buffered Saline and Tween 20

TF Transcription Factor

Chapter 1 Introduction

1.1 DNA and Gene Expression

DNA is the foundation of biochemical systems, reactions and sustainability throughout an organism's lifetime consisting of the coding and non-coding genetic code (Alberts B, Johnson A, Lewis J, et al. 2002.). These strands of deoxyribonucleic acids (DNA) carry the coding information for proteins that 'build' cells and structures, undertake biochemical reactions and help maintain homeostasis throughout the whole cell. The elements within DNA contains the genetic code that may or may not be expressed at any given time depending upon the demands of the cell and the environment. In addition to the coding sequences, it is now known that the non-coding regions of DNA may also play vital roles in controlling the expression of coding regions. The primary expression of any DNA sequence is through transcription of a given DNA sequence to generate RNA; for coding sequences this ultimately results in the synthesis of mRNA that is then subsequently translated into a polypeptide that is then folded, and potentially processed, into a mature protein. Post-translational modifications can aid in the folding and structuring of proteins, which are spread throughout the matrix and cellular components.

1.2 DNA Transcription and its Role in Control of Gene Expression

Transcription of the genetic material is initiated by the production of RNA. Each single strand acts as a template for the new RNA complementary strand (Clancy, Suzanne, 2008.). For coding regions that code for a protein, the RNA from transcription is processed into a mature mRNA that is then translated to generate a polypeptide that will fold to yield a protein. This is called the Central Dogma of Molecular Biology. Figure 1.0 depicts this with DNA being used as a template to produce RNA, then translation of the mRNA to yield a protein.

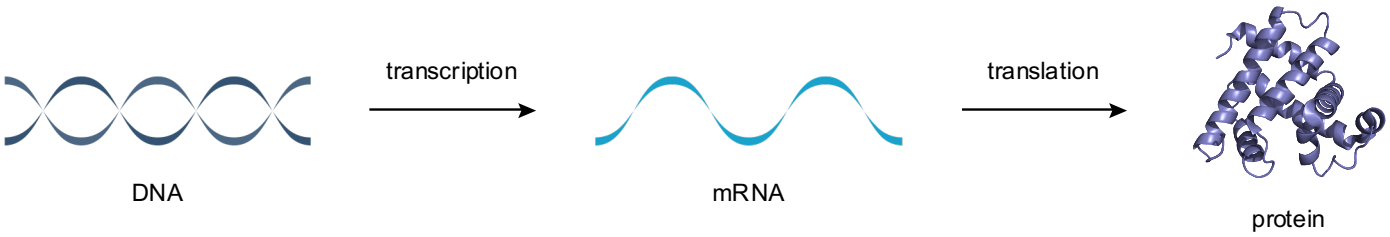


Figure 1.1 The Central Dogma of gene expression; DNA transcription and mRNA translation to produce a protein product. Image derived from "ATDBio - Transcription, Translation

and Replication." *Atdbio.com*, atdbio.com/nucleic-acids-book/Transcription-Translation-and-Replication.

Gene expression is controlled at all levels but particularly at the level of transcription. The structural genes contain sequences that encode for the protein products along with enhancement or repressor regions to regulate the transcription of the gene and thus ultimately control protein production levels within a cell by controlling the amount of mRNA transcript present. Transcription factors (TFs) are proteins that aid in transcribing DNA into RNA by enhancing or repressing the rate or amount of transcription. They do this by binding to specific transcription factor binding sites that then subsequently impact promoter activity of the target gene. Generally, TF's regulate and initiate the transcription of genes by interacting with the DNA and transcriptional machinery; binding to the DNA is via specific binding domains within the TFs that bind to specific DNA sequences. The binding domains bind to DNA transcription factor sequences and can subsequently stimulate or repress expression of the *gene of interest* (GOI) by interaction with the transcriptional machinery. Regulation of transcription of DNA to generate RNA, is the most common form of gene expression control. Such control allows for regulated expression of specific genes only when the products are required such that genes can be silent or active, for example during development or the cell division cycle. Some of the general key players in the control of DNA transcription are outlined in Figure 1.2. These include enhancer regions, where the DNA helix can fold onto itself, enhancing the transcription of the gene. A TATA box is a DNA sequence associated with many core promoter sequences to which the polymerase is recruited and also is close to where the gene transcription is initiated (Phillips. 2008). It is also identifying the direction of the transcription activity through the recruiting of the polymerase enzyme. As well as being used to coordinate expression of genes essential for cell survival and development, transcription factors are also important in activating genes and pathways required to respond to various stresses as they are perceived by cells. In this way, internal and external influences can alter the function of cellular processes and affect the controlling factors of transcription.

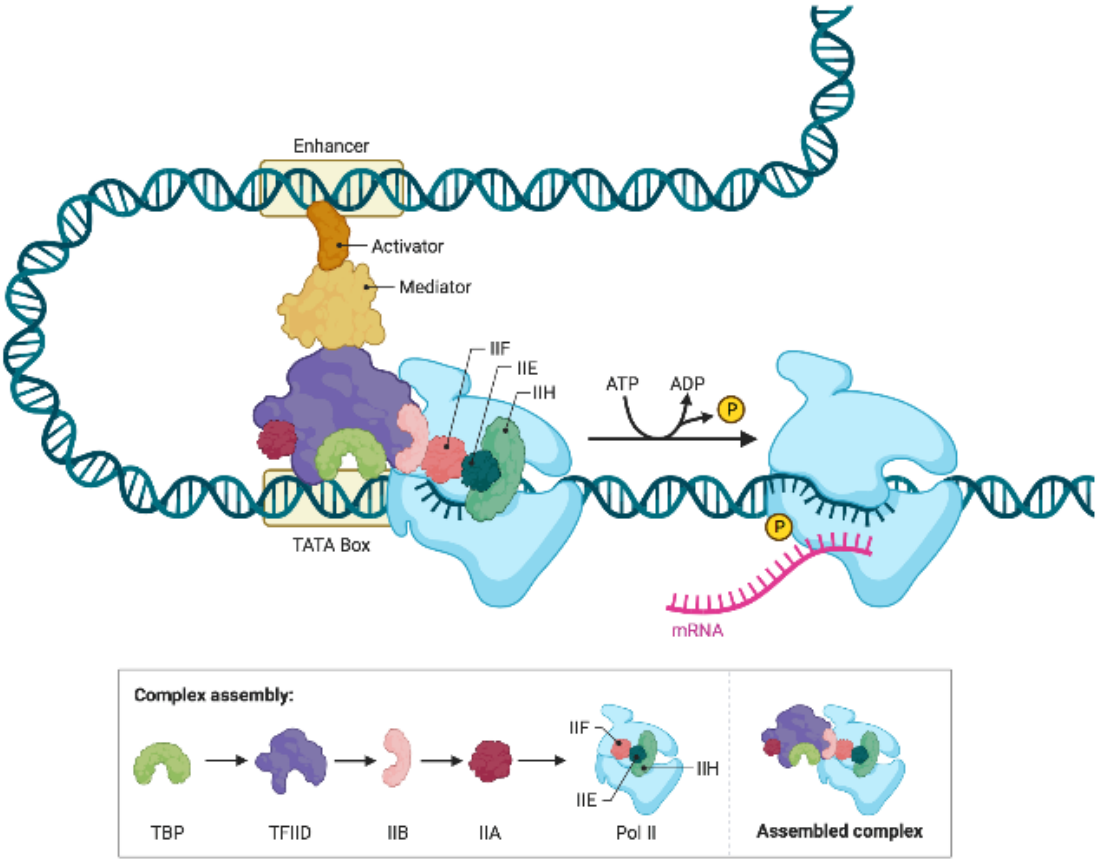


Figure 1.2 How transcription factors can regulate DNA transcription and control gene expression. The example shown is of enhancers and indicates how a transcription factor (activator) can bind a specific DNA enhancer sequence and then via interaction with a mediator and the transcriptional machinery enhance recruitment of the machinery to the core TATA promoter box and increase transcription of the downstream gene. Also shown are other key components of the transcriptional machinery including initiation and regulatory regions, such as enhancers, TATA box, activators and mediators. (Image created using biorender.com).

1.3 Transcription Factors, Transcription and Control of Gene Expression

A cell needs to coordinate hundreds or thousands of cellular activities going on at any one time and to respond to perceived perturbations and demands. As described in *Section 1.2* above, transcriptional control is essential to tight control of gene expression such that genes are only expressed when required to be. As outlined above, the expression of a coding gene starts with transcription of the DNA to yield an RNA copy by RNA polymerases. RNA polymerases are recruited to promoters in the 5' direction to the sequence of genes (Borkotoky, S., & Murali, A., 2018). Promoters are DNA sequences that recruit the protein machinery that undertakes the transcription of the DNA to generate RNA. Transcription factors (TFs), as outlined in *Section 1.2*,

are specialized proteins that help either turn transcription of a particular gene on or off, or modulate the level of transcription, by binding to DNA sequences usually nearby or directly adjacent to the promoter sequence that recruits the RNA polymerase (Borukhov, S., & Nudler, E., 2008). TFs thus bind so termed DNA regulatory sequences and can activate (enhancers) or repress (silencers) gene expression.

A DNA regulatory motif is a sequence on the DNA strand to which binding of the TF proteins occurs. These can vary in name and length as most genes have different transcriptional elements, or combinations of these, although there are several well characterised transcription factor families whereby members of the family share some structural characteristics (Remenyi, A., Scholer, H., & Wilmanns, M., 2004). As stated above, the binding sites on the DNA for TF's can be located near a gene's promoter, however, some studies suggest that spacers between the motif and activation sequences enhance transcription of genes and some sites can be distant (Guéroult-Bellone et al. 2017). Transcription factors and their associated DNA motifs are therefore essential for regulating gene transcription and ultimately protein synthesis and cellular function.

1.4 Recombinant Gene Expression and its Transcriptional Control

DNA sequences of encoded genes can be synthesised or amplified to produce templates for protein expression. These proteins, when expressed in a foreign cell type or even via exogenous DNA in the native cell type are called recombinant proteins. Proteins can be a readout of the ultimate product of coding gene expression and specific proteins are often used as reporters of gene expression activity. Thus, an increase in transcriptional activity of a recombinant gene is often translated into an increase in protein expression. Reporter gene proteins often used include green fluorescent protein (GFP) due to the fluorescent nature of the protein which can be easily detected and monitored. This can then be used to investigate different promoter sequences and TFs to determine if they increase, decrease or are negligible in gene expression by readout of GFP expression when such sequences are placed upstream of the GFP gene. In this way, control of these elements over recombinant gene expression can be evaluated rapidly, and in real time.

1.5 Cloning and Design of Recombinant Genes for Expression in Mammalian Cell Systems

Transferring a fragment of DNA or gene of interest (GOI) from one organism to a genetic expression vector, which can be a plasmid, is referred to as the process of *cloning*. Figure 1.3 depicts this process using an example from the initial cloning steps to produce one of the desired plasmid vectors in the work reported here. Using specific restriction enzyme sites within the plasmid vector, the GOI can be cloned to generate a transformed recombinant plasmid (Figure 1.3). For efficient cloning and subsequent expression of the GOI, an expression vector must contain key elements. These include: the GOI expression cassette (a promoter region and a termination sequence or pol(A) signal), origin of replication (ori), antibiotic resistance cassette for the host, and an antibiotic resistance gene for *Escherichia coli* (as this is usually used to amplify the plasmid DNA). For the purposes of this study, the plasmid vector contained a hygromycin antibiotic resistance gene for use in mammalian cells to select for cells that had the plasmid integrated into the genome, along with several supplementary clonal elements. The additional elements also included a multiple cloning site (MCS), a site where the sequence contains numerous restriction sites. Figure 1.3 also illustrates epitope tags for antibodies being incorporated in the first steps in the cloning process such that proteins with these tags at the terminal could be detected using immunological techniques and antibodies specific to these tags rather than requiring an antibody specifically for the protein encoded by the GOI.

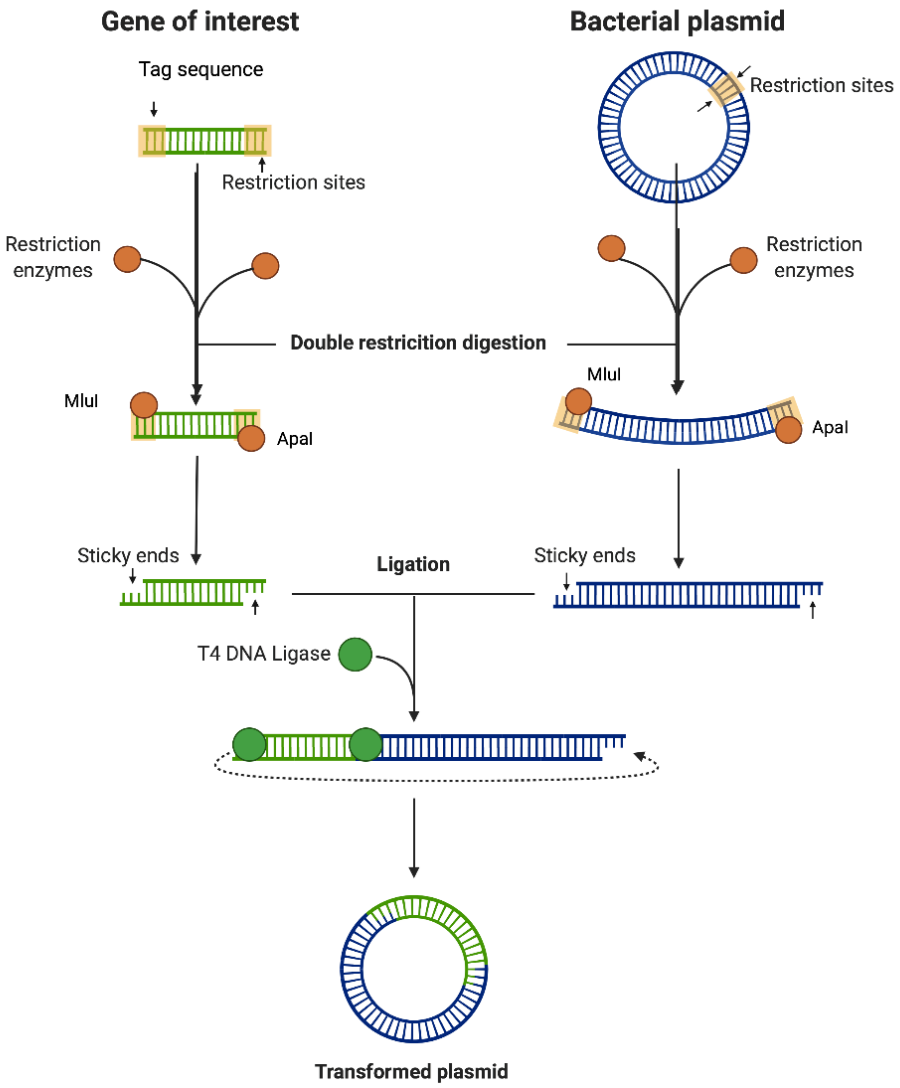


Figure 1.3 Cloning of a GOI into a genetic plasmid vector. Using restriction enzymes, here *MluI* and *ApaI*, the Tag sequence can be cut out of the PCR product, and the bacterial donor plasmid, to produce DNA sticky ends. These sticky ends can then, with the help of a Ligase enzyme, anneal the two products together to form the desired transformed plasmid. (Image created using biorender.com).

Epitope tags are predominantly used to allow for easy detection or even purification of the generated protein of interest. This is achieved by the fusion of a short coding sequence with the GOI, most often at the N- or C-termini of the sequence. There are multiple types of tags with varying application systems such as: use for detection, purification, and cleavage. Some tags serve as sequences to facilitate purification and others as protease recognition sites for the removal of

the tag sequence after detection purposes such that an antibody that recognised the included tag would allow immunodetection of the target protein without the need to raise or find an antibody specific to the target protein.

Gene expression and introduction of a recombinant gene into mammalian cells is usually achieved by cloning the appropriate GOI into a bacterial plasmid. Plasmids already have various elements required to amplify the DNA in a bacterial system and then express the required gene(s) in the mammalian system of choice. This can include a bacteriophage T7 promoter within the plasmid vector allowing for a high level of expression in bacterial systems compared to other expression systems in bacterial plasmids. Table 1.1 depicts the various types of promoter systems often utilised for gene expression and protein production in bacteria in plasmids of interest.

Table 1.1 Bacterial promoter recombinant gene expression systems offered by Thermo Fisher Scientific. (Table derived from Thermo Fisher).

Expression system	Expression level	Tight regulation of expression	Titratability
pTrc	++	+	++
T7	++/+++	+	+
Champion pET	+++	++	+
pBAD	+	+++	+++

The T7 expression system is ideal for expressing recombinant proteins in *E. coli*. This small sequence in the plasmid controls expression of genes through the actions of the RNA polymerase which transcribes the gene of interest. Other systems such as additional promoter regions can aid in the expression.

For mammalian cell expressions, promoters that are active in mammalian systems are required and more so for recombinant protein expressions, usually stronger *viral* promoters are used. Common constitutive promoters (always on, not regulated) used in mammalian systems include the cytomegalovirus (CMV), EF-1 alpha (human elongation factor alpha), UbC (human ubiquitin C),

and SV40 (simian virus 40) promoter sequences. For the purposes of this study, CMV and SV40 promoters were used to drive the recombinant gene expression in CHO cells. The CMV promoter sequence is known to be a generally strong promoter sequence in mammalian cells and is thus widely used (Wang et al. 2017). The SV40 promoter has decreased promoter activity in comparison to the CMV promoter and is therefore used when reduced expression is required than that delivered from the CMV promoter (Wang et al. 2016). Both CMV and SV40 promoter sequences are documented to be stable in mammalian CHO cell lines with transgene expression levels being high. Wang and colleagues documented that CMV yielded one of the highest expression levels of recombinant genes with the highest efficiency in transfections. They also demonstrated the SV40 promoter as being an affective promoter during long-term cultivation or for stable cultures. These promoters are constitutive, meaning that the genes are continuously expressed, and timing of the gene expression is not a critical factor for analysis. The alternative is to use inducible promoters that only 'switch on' gene expression in the presence or absence of a compound or external pressure that removes repression of the promoter. Such promoters were not utilised in this study but are useful when it is desirable to be able to control when, and for how long, recombinant gene expression is active for.

1.6 E. coli for Recombinant Gene Expression Studies

Due to the ease of use, production of high yields and simple scalability, prokaryotic systems are often utilized for the expression of recombinant genes and proteins for both study and biopharmaceutical purposes. *Escherichia coli* is often chosen for recombinant protein expression as the host organism. *E. coli* grows fast, is generally inexpensive to handle and can yield very high amounts of recombinant proteins. Different strains of *E. coli* are available with differing attributes that can be specific to protein expression. Toxicity of some recombinant protein productions can affect the inducibility of the plasmid, therefore many strains contain genetic markers to help maintain its inducibility without any degradation. Many of these aspects can be altered when designing the plasmid of interest and can control protein expression by the amount of inducer added, which allows titratability. Complex human proteins that require specific folding, assembly and post-translational modifications are often unable to be made at the quality and amount required due to the inability of *E. coli* to undertake such processes and are hence made in eukaryotic systems, including mammalian cell expression systems. The focus of this study is on development

of reporter systems in mammalian cell systems that are used for the production of biotherapeutic proteins such that the stresses and responses in such systems can be monitored and therefore the majority of the focus in this thesis is on mammalian cells systems.

1.7 Molecular Cloning of Genes of Interest into Expression Plasmids

During the cloning process, genes of interest and variants/mutants can be prepared using PCR, isolated/amplified to clone into an acceptor vector, or be commercially synthesized as a string or library. For the generation of the genes of interest investigated in the study described here, all the above techniques were utilized for the construction and amplification of the target genes or transcription factor sequences. The approaches are detailed in the methods section of the thesis. Once the design and construction of the required plasmids is complete, they need to be amplified for use using an appropriate *E. coli* strain. The process of introducing the DNA into the cell is called transformation. This can be done in various ways, chemically, or through electrocompetent *E. coli*. Chemical transformation is whereby the plasmid DNA is added to the cells and a heat shock opens pores in the *E. coli* cells, allowing the plasmid to enter into the cell. This method is predominantly used as it is inexpensive, in comparison to the electroporation method of transformation. Electroporation uses an electrical shock to the competent cells, making the membrane more permeable to allow uptake of the plasmid vector. Figure 1.4 shows transformation in chemically competent cells, as this method was extensively used throughout this study.

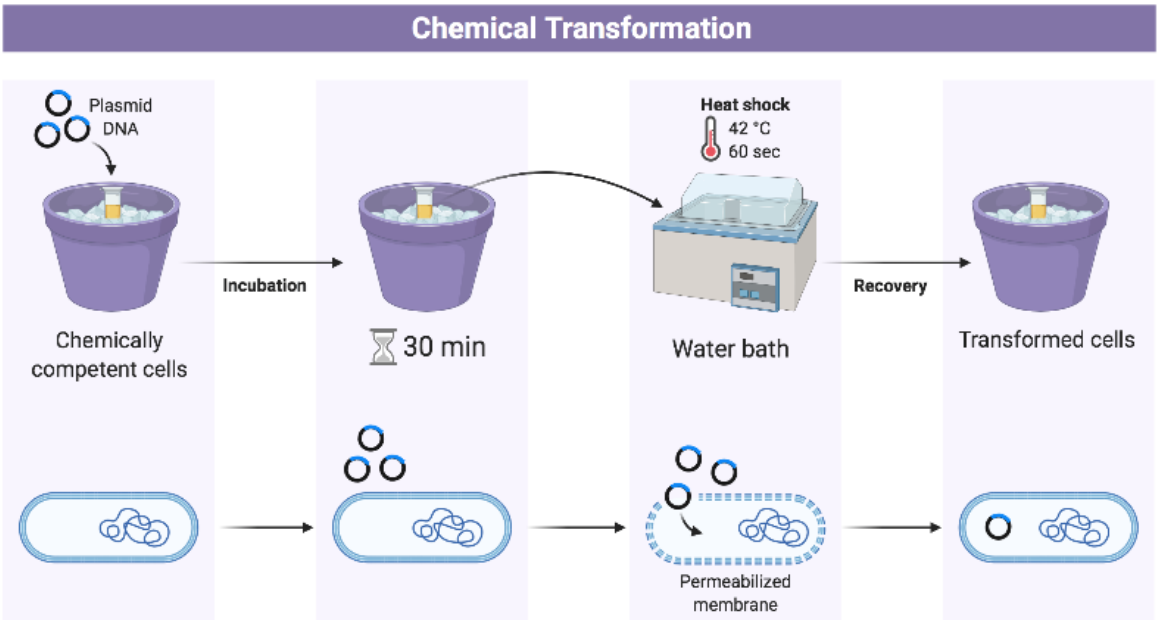


Figure 1.4 Bacterial transformation process through the use of chemically competent cells. The cellular structure depicted at the bottom is a representation of a bacterial cell encompassing its own genomic DNA. Once the membrane is made permeable through heat shocking the cells, it is more able to uptake the plasmid DNA. Image created using biorender.com.

1.8 Recombinant Biotherapeutic Protein Expression and Chinese Hamster Ovary (CHO) cells

Mammalian cell-based recombinant protein expression is a preferred expression platform for the production of many biotherapeutic proteins as they are able to produce correctly folded and assembled products with human-like post-translational modifications, such as glycosylation. The majority of biotherapeutic proteins developed in the last 10-15 years are manufactured in CHO cells (Dumont, J., Euwart, D., Mei, B., Estes, S., & Kshirsagar, R., 2016) in a multi-billion-dollar business. Many of these biotherapeutic proteins are monoclonal antibodies or based on monoclonal antibodies (see Table 1.2). The predominance of mammalian expression systems has continued to grow for the production of biopharmaceuticals as shown in Figure 1.5. Of the mammalian cell expression systems in use commercially, the Chinese hamster ovary (CHO) cell line is by far the most widely utilised (Vcelar, S. et al, 2018). This has become the mammalian cell line of choice because of its ability to produce high yields (>5 g/L) of high-quality recombinant material, its regulatory acceptance and scalability amongst other traits. Despite this, there remains a need to further develop the system to produce more *difficult to express* (DTE) proteins where yields are still low and understand the stress or pressure synthesis of such molecules places on the CHO cell.

Then, this information can be used to further develop or engineer the cell line to limit such stresses and the limitation these might place upon DTE protein yields and quality.

Table 1.2 Monoclonal antibodies that were on the market as of 2017 (from Grilo and Manaaalaris, Trends in Biotechnology (2019) 37(1), 9-16).

mAb	Commercial name	Company		Approval date		Disease target	Refs
		USA	EU	EMA	FDA		
Raxibacumab	AbThrax	GSK[a]	–	–	2012	Anthrax	iii
Tocilizumab	ACTEMRA	Genentech		2009	2010	Rheumatoid arthritis, giant cell arteritis, polyarticular juvenile idiopathic arthritis, systemic juvenile idiopathic arthritis, cytokine release syndrome (USA only)	iv
Adalimumab	AMJEVITA (USA)/ SOLYMBIC (EU)	Amgen		2017	2016	Same as Humira® except fingernail psoriasis; hidradenitis suppurativa and uveitis are EU only	v
Obiltoxaximab	ANTHIM	Elusys Therapeutics	–	–	2016	Anthrax	vi
Ofatumumab	ARZERRA	GSK		2010	2009	Chronic lymphocytic leukemia (CLL)	vii
Bevacizumab	AVASTIN	Genentech		2005	2004	Metastatic colorectal cancer, non-squamous non-small cell lung cancer, recurring glioblastoma, metastatic renal cell carcinoma, cervical cancer, ovarian, Fallopian tube, or peritoneal cancer, breast cancer (EU only), metastatic kidney carcinoma	viii
Avelumab	BAVENCIO	Pfizer/Merck Serono (not disclosed)		2017	2017	Merkel cell carcinoma (MCC)	ix
Belimumab	BENLYSTA	GSK		2011	2011	Lupus erythematosus	x
Blinatumumab	BLINCYTO	Amgen		2015	2014	Acute lymphoblastic leukemia (ALL) in adults and children	xi
Reslizumab	CINQAIR	Teva		2016	2016	Severe asthma	xii
Secukinumab	COSENTYX	Novartis		2015	2015	Plaque psoriasis, psoriatic arthritis, ankylosing spondylitis	xiii
Adalimumab	CYLTEZO	Boehringer		2017	2017	Same as Humira® except fingernail psoriasis; hidradenitis suppurativa and uveitis are EU only	xiv
Ramucirumab	CYRAMZA	Eli Lilly		2014	2014	Gastric or gastroesophageal junction adenocarcinoma, metastatic non-small cell lung cancer, metastatic colorectal cancer	xv
Daratumumab	DARZALEX	Johnson & Johnson		2016	2015	Multiple myeloma	xvi
Dupilumab	DUPIXENT	Sanofi		2017	2017	Atopic dermatitis	xvii
Elotuzumab	EMPLICITI	BMS		2016	2015	Multiple myeloma	xviii
Vedolizumab	ENTYVIO	Takeda		2014	2014	Adult ulcerative colitis, adult Crohn's disease	xix
Cetuximab	ERBITUX	Elli Lilly	Merck Serono	2004	2004	Head and neck cancer, colorectal cancer	xx
Benralizumab	FASENRA	AstraZeneca			2017	Severe asthma	xxi
Infliximab	FLIXABI (EU)/ RENFLEXIS USA)	MSD	Biogen	2016	2017	Same as Remicade®	xxii
Obinutuzumab	GAZYVA (USA)/ GAZYVARO (EU)	Genentech		2017	2013	CLL, follicular lymphoma	xxiii
Emicizumab	HEMLIBRA	Genentech	Roche		2017	Reduce frequency of bleeding episodes in hemophilia A patients	xxiv
Trastuzumab	HERCEPTIN	Genentech		2000	1998	Breast cancer, gastric or gastroesophageal junction adenocarcinoma	xxv
Adalimumab	HUMIRA	Abbvie		2003	2002	Rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, Crohn's disease (adult and children), ulcerative colitis, plaque psoriasis, juvenile idiopathic arthritis, hidradenitis suppurativa, uveitis, fingernail psoriasis (USA only)	xxvi

Table 1. (continued)

mAb	Commercial name	Company		Approval date		Disease target	Refs
		USA	EU	EMA	FDA		
Canakinumab	ILARIS	Novartis		2009	2009	Periodic fever syndromes, systemic juvenile idiopathic arthritis	[xxvii]
Durvalumab	IMFINZI	AstraZeneca	–	–	2017	Metastatic urothelial carcinoma, non-small cell lung cancer	[xxviii]
Adalimumab	IMRALDI	–	Biogen	2017	–	Same as Humira® except fingernail psoriasis	[xxix]
Infliximab	INFLECTRA (USA)/ REMSIMA (EU)[b]	Pfizer	CellTrion	2013	2016	Same as Remicade®	[xxx]
Infliximab	IXIFI	Pfizer	–	–	2017	Same as Remicade®	[xxxi]
Sarilumab	KEVZARA	Sanofi		2017	2017	Rheumatoid arthritis	[xxxii]
Pembrolizumab	KEYTRUDA	MSD		2015	2014	Melanoma, non-small cell lung cancer, head and neck cancer, Hodgkin lymphoma, urothelial carcinoma, microsatellite instability-high solid tumors (USA only), gastric cancer (USA only)	[xxxiii]
Olaratumab	LARTRUVO	Elli Lilly		2016	2016	Soft-tissue sarcoma	[xxxiv]
Alemtuzumab	LEMTRADA/ CAMPATH[c]	Sanofi		2013	2014	Multiple sclerosis (Lemtrada), B cell CLL (Campath, USA only)	[xxxv]
Ranibizumab	LUCENTIS	Genentech	Novartis	2007	2006	Macular degeneration, macular edema, diabetic macular edema, diabetic retinopathy, myopic choroidal neovascularization	[xxxvi]
Bevacizumab	MVASI	Amgen			2017	Same as Avastin®	[xxxvii]
Mepolizumab	NUCALA	GSK		2015	2015	Asthma, eosinophilic granulomatosis with polyangiitis (USA only)	[xxxviii]
Ocrelizumab	OCREVUS	Genentech	Roche		2017	Multiple sclerosis	[xxxix]
Trastuzumab	OGIVRI	Mylan	–	–	2017	Same as Herceptin®	[xl]
Trastuzumab	ONTRUZANT	–	MSD	2017	–	Same as Herceptin®	[xli]
Nivolumab	OPDIVO	BMS		2015	2015	Metastatic melanoma, metastatic non-small cell lung cancer, renal cell carcinoma, Hodgkin lymphoma, head and neck carcinoma, urothelial carcinoma, colorectal cancer (USA only), hepatocellular carcinoma (USA only)	[xlii]
Pertuzumab	PERJETA	Genentech		2013	2012	Breast cancer	[xliii]
Necitumumab	PORTRAZZA	Elli Lilly		2016	2015	Metastatic non-small cell lung cancer	[xliv]
Alirocumab	PRALUENT	Sanofi		2015	2015	Hypercholesterolemia, atherosclerotic cardiovascular disease	[xlv]
Idarucizumab	PRAXBIND	Boehringer		2015	2015	Reversal of the anticoagulant effect of dabigatran	[xlvi]
Denosumab	PROLIA/XGEVA[c]	Amgen		2010	2010	Multiple myeloma (XGEVA), giant cell tumor of bone (XGEVA), hypercalcemia (XGEVA), osteoporosis (Prolia), bone loss in men being treated for prostate cancer and women being treated for breast cancer (Prolia)	Prolia[xlvii] XGEVA[xlviii]
Infliximab	REMICADE	Johnson & Johnson	MSD	1999	1998	Crohn's disease (adult and children), ulcerative colitis (adult and children), rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, plaque psoriasis	[xlix]
Abciximab	REOPRO	Elli Lilly		1993	1994	Prevention of blood clotting	[l]

Table 1. (continued)

mAb	Commercial name	Company		Approval date		Disease target	Refs
		USA	EU	EMA	FDA		
Evolocumab	REPATHA	Amgen		2015	2015	Hypercholesterolemia, myocardial infarction, stroke and coronary revascularization risk reduction	li
Rituximab	RITUXAN/MabThera	Genentech		1998	1997	Non-Hodgkin's lymphoma, chronic lymphocytic leukemia, rheumatoid arthritis, granulomatosis with polyangiitis	lii
Rituximab	RIXATHON	–	Novartis	2017	–	Same as Rituxan/MabThera®	liii
Brodalumab	SILIQ/KYNTHEUM	Valeant	LeoPharma	2017	2017	Plaque psoriasis	liv
Golimumab	SIMPONI/SIMPONI ARIA (USA only)	Johnson & Johnson	MSD	2009	2009	Rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, ulcerative colitis, juvenile idiopathic arthritis (EU only)	lv
Basiliximab	SIMULECT	Novartis		1998	1998	Prevent transplanted kidney rejection	lvi
Eculizumab	SOLIRIS	Alexion		2007	2007	Paroxysmal nocturnal hemoglobinuria, hemolytic uremic syndrome	lvii
Ustekinumab	STELARA	Johnson & Johnson		2009	2009	Plaque psoriasis, psoriatic arthritis, Crohn's disease	lviii
Siltuximab	SYLVANT	Johnson & Johnson		2014	2014	Castleman's disease	lix
Palivizumab	SYNAGIS	AstraZeneca	Abbvie	1999	1998	Respiratory syncytial virus inhibition in children	lx
Ixekizumab	TALTZ	Elli Lilly		2016	2016	Plaque psoriasis, psoriatic arthritis (USA only)	lxi
Atezolizumab	TECENTRIQ	Genentech		2017	2016	Urothelial carcinoma, metastatic non-small cell lung cancer	lxii
Guselkumab	TREMFYA	Johnson & Johnson	–	–	2017	Plaque psoriasis	lxiii
Rituximab	TRUXIMA		CellTrion	2017	–	Same as Rituxan®/MabThera®	lxiv
Natalizumab	TYSABRI	Biogen		2006	2004	Multiple sclerosis, Crohn's disease (USA only)	lxv
Dinutuximab	UNITUXIN[d]	UTC		2015	2015	Pediatric neuroblastoma	lxvi
Panitumumab	VECTIBIX	Amgen		2007	2006	Colorectal cancer	lxvii
Omalizumab	XOLAIR	Genentech	Novartis	2005	2003	Asthma, chronic idiopathic urticaria	lxviii
Ipilimumab	YERVOY	BMS		2011	2011	Melanoma	lxix
Daclizumab	ZINBRYTA[e]	Biogen		2016	2016	Multiple sclerosis	lxx
Bezlotoxumab	ZINPLAVA	MSD		2017	2016	*Clostridium difficile* infections	lxxi

[a]Rights recently sold to Emergent Biosolutions.
[b]Also approved under the names of TRUXIMA, BLITZIMA, RITEMVIA, and RITUZENA by the EMA.
[c]Trade name depends on clinical use as described under clinical indications.
[d]Voluntary withdrawal from the European market in 2017.
[e]Voluntary withdrawal following reports of brain inflammation.

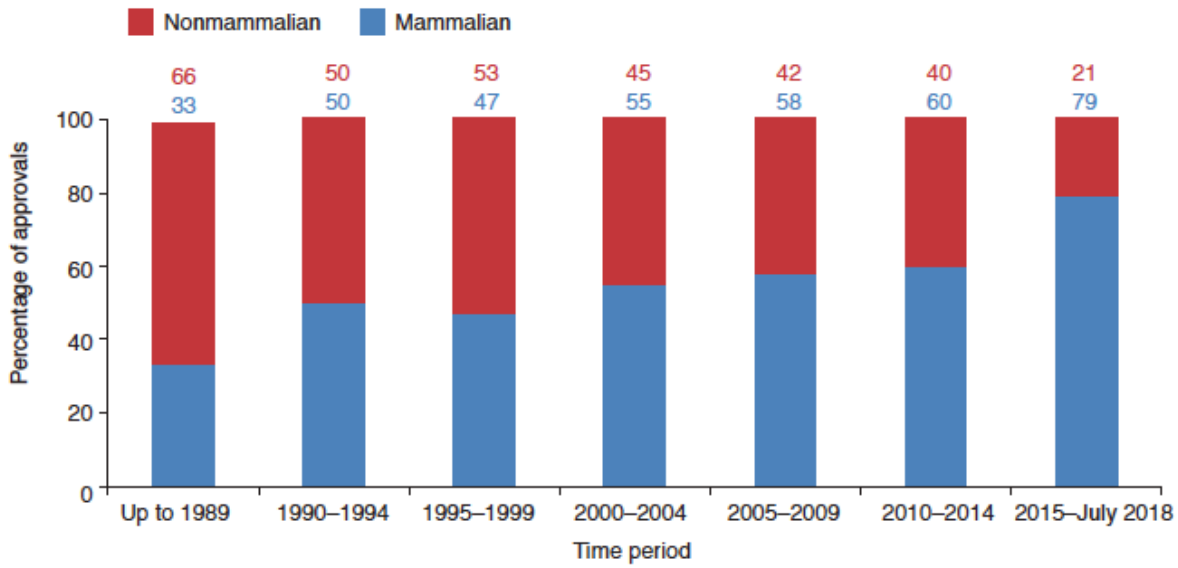


Figure 1.5 Data expressed in mammalian and non-mammalian cells in drug conjugates. The increase in the use of mammalian cell expression systems for new biotherapeutic approved products over time as indicated showing the increasing dominance of mammalian cell expression systems. Image derived from Walsh (2018) *Nature Biotechnology* 36(12) 1136-1145.

Introduction of the target gene (or genes in the case of multiple polypeptide recombinant biotherapeutic proteins such as monoclonal antibodies) is undertaken usually via a plasmid encompassing mammalian genes for the protein production but also, for stable integration into the host genome, a selection marker. Transfection into a host cell can be performed either transiently (where integration does not occur and thus gene expression is for a limited period of time) or to generate and isolate stable cell lines that express the GOI constitutively and permanently. Transient expression results in expression for 2-10 days at most in most cases, which is ideal for rapid protein production of small amounts of material for early research purposes. Stable expression systems require the recombinant genes to be integrated into the host genome. Generation of such cells is time consuming and lengthy (usually anywhere from 3-12 months).

As indicated, the generation of stable cell lines is usually achieved using a metabolic selection marker. This means cells are grown in the absence of a particular metabolite that they cannot survive without, but the introduced plasmid contains a gene that codes for the metabolite. In this way, only cells that have successfully integrated the plasmid DNA into their genome can survive long term via synthesis of the required metabolite whilst those that lack the plasmid DNA die

(Büssow, Konrad., 2015). This gives an indirect selection of cells producing the recombinant protein of interest as the assumption is that those cells that have integrated the election marker will also express those recombinant genes on the same plasmid that was introduced. This results in isolated collections of cells where the cells can express different amounts of the target product and then be further analyzed. These cells must then be cloned, screened to identify and isolate high producing cells from lower producing ones, thus selecting for high generating yields of the product of interest. One of the most common metabolic selection genes utilised in CHO cells is the glutamine synthetase (GS) marker whereby cells are grown in the absence of the amino acid glutamine, but the glutamine synthetase gene is included on the plasmid introduced into the CHO cells for selection purposes (Vcelar, S. & Jadhav, V., et al., 2018).

Various mammalian cell lines can be potential host cells and are being engineered as next generation production cell lines, however, two cell lines have been successful in the research development of many studies (Kim, J. Y., Kim, Y. G., & Lee, G. M., 2012). The use of mammalian cells has been extensively documented in Chinese hamster ovary cells (CHO) as outlined previously, but also in HEK293 (or Human Embryonic Kidney cells). These lines are commonly used in research and industrial settings for their protein compacity for production (S. Estes & Melville, 2014; Hacker et al., 2013; Swiech Picanco-Castro, & Covas, 2012). Whilst the use of CHO and HEK cells has been greatly researched and transformed for the production of stable expression systems, this process is very time consuming. Transient systems provide a viable alternative to producing early-stage quantities of recombinant protein. This method now generally utilizes cells in suspension, meaning that the mammalian culture can be scaled and controlled in a dynamic environment. CHO cell lines are often termed the current gold-standard from biopharmaceutical protein production. Due to the nature of this study, CHO cells were considered for use. These cells, when in optimal conditions and in suspension, have an average doubling time of around 24 hours and can grow to cell densities upwards of 5 million cells per mL. CHO cells can be transfected to produce grams of protein per liter of culture, generated by transient expression systems (Huang et al., 2010; C. Liu, Dalby, Chen, Kilzer, & Chiou, 2008).

Both transient and stable expression of recombinant proteins in cells places an additional gene expression burden on the cells that is not natural or a requirement for the cells to make. Cells

respond to this in a variety of ways by activating pathways to reduce any stress or burden that may be imposed on them, but this can also compromise the potential yields from the cells (Hussain et al 2021). In order to identify when such cell stresses are imposed and which may be limiting in terms of recombinant protein production, 'omic' type studies such as proteomic and transcriptomic approaches have been undertaken to identify bottlenecks in production (Orellana, C. A., 2015 & Aggarwal R. S., 2014). However, these give a static representative readout of cell status in different cell lines at a given time and do not allow real-time monitoring of the 'stress' state of a cell when a recombinant protein burden is placed upon it. This project sets out to develop reporter gene assays that could report on the state of a cell in real time and allow investigations into which stress are perceived and when during recombinant protein production from CHO cells such that, once identified, engineering or bioprocess approaches to remove or reduce these could be designed and tested (Hansen H.G. et al., 2017).

1.9 What Stresses are Perceived during Production and How the Cell Responds

Cells endure many types of stressful events during processing proteins. Temperature, pH, and pressure all affect the nature of the productivity of the cell (Hoffmann, F., & Rinas, U., 2004). During stressful events, the cell can halt protein synthesis caused by temperature shifts, misfolding of protein and oxidative damage. If the protein synthesis machinery would continue to produce protein products during these types of stress, it could cause build-up of products that lead to disease and toxic environments and in the context of biopharmaceutical production, poor quality protein and low yields (Christofk, H. R. et al., 2008 & Vander Heiden, M. G., 2010). The goal for a cell during stress is to maintain homeostasis, and it can accomplish this by employing chaperone molecules to refold the damaged proteins or target them for destruction (da Fonseca, A. C. C., et al. 2021). The protein synthesis pathway is attenuated until the stress dissipates otherwise unfolded proteins can aggregate and impact cell fitness (Reid, David W., et al., 2014). If the stress becomes more prominent and does not cease, such as with the production of helper molecules, cellular senescence may occur or apoptosis (programmed cell death) be initiated (González-Gualda, E., Baker, A. G., Fruk, L., & Muñoz-Espín, D., 2021). Senescence is the arrest of growth which cannot be reversible, this often leads then to cellular death/ apoptosis.

1.9.1 The ER Unfolded Protein Stress Response

All *in vitro* cultured cells will be exposed to or perceive a stress at some stage. When an organism endures a stress, such as a novel chemical occurrence or a physical disruption, many biological defenses are activated/deactivated to help maintain cell homeostasis. Eukaryotes and prokaryotes have systems that can adapt to these stresses by triggering compensatory responses or sending signals for induction of biological stress response mechanisms. Some examples of cellular stress will be discussed to broaden the purpose of the thesis. Such stresses are also encountered in *in vitro* culture cells and similar responses initiated. For example, in CHO cells the recombinant proteins are directed into the ER where they are folded and assembled before passing on through the secretory pathway. This can place a burden on the ER in particular and activate the ER stress response pathway (Maldonado-Agurto R and Dickson AJ., 2018).

Indeed, organelles within the mammalian cell have specific functions and systems to maintain homeostasis. In eukaryotic cells, the Endoplasmic Reticulum (ER) is a central intracellular organelle that is the entry point for the secretory pathway. The ER is where ER tethered ribosomes undertake protein synthesis, feeding the growing polypeptide into the ER (in cases of co-translational translocation), and where folding and some post-translational modifications are undertaken before protein translocation to the Golgi for non-ER resident proteins (Schwartz, et al., 2016). It also serves an important role in calcium storage and lipid metabolism of the cell. Thus, for secreted biotherapeutic proteins made in CHO cells the ER is a crucial organelle that can have stress imposed upon it if the capacity is overwhelmed by recombinant protein demands.

Signalling of the demands on this organelle is coordinated with the Golgi and other parts of the cell as protein production and folding triggers a variety of signalling pathways. Improperly folded proteins can become very problematic if not dealt with, so much so that it can cause cellular death or apoptosis (González-Gualda, E., Baker, A. G., Fruk, L., & Muñoz-Espín, D., 2021). Eukaryotic cells and the ER stress response (ERSR) have been extensively studied as they have an evolutionary adaptive mechanism that can restore homeostasis during times of stress. The ER has several mechanisms that deals with unfolded proteins when they are sensed by signal molecules, called the UPR, *unfolded protein response.* Accumulation of unfolded proteins in the ER causes stress resulting in transcription of genes, and various protein/chaperones that regulate cellular

homeostasis. Under such stresses, a signalling pathway called the unfolded protein response (UPR) is initiated (Figure 1.6).

The sensing of ER stress in mammalian cells is thought to be via a protein called Bip that associates with receptors in the ER membrane or binds to hydrophobic patches on unfolded or misfolded polypeptides and proteins (e.g., ATF6, Figure 1.6). As Bip dissociates from the receptors to bind to unfolded protein, the receptors are activated, resulting in activation of the UPR. The genes activated in response to ER stress include transcription motifs termed ERSE or ER stress response elements, to which transcription factors bind, resulting in induction of transcription of these genes. Such genes include repeated units of the ERSE upstream of the promoter sequence of target UPR genes, as depicted in Figure 1.6b. In this way, ERSE, a consensus sequence of CCAAT-N9-CCACG, can be found to regulate ER homeostasis (Kokame, Kato and Miyata, 2001). Under stress, specific transcription factors that recognise this sequence bind to the ERSE sequence, to promote target gene transcription and activate promoter activity. The multiple copies of ERSE facilitate stronger induction in response to the UPR than a single copy to help address the disturbance of the ER. There are many proteins and chaperones produced in response to ER stress, not discussed here, in response to binding of TFs at such sequence elements. Figure 1.6 compares the evolutionarily conserved tripartite structure of the ERSE units in mammalian cells with yeast such structures.

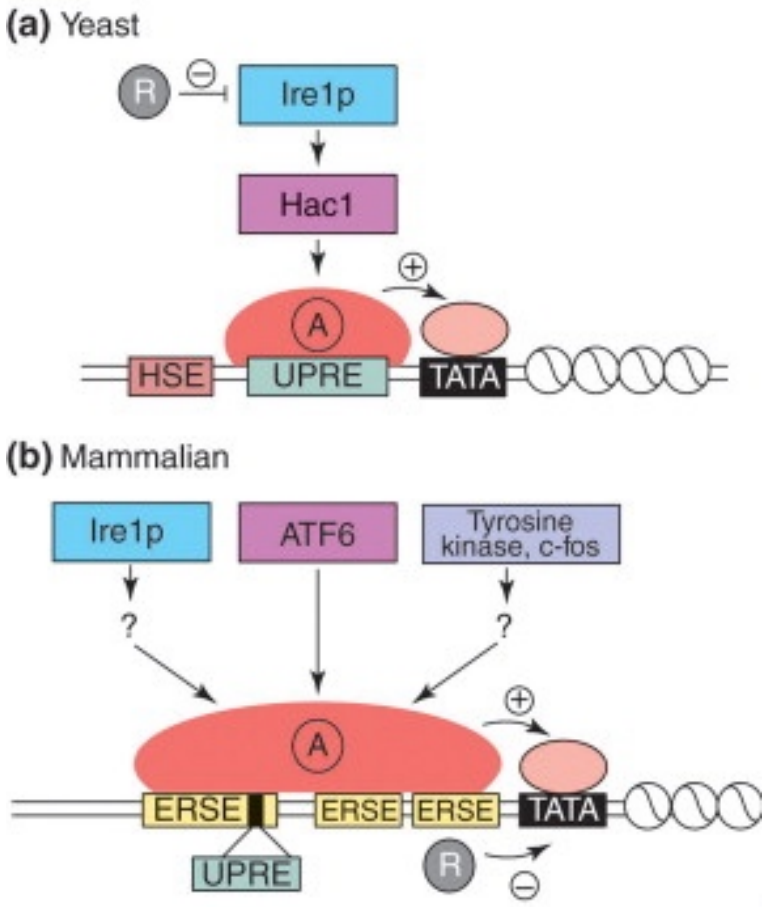


Figure 1.6 Mechanisms of induction of the unfolded protein response in the ER within yeast and mammalian cells. ER stress leads to these players to activate mRNA splicing and translation. Yeast (a) does not contain any ERSE elements as the mammalian (b) cells do. Many downstream targets and pathways are still unknown. Image created using biorender.com.

1.9.2 Cholesterol and Fatty Acid Synthesis Stress Responses

In addition to ER stress, another stress that may be sensed in mammalian cells during bioprocessing is lipid stress. Levels of cholesterol are critical to the structure of the cellular plasma membrane and other membranes of organelles in mammalian cells. Cholesterol maintains fluidity of the membrane and stability, creating a solid structure protecting the inner components of the cell. During production of secretory proteins in CHO cells, vesicle transport can be overloaded as large amounts of recombinant protein are moved through the secretory pathway and such vesicles are made of different lipid constituents. Indeed, manipulation of lipid biosynthesis in CHO cells has recently been reported to influence secretory recombinant protein yields from CHO cells (Budge J.D., 2019). The levels of cholesterol are regulated by a feedback loop mechanism that

controls the transcription of genes for cholesterol synthesis. A stretch of nucleotides upstream of genes called the sterol regulatory element (SRE) is a motif that was found to be obligatory for the transcription of these genes (Inoue et al. 1998). As well as cholesterol, SRE sequences have also been found to have regulatory aspects in general fatty acid synthesis (Figure 1.7) (Kawabe et al. 1996). In human cells, it was found that two SRE motifs and a promoter box sequence termed as the CCAAT box were responsible for the sterol-regulatory transcription of genes in times of stress (Inoue et al. 1998). Low sterol concentrations result in transcription factors binding/interaction with this motif and activation of the genes involved in cholesterol biosynthesis and lipid homeostasis (Amemiya et al. 2002).

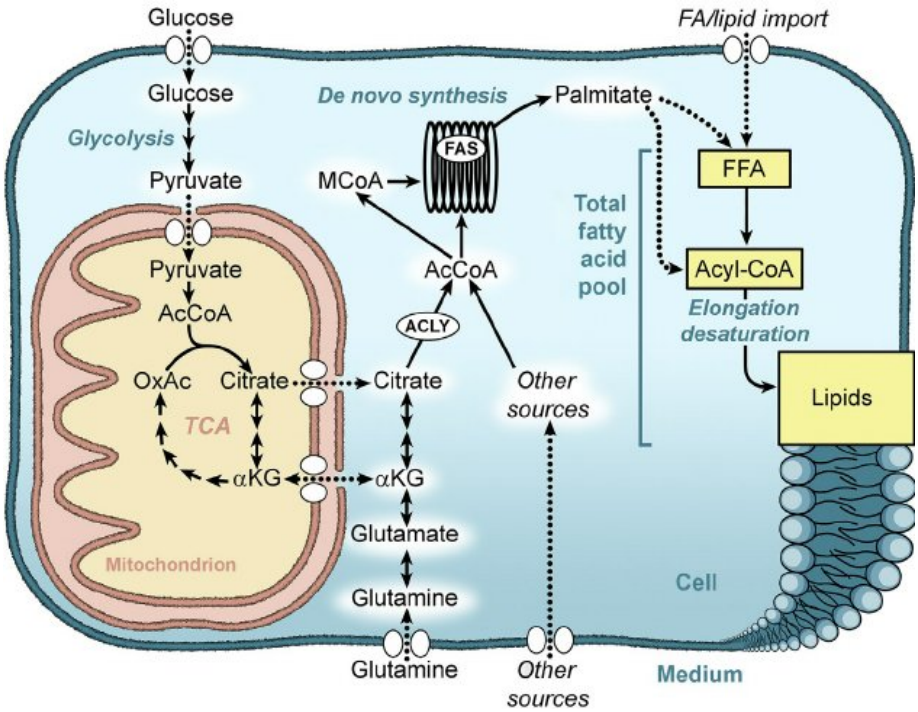


Figure 1.7 Regulation of fatty acid synthesis in a cell. The known pathways associated in fatty acid synthesis involves glucose, glutamine and other substrates. The production of Acetyl-CoA through the TCA cycle in the mitochondrion facilitates the activation of the fatty acid synthase (FAS). The resulting palmitate, a long-chain fatty acid, is elongated and desaturated to produce the free fatty acid (FFA) found in the lipid bilayer of cells. Image derived from Tumanov et al. 2015.

Fatty acids are also important constituents of mammalian cells and play a role as structural elements of the cell membrane, storage of energy, and work as signal molecules. Fatty acid synthase (FASN) is a key metabolic enzyme in fatty acid synthesis. FASN catalyzes the synthesis of long-chain fatty acids from acetyl-CoA and malonyl-CoA (Suburu et al., 2014). When cells

sense a lack of, or large demand for, fatty acids a stress mechanism is activated that can cause negative effects in the cell, halting fat synthesis can accumulate toxic products within the cytoplasm. Indeed, deregulation of fatty acid metabolism has been correlated with disease and cancer (Baenke, Peck, Miess, & Schulze, 2013; Khandekar, Cohen, & Spiegelman, 2011; Zaidi et al., 2014). The cell has controlled feedback loops that aim to promote cell growth and development through transcription factor activation of genes involved in fatty acid biosynthesis. Receptors on the nucleus were discovered to be ligand-activated TF's that mediate nutrient and hormonal responses (McKenna NJ. 2009). These nuclear receptors can sense levels of nutrients and hormones to direct respective transcription of genes, including fatty acids. One subtype of the nuclear receptors includes peroxisome proliferator-activated receptor-α (PPARα), which plays a role in sensing and signalling fatty acid synthesis. These receptors, when activated, transcribe genes for fatty acid catabolism (including FASN), clearing unwanted remains in the cell (Tzeng J., 2015). Fatty acid regulation is required for proper cellular function and is a preventative factor in stress. Figure 1.7 depicts the steps involved in how fatty acids are utilized. Key drivers in this mechanism and maintaining appropriate fatty acid synthesis is a good carbon source, glucose and fat source.

1.9.3 Carbon source sensing, particularly glucose, and stress responses in mammalian cells

Glucose is the main sugar and carbon source that drives cell function, prosperity, and growth and is the main carbon source used during the *in vitro* culturing of CHO cells for biotherapeutic protein production. Thus, a further potential stress on CHO cells during bioprocessing is the availability, allocation and regulation of glucose sugars. Protein production relies heavily on availability of energy requirements to sustain such exogenous protein production that is not a necessity for the cell. Enhancement of protein production from transient and stable CHO cell lines can be achieved by supplementing with glucose feds, which can also affect post-translational modifications of proteins (Iacobini, C., Vitale, M., Pugliese, G., & Menini, S., 2021). Further, importantly, during oxidative folding in the ER requires control of this process and one protein involved in this is Thioredoxin (Txnip). One of the roles of this protein is in regulating the redox state in the cell (Yu F-X, Luo Y., 2009). Upstream of the promoter of the gene for Txnip is a regulatory sequence that was identified as the carbohydrate response element (ChoRE), which cooperatively functions in line with glucose availability. ChoRE, a consensus sequence, is

comprised of two E-boxes inserted with base pair nucleosides, which drives the activity of transcription of downstream genes (described more later in the thesis). During stressful occurrences, such as low or very high glucose, this response element controls the transcription of downstream genes. Glucose is also known to be a precursor in fatty acid synthesis, as Figure 1.7 shows, thus linking multiple potential stresses in the cell.

1.10 Aims of this project

As described above, there is a need to further develop CHO cells to improve their ability to produce difficult to expression biotherapeutic proteins. In order to do this, a better understanding of the limitations upon recombinant protein production from these cells needs to be established. One approach to this is to monitor the stress perception on CHO cells during recombinant protein production in real time (i.e., as they are placed under the stress of making large amounts of exogenous recombinant protein). As outlined above, during bioprocessing and production of such molecules the cells can potentially be exposed to, and sense, aa number of stresses that may limit their ability to produce the target proteins. This project set out to develop and begin investigating reporter gene constructs that contained specific transcription factor binding sites upstream of the promoter sequences to drive their expression that would 'report' on the stress state of the cells during culture and recombinant protein production. By placing stress induced transcription factor binding sequences upstream of the promoters, the reporter expression should respond to the perception of stress in the cell and allow identification of when and where such stresses were perceived in CHO cell. Such an approach has already been described to start looking at monitoring some ER stress in CHO cell during protein production (Roy G, Zhang S., 2017). He we looked to establish multiple fluorescent reporter constructs for ER, lipid and energy stress in CHO cells such that these could be monitored and a more complete analysis and evaluation of the different stresses upon CHO cells during recombinant protein production determined and the timing and connectivity between these. Alternately, this would aid in the development of new bioprocesses and cell engineering strategies to help overcome such stresses and enhance recombinant protein yields and quality from CHO cells.

Chapter 2: MATERIALS AND METHODS

2.1 *ESCHERICHIA COLI BASED METHODS*

2.1.1 Preparation of LB Media

Luria-Bertani (LB) (Bertani, 1951) media was used to promote growth of the *E. coli* strain. To prepare 500 mL of liquid media, a 1 L flask was used to add 5 g of Tryptone, 2.5 g of Yeast, 5 g of NaCl, 7.5 g of Agar, and this was then made up to 500 mL with Milli-Q water. The solution was then stirred to mix and then autoclaved at 121°C for 30 mins before using. Ampicillin (Amp) is sensitive to degradation at high temperatures; therefore, it was added sterilely (using a flame) to the media once it had cooled at a final concentration of 50 µg/mL. The media was then poured into petri dishes (100 x 5 mm) and allowed to set. Plates were then stored in a cold room or refrigerator and used within 1 month. For LB liquid broth, no Agar was added.

2.1.2 Preparation of Calcium Competent Cells

E. coli strain DH5α was streaked onto an agar plate without antibiotic (no Amp) and incubated overnight at 37ºC. From the overnight plate, a single colony was picked and inoculated into a 5 mL tube with LB media. This was incubated overnight at 37 ºC with shaking at 200 rpm. The following morning, 1 mL of the starter culture was added to a 250 mL Erlenmeyer flask containing 50 mL of LB media. The culture density was then monitored at 600 nm and allowed to grow to an A_{600} of 0.4-0.6. The cells were then cooled on ice and centrifuged at 3500 rpm for 15 min at 4ºC. The pellet was resuspended in 10 mL of filter sterilized 100 mM $CaCl_2$ and incubated on ice for 30 mins. The suspension was then again centrifuged for 15 min and the cell pellet then resuspended in 2 mL of prechilled 100 mM $CaCl_2$. The cells were then stored frozen at -80ºC until required for use.

2.1.3 Transformation of Competent E. coli Cells

The transformation of ligation reaction products was performed into calcium competent cells. Aliquots of 50 µL of Ca^{+2} competent cells were put on ice to thaw. Once thawed, the ligation reaction products were mixed with the cells for 30 minutes on ice. Following this, the samples were heat shocked in a water bath at 42°C for 1 minute, and immediately put

on ice. Sterile LB broth (750 µL), at room temperature, was then added to the cells and incubated with shaking in a 37°C temperature for one hour. 100 µL of the transformed products were then plated onto Amp agar plates and incubated overnight (~12-16 hours) at 37°C.

2.1.4 Growth of E. coli on Agar Plates with Amp

If overnight plates from transformation experiments (*Section 2.1.3*) resulted in colonies, it was compared to a control plate. The control plate was prepared similarly to the ligated products, however lacking plasmid DNA. Single colonies of *E.coli* growth were selected from test plates, preferably visibly larger colonies, for culturing overnight and bulking up the DNA (*Section 2.1.5*).

2.1.5 Overnight Culturing of E. coli Colonies in LB

To prepare overnight culturing tubes, 10 mL falcon tubes were used. To these were added sterile LB medium (approximately 5 mL) and 5 µL of the antibiotic Amp to give a concentration of 50 µg/mL. This was all undertaken using an aseptic technique, near a Bunsen burner flame. Subsequently, one colony from a plate was selected and inoculated into the media. The tubes were then incubated with shaking rotation at 37°C and 300 rpm, overnight (12-16 hours).

2.1.6 Miniprep Plasmid DNA Isolation from E. coli

Plasmid DNA was recovered from cultures grown as described in *Section 2.1.5* using a commercially available Qiagen Mini Prep kit following the manufacturer's instructions. Cells were recovered by centrifugation of the cultures at 3000 rpm for 20 minutes at room temperature. The cell pellet was resuspended with P1 solution from the QIAGEN® Mini Prep kit. The DNA was then purified directly using the QIAGEN® kit instruction methods.

2.2 GENERAL MOLECULAR BIOLOGY METHODS

2.2.1 Agarose Gel Electrophoresis

1X TAE (Tris-EDTA-Acetate) buffer was used for the preparation of 1-2% (w/v) agarose gels. Unless otherwise stated, 1.5% gels were used, and run at 80 mV for 30-50 minutes in 1X TAE buffer. DNA was visualized by the addition of 0.005% (v/v) ethidium bromide into the gel before setting.

2.2.2 Purification of DNA from Agarose Gels

The appropriate bands were physically excised from the gel using a sterile razor blade. The DNA was then recovered and purified using the commercially available Wizard® SV Gel and Clean-Up System (Promega Express) following the manufacturer's instructions.

2.2.3 Quantification of Nucleic Acid Concentrations

A Nanodrop UV/Vis Spectrophotometer was used to quantify DNA concentrations during cloning at A_{260} nm and of plasmid DNA concentrations for transfections.

2.3 CLONING

2.3.1 PCR Amplification

Each PCR was altered based on the DNA strands being amplified. PCR reactions were generally set up by the addition of the following reagents into a PCR tube unless otherwise stated:

5X HF Buffer	20 µL
dNTP (10 mM)	2 µL
DNA (1 mg/mL)	2 µL
Forward Primer	4 µL
Reverse Primer	4 µL
H_2O	67 µL
Polymerase	1 µL

The Phusion Polymerase enzyme was the last component of the reaction for PCR to be added, as it activates the reaction. The reaction was programmed to proceed for 35 cycles (subject to change based on reaction details). PCR was carried out using the thermal cycle conditions given below:

98°C	30 seconds	
98°C	**10 seconds**	35 cycles
(T_m - 2°C)	**30 seconds**	
72°C	**(30 sec/kbp)**	
72°C	7 minutes	
4°C	HOLD	

The steps above in bold were variable depending upon the size fragment of DNA being amplified and the melting point (T_m) of the primers. Annealing temperatures of primers were estimated according to Tm as above. PCR products were purified directly using the commercially available Wizard® SV Gel and Clean-Up System (Promega Express). DNA for PCR master mix was always at a concentration of 10 ng/µL, diluted with NF H_2O.

2.3.2 Enzymes and Buffers

Restriction and ligation enzymes were obtained from Thermofisher, Promega, and/or NEB and are as stated in the results section. Purified plasmids or DNA string sequences were digested using appropriate restriction enzymes according to the suppliers' instructions and protocols. Depending on the enzyme supplier being used, their appropriate buffers were used.

2.3.3 Restriction Endonuclease Digests

Reactions were carried out with the appropriate buffer solution for each enzyme in a microfuge tube. Conditions for the reaction included using the appropriate temperature for the particular enzyme, based on the manufacturer's specifications. Variations in the quantity of DNA was dependent upon the concentration of inserts and vectors available.

The volume was large enough to ensure a 10-fold dilution of added enzyme. The volumes of the DNA backbone, insert, and water addition varied depending on the nucleic acid concentrations. The total volume was for each digest reaction was 20 µL. Enzymes were added at the concentration provided by the supplier either as 1.0 µL or 1.5 µL amounts to the insert and backbone mixtures, respectively. Digests were allowed to proceed for 30 min-1 h in a warm water bath at 37°C. The digested product fragments were then analysed and separated on an agarose gel before being visualized under UV light.

2.3.4 Gel Extraction of DNA

Digested products were separated on a 1.0-2.0% agarose gel run with 1X TAE buffer solution until the dye visually was more than halfway down the gel (~30-50 minutes). The settings for the electrophoresis apparatus were consistent at 80 V. The appropriate bands from digests were then excised and purified by agarose gel extraction, using the commercially available Wizard® SV Gel and Clean-Up System (Promega Express).

2.3.5 Ligation Reactions

Double digested DNA sequences were ligated with the acceptor plasmid backbone at a 3:1 ratio, or at varying ratio's dependent upon the DNA concentrations after digestion and recovery of digest products. Ligation mixtures were incubated for 2-3 hours on the benchtop at room temperature or overnight at 4°C. After heat deactivation of the T4 ligase enzyme at 42°C in a water bath, Ca^{+2} competent cells were transformed with the ligation products, and subsequently spread on an Amp agar plate. The plate was incubated overnight to yield colonies for culturing and screening to determine if the correct insert had been successfully ligated into the acceptor plasmid backbone (*Section 2.1.5*).

2.3.6 DNA Sequencing

Final sequencing verification of DNA products (generally plasmids to confirm the correct insert was present) was commercially provided by *GeneWiz* Sanger Sequencing. Samples were submitted after all cloning protocols were completed.

2.4 ANALYTICAL METHODS

2.4.1 *Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)*

Samples for SDS-PAGE analysis were prepared by lysis and Bradford assay concentrations in a non-reducing buffer solution. The samples were boiled to 95°C for 5 mins. 1X Tris-Glycine-SDS buffer was used for running SDS-PAGE by loading 4-5 μg of total protein per well. Gels were run at a constant voltage of 100 mV for 1-2 hours until the dye front reached the bottom of the gel. The gels were prepared as 12% acrylamide gels unless otherwise stated with a 4% stacking gel. A molecular weight marker was run on all gels to estimate the size of protein bands observed.

The 12% SDS-PAGE was composed of the following materials: 8.2 ml H_2O, 10 ml 30% polyacrylamide, 6.3 ml 1.5 M Tris pH 8.8, 250 ul 10% SDS, 250 ul ammonium persulfate, and 10 ul TEMED.

2.4.2 *Western Blotting*

Western blotting was undertaken by transferring proteins from SDS-PAGE gels after separation (as described in *Section 2.4.1*) to nitrocellulose membranes. Buffers for blotting consisted of 1X Transfer Buffer (50mM Tris base, 144g/L Glycine and 20% Methanol) at pH 8.8. Transfer was run at 100V for 30 mins and 150V for 1-1.5 hours. After transfer, membranes were blocked with blocking solution consisting of 5% (w/v) milk powder or BSA in TBST (0.1% Tween 20). Anti-mouse antibody was used as the secondary antibody for incubation 1:5000 in TBST. All primary antibodies used were raised in mouse. The primary antibodies used were eGFP+V5 in a dilution of 8 mL of BSA in 1.6 μL of the primary antibody.

TBS buffer at 10X for 1 liter was derived from the following components: 80g NaCl (150mMx10), 60g Tris (50mMx10), 2g KCl (2mMx10) and MilliQ H_2O.

50X TAE buffer was composed of: 242 g Tris-Base, 57.1ml Acetate, 100 ml 0.5 M Sodium EDTA, and MilliQ H_2O.

2.4.3 *Preparation of Plasmid DNA for Transfection of CHO Cells and Subsequent Transfection Protocol*

For transfection of plasmid DNA into CHO cells, plasmid constructs were bulked up using QIAGEN Mini Prep purification after transformation of successful digests. The purified samples were quantified using a Nanodrop instrument. Based on their respective concentrations, electroporation buffer (EB) was added to each construct to give a concentration of 40 µg of DNA in a 100 µL. This was used to transfect 1 x 10^7 viable CHO cells in 700 µL of EB. Transfection of the plasmids were performed using Electroporation under sterile conditions in a laminar flow hood. The settings for the electroporator are listed below:

Exponential Protocol	
Voltage (V)	300z
Capacitance (µF)	900
Resistance	∞
Cuvette	4 mm

Electroporation device: Gene Pulser X-cell Electroporation system, BIO-RAD. After the electroporation the cell solution was added to CD-CHO media containing 4 mM L-glutamine and cultured at 37ºC in a 5% CO2 environment with shaking at 200 rpm until samples were harvested for analysis.

2.4.4 *Cell Tissue Culture*

CHO-S mammalian cells were counted and split every 3 days for optimal growth and sufficient viability prior to electroporation, and other tissue culture experimentations. Seeding was performed every 3 days with concentrations at 0.2x10^6 cells/mL. Temperatures were kept at 37°C with a CO_2 concentration of 8% and shake speed of 200 rpm. Cell maintenance and transfections used pre-warmed media to 37°C in a water bath.

CD-CHO medium. The media was enhanced with L-Glutamine at 4 mM and anti-clump solution according to the manufacturer's instructions; at a 1:1000 dilution.

Chapter 3: RESULTS

3.1 CONSTRUCTION OF INITIAL PLASMIDS FOR USE IN THE PROJECT

The work in this thesis describes the development of reporter systems in mammalian cell lines that will respond to different cellular stresses and thus report on the state of the cell under different conditions. In order to develop these systems, a suitable plasmid backbone and design was required so that appropriate response elements could be investigated. Therefore, initially a series of plasmids were designed that would allow for insertion of specific features for study. The design and structure of these initial plasmid backbones comprised of: (a) peptide tag sequences that could be used to detect a gene product by western analysis (i.e., tags for which commercially available antibodies are available), (b) a flexible promoter region with flanking restriction sites, which have the potential to be altered, (c) potential transcription factor binding sites to be added (or removed), and (d) fluorescent reporter gene sequences to visualize the end product. The commercially available *pcDNA™3.1/Hygro*$^{(+)}$ plasmid (Figure 3.0) was used as the starting vector due to its ability to give high-level expression in mammalian cell lines and circumstantial features such as the selection marker Hygromycin (used for selection in mammalian cells) and ampicillin resistance (for selection in bacteria systems) genes. The extensive multiple cloning site (MCS) in this plasmid facilitated the construction of constructs containing the elements desired in the new backbones and plasmid constructs (different promoters and regulatory regions, tags and a reporter gene) *pcDNA™3.1/Hygro*$^{(+)}$ already contains a strong CMV enhancer and promotor site (Figure 3.0). However, for the purpose of this study these were also replaced with other sequences, as outlined below, that would respond to different cellular stresses.

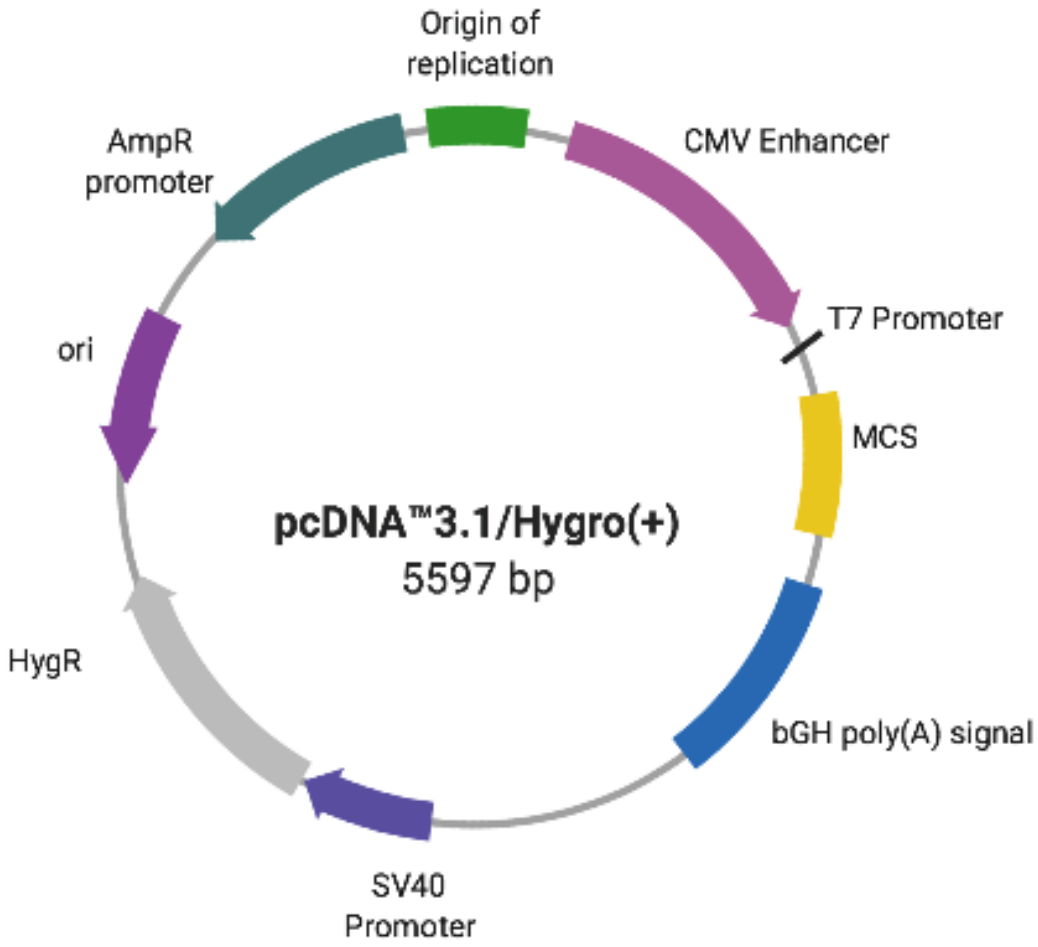


Figure 3.0 The donor plasmid *pcDNA™ 3.1/Hygro*$^{(+)}$ used for the initial steps in cloning. This vector contains all necessary sites for proper cell function and its easy-to-manipulate sites into the vector plasmid. This plasmid contains approximately 5600 base pairs (bp).

3.1.1 Design of Tag sequences

The first stage of the plasmid construction process involved the design and insertion of the tag sequences into the starting plasmid. Due to the differing sequences of each of the four types of peptide tags selected: Flag, His, Strep, and V5, they were cloned independently into the vector plasmid. These tags were selected as they are widely used and validated commercial antibodies that specifically recognise these tags are available. Further, the His and Strep tags can be used for affinity purification purposes if required. The design of the tag sequences incorporated multiple restriction sites, spacers from the plasmid, and the tag itself. Using restriction sites *MluI* and *ApaI*, the sequence of each tag was inserted such that it removed the current CMV promoter and

introduced a new MCS (such as in Figure 3.0). Importantly this section of plasmid DNA also included a T7 promoter region, which was retained to aid in PCR cloning screens. The designed tag sequences and MCS were generated commercially as DNA strings *de novo* using the *GeneArt* gene synthesis program at Thermofisher Scientific. The linear double stranded DNA fragments were then provided ready to be cloned directly into the vector plasmids as strings. Amplification of the strings using PCR provided additional material for cloning experiments; in order to allow this, specific primers were designed and the details of these are provided in Table 3.1.

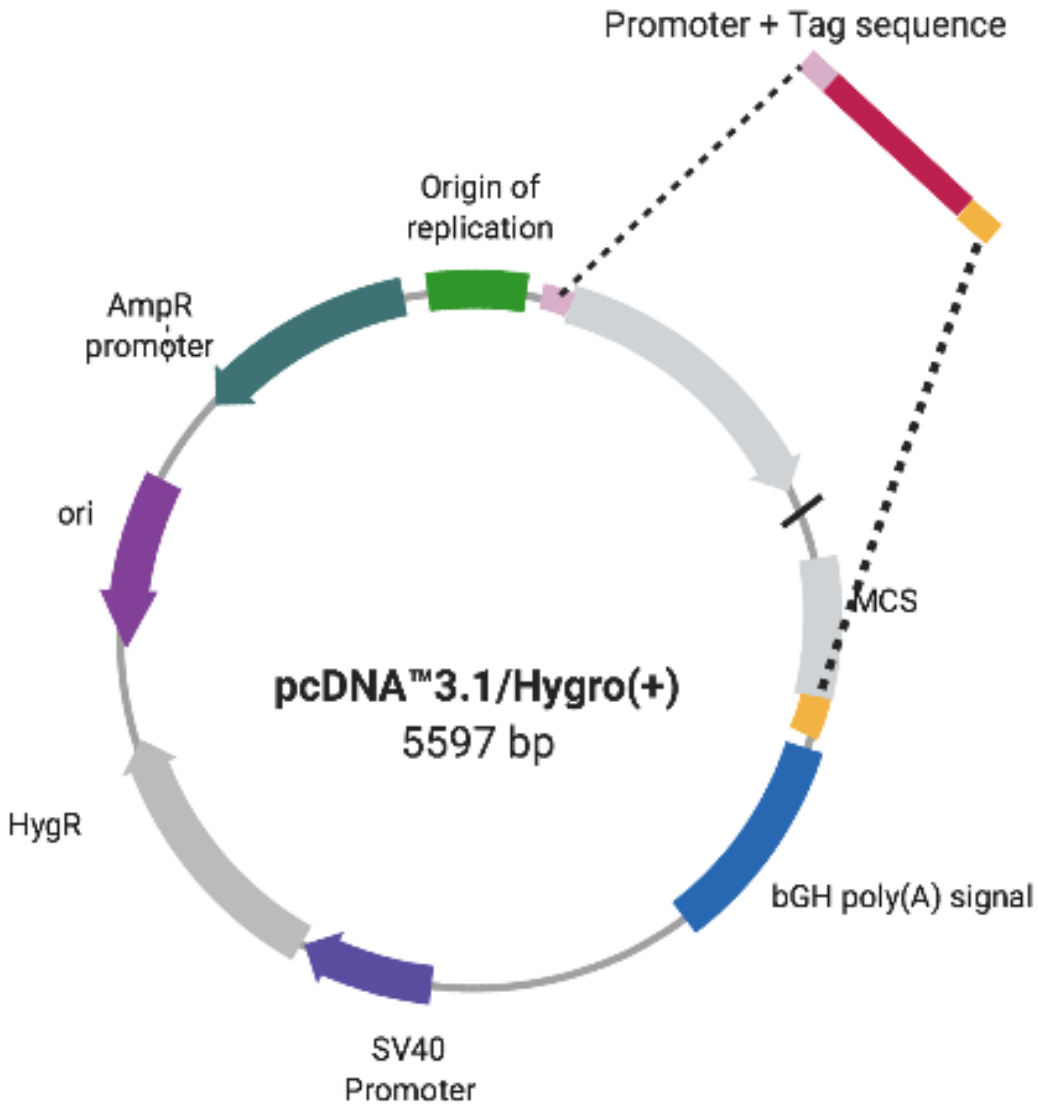


Figure 3.1 Plasmid vector *Hygro 3.1*; being cloned with a tag sequence and a promoter region in the MSC site. Along with this first cloning, the existing promoter regions were ejected.

3.1.2 PCR Amplification of Tag Sequences and Gel Analysis

Using the oligonucleotide forward and reverse primers for each tag (See Table 3.0) the desired sequence was amplified from the commercially provided samples. PCR conditions were as set as described in *Section 2.3.* Dilutions of the string sequences were made according to the manufactures' protocol (IDT) to prepare 100 ng/µL, i.e., adding 100 µL of ddH_2O to the powered sample.

Table 3.0 Oligonucleotides generated for PCR analyses of the Tag sequences

Tag Name	Forward Primer sequence	Reverse primer sequence
V5	5'- TAT GCT AGC GAC TAC AAA GAC GAT GAC GAC AAG -3'	5'- CTT GTC GTC ATC GTC TTT GTA GTC GCT AGC ATA -3'
Flag	5'- TAT GCT AGC GAC TAC AAA GAC GAT GAC GAC AAG -3'	5'- CTT GTC GTC ATC GTC TTT GTA GTC GCT AGC ATA -3'
His	5'- TAT GCT AGC CAT CAT CAC CAT CAC CAT -3'	5'- ATG GTG ATG GTG ATG ATG GCT AGC ATA -3'
Strep	5'- TAT GCT AGC TGG TCC CAC CCC CAG TTC GAG -3'	5'- CTC GAA CTG GGG GTG GGA CCA GCT AGC ATA -3'

The PCR products were then immediately cleaned up using the protocol described in the *Methods* section and the concentrations determined using a Nanodrop UV/Vis Spectrophotometer. The concentrations were as follows:

Table 3.1 Concentrations of post PCR clean-up of amplified Tag sequences as determined by measurement at A_{260} nm

Sample Tag name	Concentration (ng/µL)
V5	271.7
FLAG	254.4
STREP	268.9
HIS	237.4

3.1.3 Double Digestion of Amplified Tag Sequences and Ligation into the pcDNA3.1 Plasmid

The individual tag sequences were digested with the restriction endonucleases *MluI* and *ApaI*, in a double digest. Likewise, the vector plasmid was separately double digested with the same enzymes to generate sticky ends of both the insert fragments and vector for ligation. The tag sequence inserts had a size of ranging between 271-237 bp's and the vector ~5.6 kb. After digestion the fragments were subsequently run on an agarose gel, the appropriate bands cut out of the gel and cleaned up using a PCR cleaning kit. The PCR cleaning kit was used instead of a gel extraction kit as it contains similar reagent mechanisms. The tag and plasmid backbones were ligated together with a 3:1 ratio respectively, transformed into competent *E. coli* DH5α cells and transformant colonies selected for miniprep purification. For plasmids containing V5, Flag, and Strep, three colonies were selected each and purified. For His plasmid constructs, six colonies were selected.

3.1.4 Verification of Successful Ligation of Tag Inserts into the Backbone Vector via Digest Screening

To verify successful ligation of the inserts into the host plasmid, a digest screen was completed. The screening process used the same enzymes as the digestion protocol to release the insert. Figure 3.2 displays the digest screen products on a 1.5 % gel with a 1 kb ladder on the left most lane, for the V5, Flag and Strep colonies screened. Two brightness contrasts are shown so that the vector backbone and released insert can be seen. The boxes on the figures highlights the insert fragment position on the gel. All colonies selected proved positive for successful insertion

of the MCS and tag sequences. However, as seen in Figure 3.3 for the His tag ligations, not all colonies showed the presence of the expected insert although a number did and therefore one of these was selected for sequence verification and further cloning.

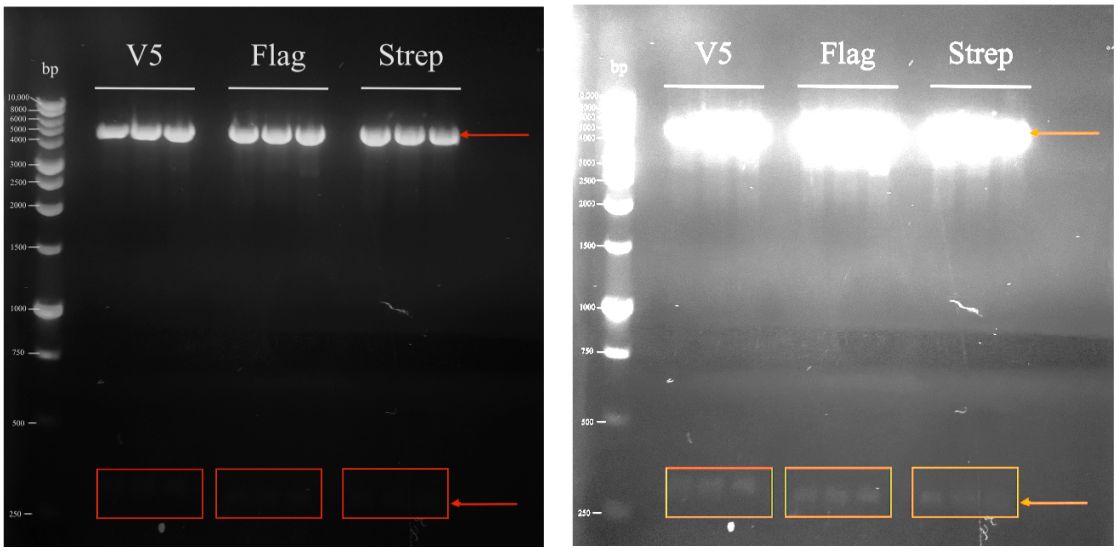


Figure 3.2 Digest screen of V5, Flag, and Strep tag sequence ligations using enzymes *MluI* and *ApaI*. Analysis was undertaken on a 1.5% agarose gel. Successful ligations showed the presence of the plasmid backbone at ~5 kbp (top band, indicated by the arrows) and the insert at the highlighted box (~340 bp). Left and right gels are the same at different brightness to reveal the presence of the lower insert band.

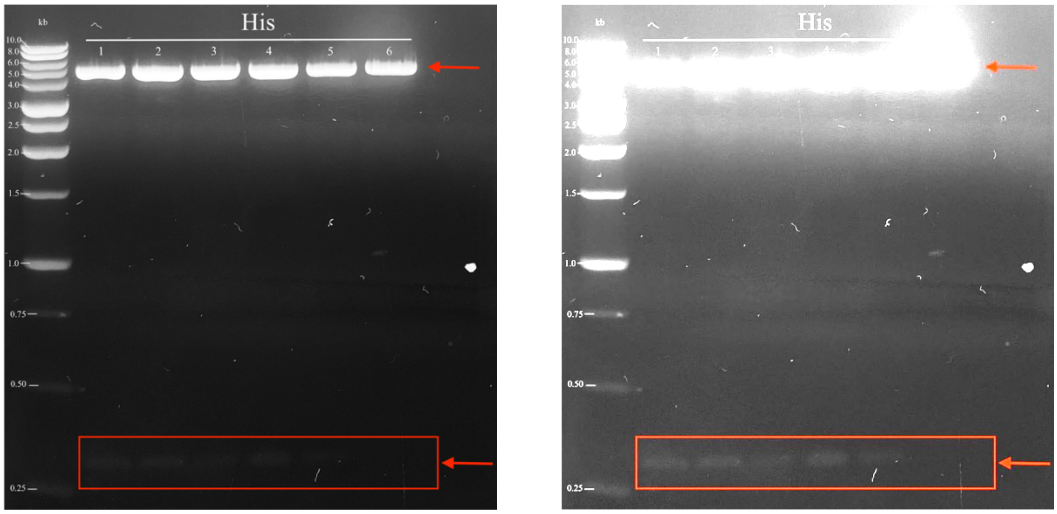


Figure 3.3 Digest screen of the His tag sequence ligation using enzymes *MluI* and *ApaI*. Analysis was undertaken on a 1.5% agarose gel. Successful ligations showed the presence of the plasmid backbone at ~5 kbp (top band, indicated by the arrow) and the insert at the highlighted box (~340 bp). Left and right gels are the same at different brightness to reveal the presence of the lower insert band.

In order to confirm that the inserts amplified by PCR were ligated correctly, the products were also analysed by agarose gel electrophoresis and then again after the double digestion. An example of these analyses for the His tags are shown in Figure's 3.4 and 3.5 below. The His tags were performed separately as the DNA constructs were received by the laboratory later in the week. To compensate for time, His tag cloning was delayed. Hence, the separate gel analyses (Figure 3.4 & 3.5). The gel below confirms amplification of the tag insert by generating a band once digested.

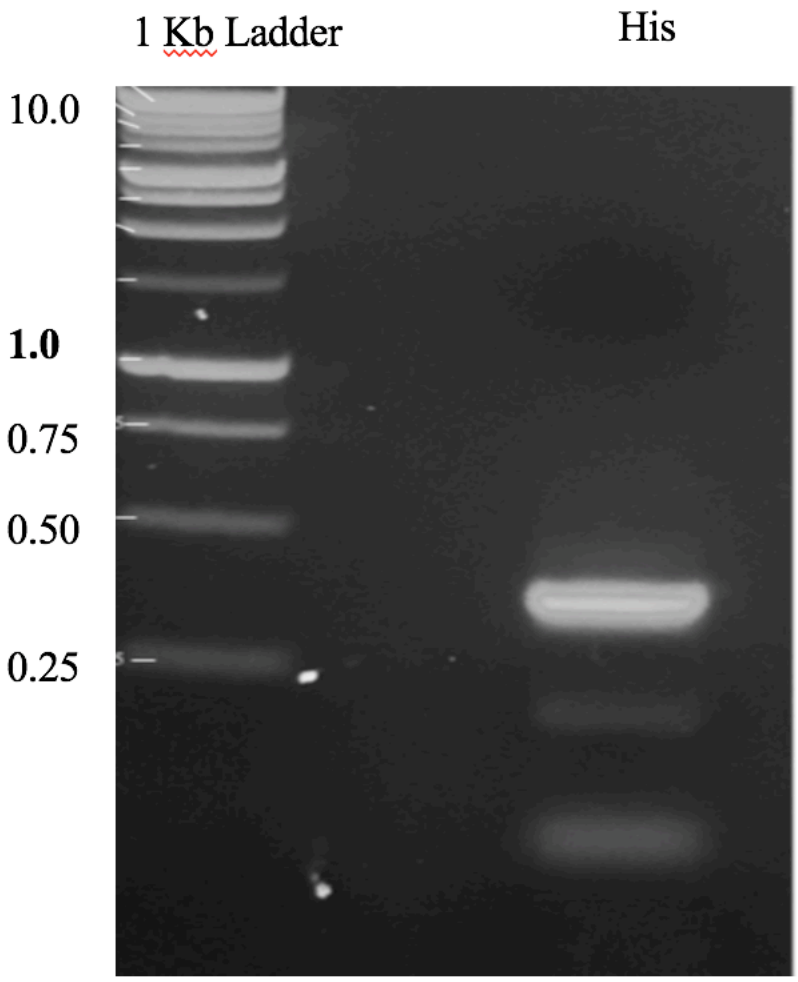


Figure 3.4 His PCR product after amplification from the commercial gene string analysed on a 1.5% agarose gel. The expected band is of approximately 340 bps in size as seen in the gel post-PCR.

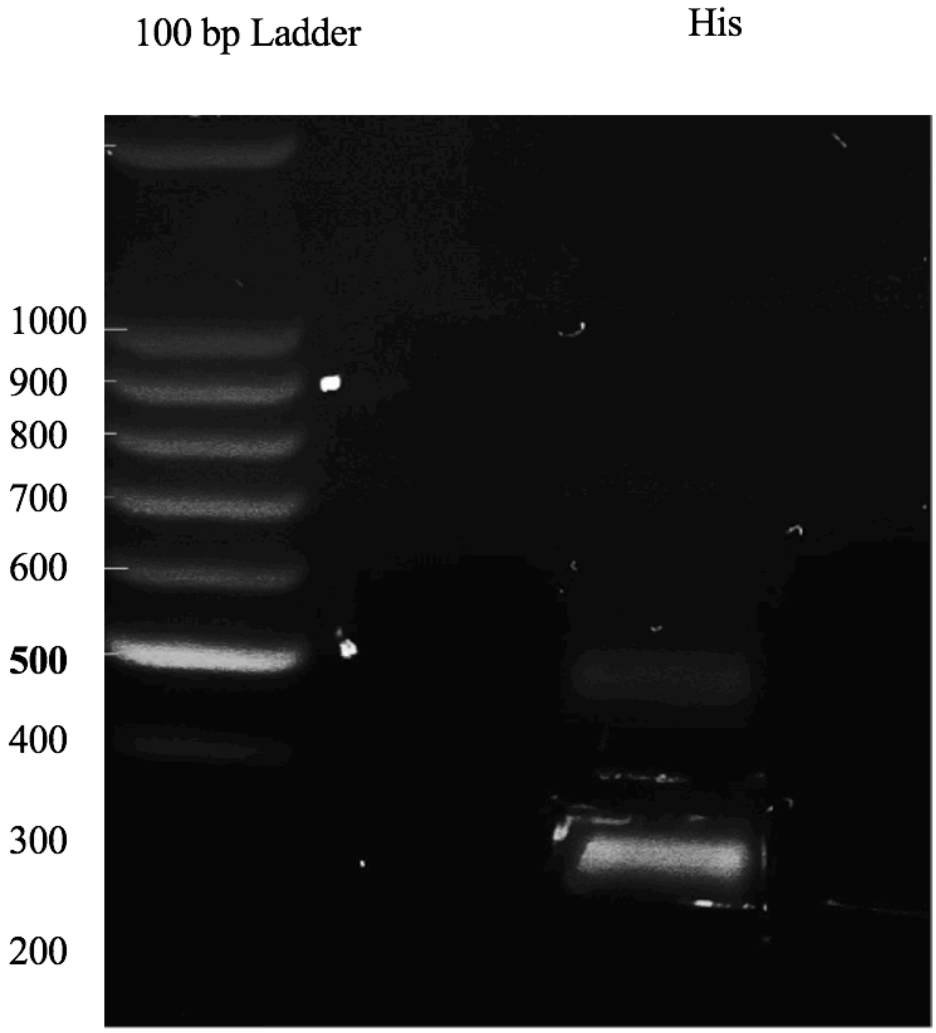


Figure 3.5 His PCR product after amplification and subsequent digestion with the restriction enzymes MluI and ApaI analysed on a 1.5% agarose gel. The expected band is of approximately 340 bps in size as seen in the gel.

3.2 CLONING OF PROMOTER REGIONS

3.2.1 PCR Amplification of Target Promoters

The next step in the design of the constructs was to re-insert promoters of interest, in this case the strong CMV and 'weaker' SV40 virus promoter regions were selected based on their differing transcriptional activity. The CMV core promoter (without enhancers) is a ~580 bp fragment and the SV40 promoter a ~330 bp fragment that drive transcription of downstream target

genes. These sequences were amplified from plasmids already available with the following primers (note restriction site sequences included in these primers are underlined, *BsiWI* and *KasI* for forward and reverse primers respectively):

Table 3.1.1: Oligonucleotides generated for PCR analyses of CMV and SV40 sequences

Name:	Forward Primer sequence	Reverse primer sequence
CMV	5'-TAT CGT ACG GAC ATT GAT TAT TGA CTA GTT ATT AAT AGT AAT CAA TTA CGG-3'	5'-ATA GGC GCC AGC TCT GCT TAT ATA GAC -3'
SV40	5'-TAT CGT ACG GTG TGT CAG TTA GGG TGT -3'	5'-ATA GGC GCC TTT GCA AAA GCC TAG G-3'

The promoter fragments were individually amplified by PCR and then cloned into each tag sequence vector described in *Section 3.1* above.

3.2.2 Analysis of Double Digestion of PCR Amplified Promoter Sequences and Tag-Plasmids for Ligation

The general strategy for cloning was digestion of the CMV and SV40 core promoter fragments using the appropriate restrictions enzymes followed by ligation with the tag containing expression vector prepared by digestion with the same enzymes. The amplified PCR products were therefore initially double digested with *BsiWI* and *KasI*. Along with the promoter regions, the successful tag plasmid constructs were digested with the same enzymes. Analysis of the digestions on a 1.5% agarose gel depicting each fragment and vector size is shown in Figure 3.6. As seen in Figure 3.6, the observed inserts were appropriately positioned on the gel for their expected individual sizes. The fragments were then ligated together with T4 ligase and transformed into Ca^{+2} cells. Colonies were grown and the plasmid DNA was then purified by miniprep purification for analysis to determine if ligations had been successful.

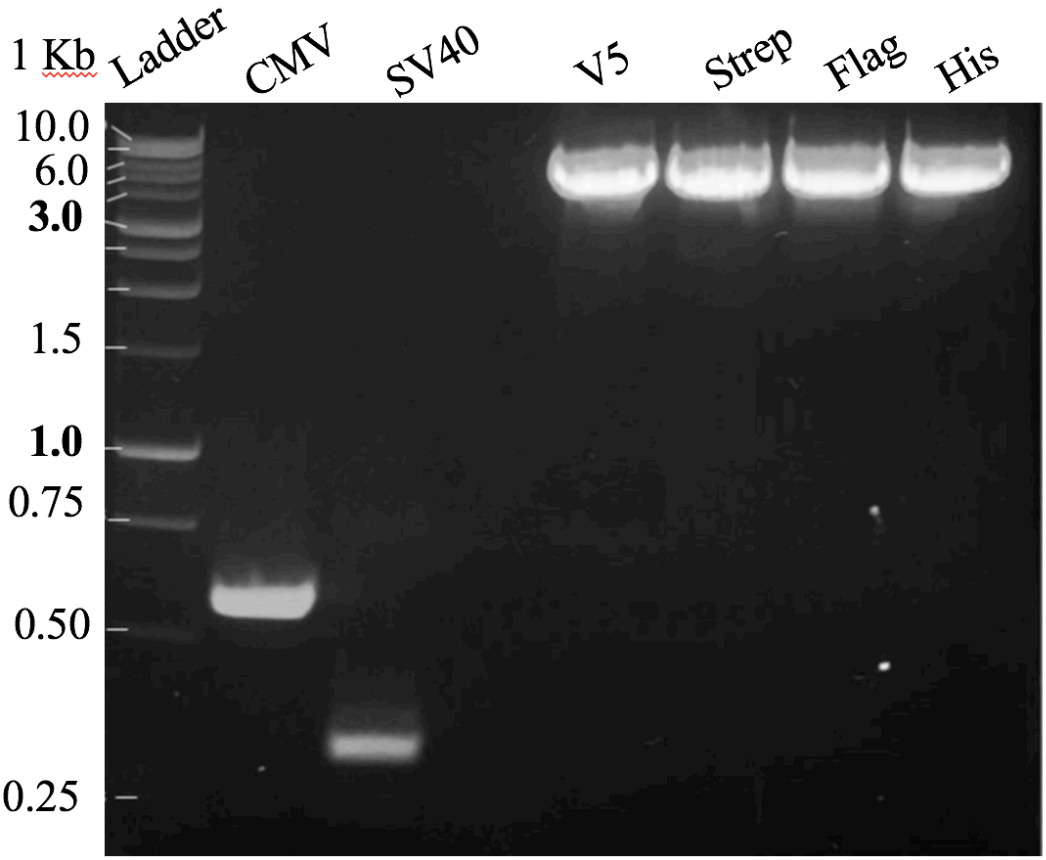


Figure 3.6 Double digestion of tag containing vectors and promoter inserts with *BsiWI* and *KasI*. The core CMV promoter insert has a length of ~580 bps and the SV40 ~330 bps in line with that of the appropriate markers in the 1 kb DNA ladder (Promega) on the left-most lane.

3.2.3 Confirmation of Ligation of Promoter Inserts in Tag Sequence Containing Plasmids

The presence of the promoter inserts after ligation was verified by digestion of the isolated plasmids with the two restriction enzymes *BsiWI* and *KasI*. Results of the agarose gel analysis of the double digested plasmids after ligation and transformation are shown in Figure's 3.7 and 3.8. A single colony from each construct was purified and screened. Whilst the agarose gel showed the appropriate insert band for the Flag, His and Strep constructs, and for the SV40 V5, His and Strep constructs, both the CMV + V5 and SV40 + Flag construct did not release a second band (Figure 3.7) suggesting no insert was present. Additional colonies from these transformants were therefore prepared and analysed for the presence of the insert (Figure 3.8). From these, successful bands

were observed. The red arrows in Figure's 3.7 and 3.8 indicate the successful cloning of the promoter regions into the tag-containing plasmid vectors.

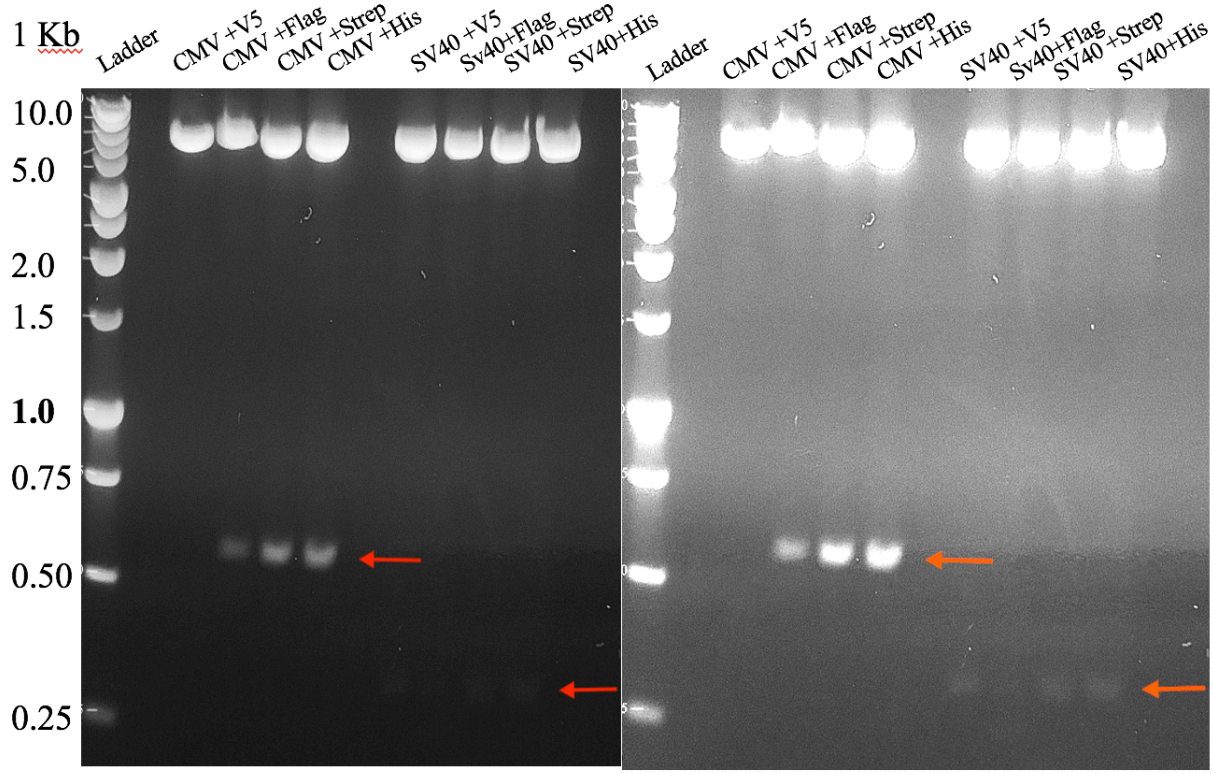


Figure 3.7 Analysis of *BsiWI* and *KasI* restriction enzyme digestions of plasmids isolated from colonies after ligation between CMV and SV40 promoter sequences and tag backbone vectors. The red arrows indicate the positions of the successful ligated plasmids with the promoter region observed (expected sizes for CMV = 580 bp, for SV40 = 330 bp). In the right-hand image, the brightness has been increased to allow visualisation of the SV40 band.

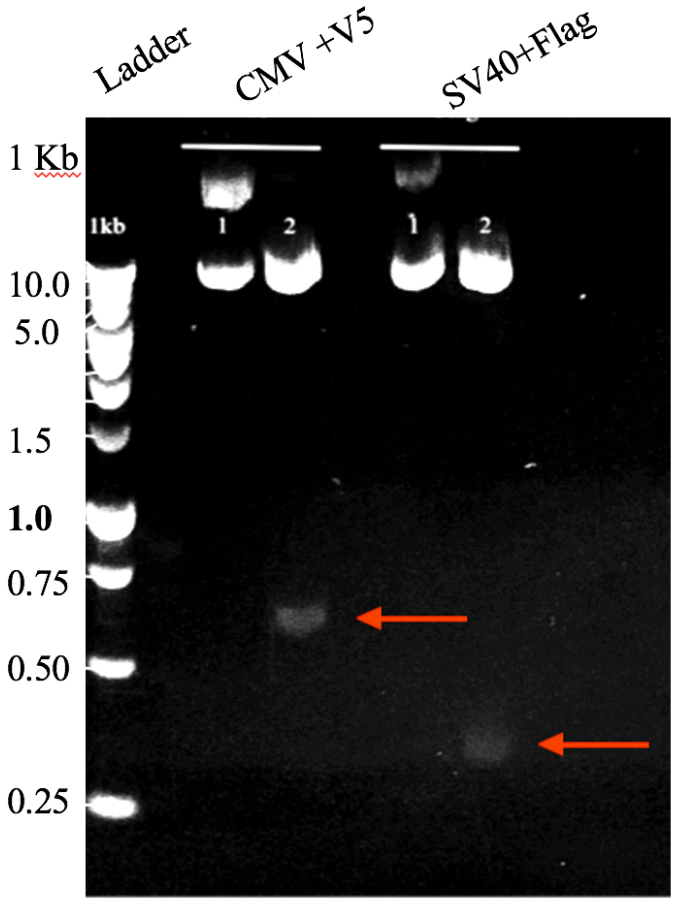


Figure 3.8 Analysis of further *BsiWI* and *KasI* restriction enzyme digestions of plasmids isolated from additional colonies after ligation between CMV and SV40 promoter sequences and tag backbone vectors. The red arrows indicate the positions of the successful ligated plasmids with the promoter region observed (expected sizes for CMV = 580 bp, for SV40 = 330 bp). In the right-hand image, the brightness has been increased to allow visualisation of the SV40 band.

3.3 CLONING OF FLUOROSCENT REPORTER GENES INTO THE PROMOTER-TAG VECTORS

The next step in the construct generation was to clone the reporter gene sequences into the vectors prepared previously containing the tag and promoter sequences.

3.3.1 Generation of Reporter Sequences

DNA sequences for the fluorescent reporter genes *eGFP* and *mCherry* were commercially ordered for synthesis as string fragments. The full nucleic acid sequences of the fluorescent reporter genes were generated using data obtained from the Addgene sequence bank. The first step

in the cloning process was to amplify the fluorescent gene of interest (GOI) to ensure sufficient material was available for cloning experiments. For this, PCR oligonucleotides were designed based on their DNA sequences and additional restriction enzymes were included, which gave an opportunity for insertion into the appropriate vector plasmids. The enzymes used were *HindII*I, for the forward primer, and *XbaI*, for the reverse primer, both underlined below:

Table 3.1.2 Oligonucleotides generated for PCR analyses of eGFP and mCherry sequences

Name:	Forward Primer sequence	Reverse primer sequence
eGFP Primer	5'- TAT AAG CTT ATG GTG AGC AAG GGC GAG - 3'	5'- ATA TCT AGA CTT GTA CAG CTC GTC CAT GC - 3'
mCherry Primer	5'- TAT AAG CTT ATG GTG AGC AAG GGC GAG - 3'	5'- ATA TCT AGA CTT GTA CAG CTC GTC CAT GCC - 3'

For subcloning and next steps, both fluorescent markers were positioned in the plasmid 5' (upstream) to the tag sequence. This was made easier by placing a no stop codon at the 3' end of the fluorescent gene sequence which was in frame with the appropriate tags so that these would be read through during translation to generate a fluorescent protein with an appropriate tag. The amplified eGFP sequence should yield a DNA product of 711 bp and the mCherry of 715 bp.

3.3.2 Ligation of Fluorescent Reporters into the Promoter and Tag Sequence Containing Vector Plasmids

In this phase of work the genes for the fluorescent markers eGFP and mCherry were inserted into the previously generated plasmid vectors containing the CMV and SV40 promoters and different tag sequences. As described previously, cloning proceeded with a double restriction digest of each vector plasmid and the amplified PCR fluorescent gene products which were then analysed by agarose gel electrophoresis (Figure 3.9). The inserts were digested with *HindIII* and *XbaI*, while the backbone vector plasmids with *HindIII* and *NheI*. The restriction enzymes *NheI* and *XbaI* produce compatible overhangs when digested. Since these sites are compatible cohesive ends, they generate neither *XbaI* nor *NheI* sites after ligation, destroying both these sites. After digestion fragments of the expected corresponding size as determined by agarose gel analysis were

observed (Figure 3.9 for CMV constructs and the individual tags). The SV40 promoter and tag construct digestions are not shown but were again as expected. Fluorescent marker digests should reveal a band at ~710 bp, and vector plasmids ~5.5 kb, as depicted in Figure 3.9.

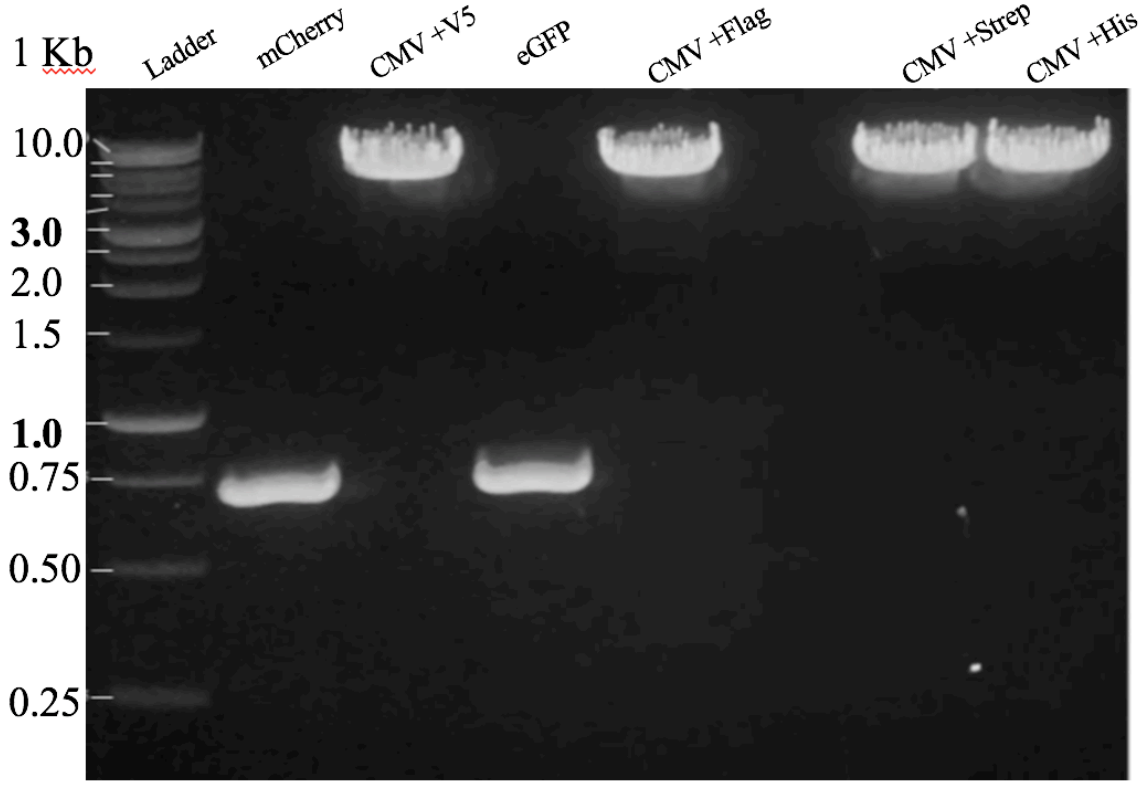


Figure 3.9 Restriction double digestion of promoter and tag containing vector plasmids and gene inserts for mCherry and eGFP. The insert fragments showed the expected bands at approximately 710 and 715 bps; the backbone plasmids showed bands at ~5.5 kb as expected. A 1.5% agarose gel with a 1 kb DNA ladder is shown.

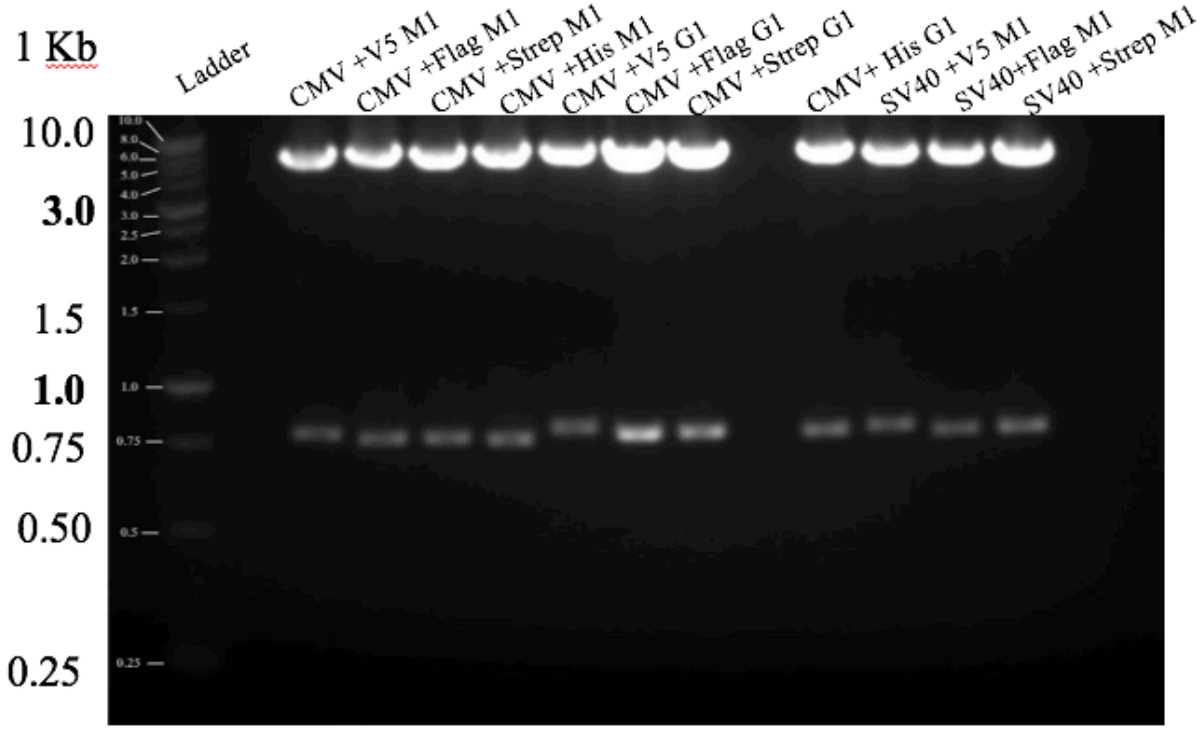
3.3.3 Generation of Stop Constructs which would not Express the C-Terminal Tags

Along with the fluorescent genes being cloned in frame with the C-terminal tags, additional constructs were generated without the tag sequences in case these impacted expression of the fluorescent tags. A simple stop codon as placed within the plasmid before the tag sequence to halt transcription to include the tag. These constructs are depicted as 'Stop' or (-) constructs, such as in Figure 3.10 B. These were generated by using primers to amplify the mCherry and eGFP genes as previously described but, where a stop codon was included.

3.3.4 Restriction Digestion Verification of Correct Fluorescent Gene Insert Ligation Success

Analysis of colonies to identify successful transformants that contained the fluorescent gene markers was undertaken using double digestion screens of the plasmid DNA from minipreps (Figure 3.10). Multiple colonies were selected for insert screening, depicted as 'M' for mCherry and 'G' for eGFP. For screening, DNA from minipreps of each colony was digested with *HindIII* and *ApaI*. *ApaI,* a restriction enzyme site within the plasmid was used instead of both *XbaI* and *NheI* as these restriction sites were destroyed upon ligation. Agarose gel analysis of the resulting digests revealed those colonies that gave rise to he expected insert band at ~780 bp for many of the colonies selected (Figure 3.10 A & B). Several of the selected colonies for different constructs were not successful or lacked the presence of the ligated products. For the unsuccessful samples, additional colonies were screened and identified for each construct as determined by the presence of the desired band (Figure 3.11).

A.



B.

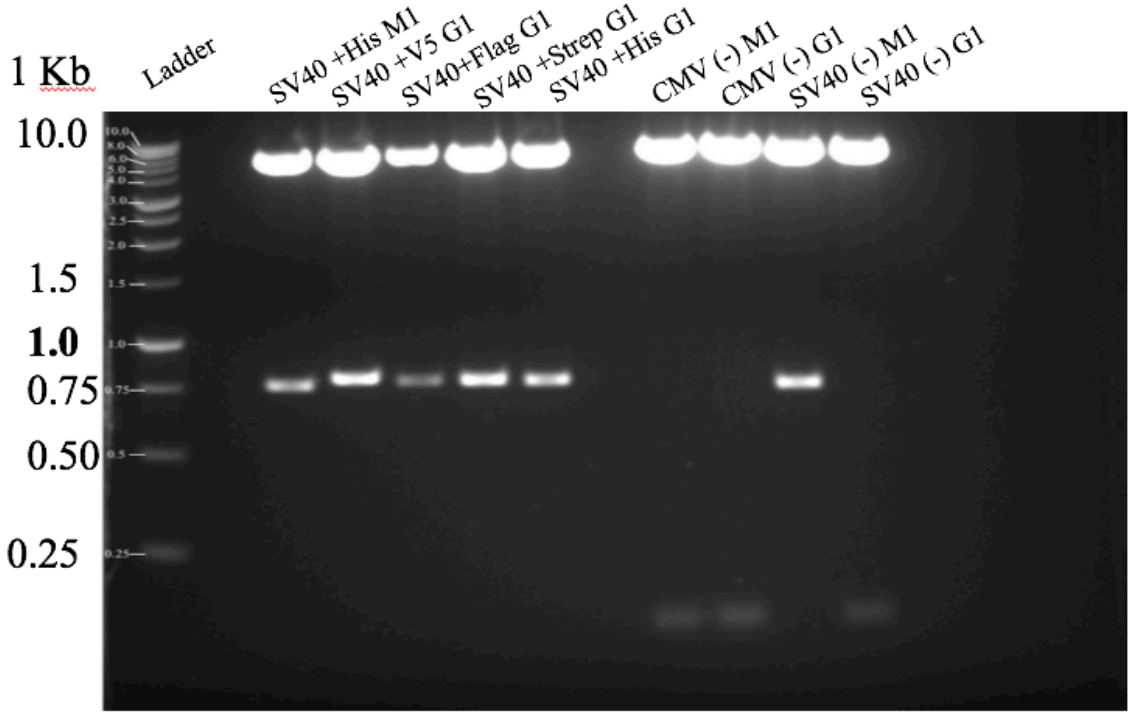


Figure 3.10 Analysis of transformant colonies from eGFP (G) and mCherry (M) ligations.

Constructs were screened with *HindIII* and *ApaI*, successful ligations containing a band at ~780 bps.

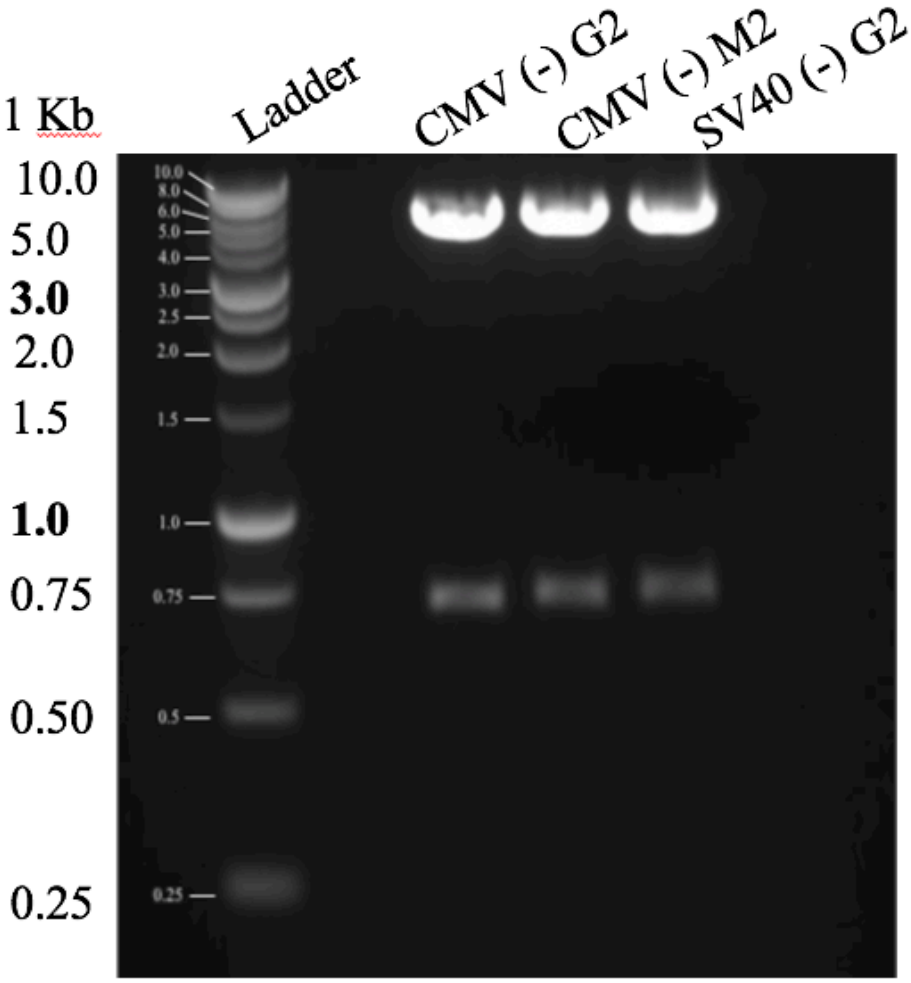


Figure 3.11 Repeated analysis of initially unsuccessfully screened transformant colonies from eGFP (G) and mCherry (M) ligations. Constructs were screened with *HindIII* and *ApaI*, successful ligations containing a band at ~780 bp. The digest screen used HindIII and ApaI for plasmids CMV + eGFP (2nd selected colony), CMV + mCherry (2), and SV40 + eGFP (2).

3.3.5 Sequencing to Confirm Presence of Correct Insert Sequences

Plasmid sequencing was performed on constructs identified as having the correct inserts by restriction enzyme mapping commercially by *GeneWiz* using Sanger Sequencing. All constructs were sent to confirm the nucleic acid sequences and whether cloning was successful, as well as assess whether introduction of any mutations or addition/deletion of nucleotides was seen. The forward primer used for sequencing was developed to provide a sequence from the MCS. A

standard bGHR reverse primer was used for this purpose (Table 3.1.3). This is a universal primer used in Sanger sequencing which was present in the plasmids used in this study, therefore making it appropriate for use.

Table 3.1.3 Oligonucleotides generated for PCR analyses of the bGHR sequence provided by Sanger sequencing

Name:	Reverse Primer sequence
bGHR Primer	5' – TAGTGCGCGAGCAAAATTTA – 3'

The constructed primer was designed 100 nucleic bases away from the start site beginning with the *MluI* restriction site in the plasmid vectors.

3.4 TRANSFECTION OF FLUORESCENT GENE CONTAINING PLASMIDS INTO MAMMALIAL CELLS TO CONFIRM EXPRESSION AT THE PROTEIN LEVEL

3.4.1 Transfections of eGFP Containing Plasmid Constructs

All constructs containing the eGFP fluorescent marker were amplified in *E. coli* and purified for transfection purposes. In total, 12 eGFP constructs were used comprising of the tags, and promoter sequences (Table 3.2). For each construct, the DNA was electroporated into CHO-S mammalian cells as described in the *Methods* section. The cells were subsequently harvested 48 hours post-transfection for determination of fluorescence and protein expression. Table 3.2 describes the amount of fluorescence visually observed under a fluorescent microscope. The intensities observed are recorded as + (low intensity) to +++ (high intensity); (-) indicates no observable fluorescence. The V5 tag constructs for both CMV and SV40 promoter plasmids had the lowest fluorescence intensities from all the tag sequences. The reason for this is unknown. Plasmids containing only the promoters were used as controls and showed no fluorescence as expected. "CMV (-)" represents plasmids with a promoter and eGFP marker gene but lacking any tag sequences. This construct was used as a positive control.

Table 3.2: Relative eGFP fluorescence intensities from constructs transfected into CHO-S cells based on visual fluorescent microscope observation. Details are provided in the text.

Construct	Intensity	Construct	Intensity
V5 + CMV	+	V5 + SV40	+
Flag + CMV	+++	Flag + SV40	+++
Strep + CMV	+++	Strep + SV40	+++
His+ CMV	+++	His + SV40	+++
CMV (-)	+++	SV40 (-)	+++
CMV only	-	SV40 only	-

3.4.2 Cell Lysis and Protein Recovery for Western Blot Analysis

Cell pellets were also collected at the 48-hour time point post-transfection for protein expression analysis by western blot. Fluorescence will reveal that functional protein is being expressed in the cells however, some protein may be expressed that is deemed non-functional (for example it may not be correctly folded and therefore may not result in fluorescence). Thus, western blot would allow investigation of total protein expressed and any differences between samples as opposed to comparing fluorescence intensity from functionally active protein. Lysis of the cells was performed with RIPA lysis buffer. A Bradford assay was performed in non-reducing buffer to determine the total protein concentrations in the cell lysate. A calibration curve was produced using known amounts of BSA protein to determine the protein concentrations in the cell lysates (Figure 3.12). A linear relationship between BSA concentration and absorbance at 595 nm was observed (Figure 3.12). The slope of the best fit linear line was used to calculate the amount of protein in the cell lysates from their respective absorbance readings (Table 3.3). Table 3.3 reports the calculated cell lysate sample concentrations using the Bradford assay and the standard curve in Figure 3.12. This quantification was used to calculate the concentration of the protein of interest.

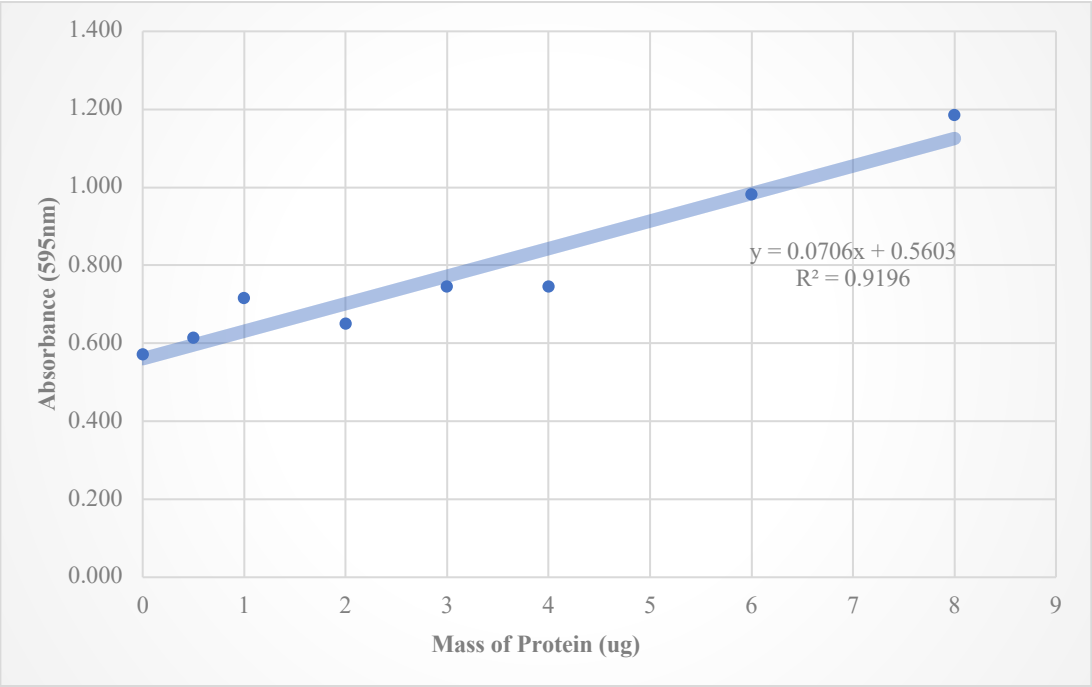


Figure 3.12 Bradford assay calibration curve. A standard curve of absorbance (595 nm) vs concentration (µg) of known BSA concentration standards was used to determine the protein concentrations of the cell lysates from cells transfected with the eGFP plasmid constructs.

Table 3.3 Calculated protein concentrations in cell lysates of samples 48 h post-transfection with eGFP constructs. Protein concentrations were determined using the calibration curve in Figure 3.12 and the respective absorbance readings of the individual lysate samples.

Sample	Average Absorbance (595 nm)	Calculated Sample Concentration (μg/mL)
CMV_V5	0.718	2228.22
CMV_FLAG	0.692	1860.05
CMV_STREP	0.656	1359.72
CMV_HIS	0.757	2789.91
CMV(-)	0.674	1609.89
CMV ONLY	0.706	2063.02
SV40_V5	1.144	8265.24
SV40_FLAG	0.867	4347.55
SV40_STREP	0.717	2218.78
SV40_HIS	0.946	5456.77
SV40(-)	0.879	4517.47
SV40 ONLY	0.785	3176.96

3.4.3 Western Blot Analysis of Protein Expression

The cell lysates were then prepared for western blot analysis. 12% SDS-PAGE gels were used to separate the proteins present, transferred onto a nitrocellulose membrane, the membrane blocked and then probed with primary and secondary antibodies followed by visualization through chemiluminescence. There was weak evidence of protein expression (Appendix Figure's 7.2 and 7.3), indicated by low intensity of bands present on the blots. The reason for this is unclear as strong fluorescence was observed under the microscope (data not included). The western blots were attempted three times, however, no suitable expression data with appropriate bands present was obtained. When editing the blots with altered illumination/contrasts, bands of the expected sizes were visually detectable. After 15 mins of exposure, some indication of protein expression can be seen at approximately 27 kDa as expected for eGFP. The top portion of the gel was probed with an anti-β-actin antibody as a control. β-actin was used as a loading control, to ensure the loading of the protein sample was the same across the gel. Since the β -actin was not visible, the data collected could not be practically used and suggests the amount of protein loaded for analysis was too low for reliable analysis and detection of bands.

3.5 GENERATION OF DESTABLISED GFP (DGFP) CONSTRUCTS

In order for the designed eGFP transcriptional reporter systems to exhibit more rapid responses to changes in transcription, a destabilized GFP (dGFP) that has a much shorter half-life of 2 hours rather than eGFP (a half-life of ~26 hours) was developed (Cornish & Smith, 1999). The dGFP was created using fusion of degradation region to the C-terminal end of the eGFP (Li X, et al., 1998). This mutation gives it the unstable form, which aids in the transient models of CHO cells.

3.5.1 dGFP Cloning

Destabilized GFP (dGFP) replaced the eGFP in the previous generated constructs due to its quicker turnover rate in mammalian cells. The dGFP DNA was obtained from Addgene (pcDNA3.3_d2eGFP plasmid) as a bacterial stab, colonies of *E.coli* within an agar tube. Cloning of the sequence was initiated by culturing the stab and purifying the resulting DNA from the bacteria. PCR primers were then designed as followed, and used to amplify the dGFP DNA sequence:

Table 3.2.1 Oligonucleotides generated for PCR analyses of the dGFP sequence

Name:	**Forward Primer sequence**	**Reverse primer sequence**
dGFP Primer	5'- TAT CTT AAG ATG GTG AGC AAG GGC GAG G – 3'	5'- ATA GCT AGC CAC ATT GAT CCT AGC AGA AGC – 3'

3.5.2 Subcloning of the dGFP Sequence into Appropriate Plasmid Vectors

The dGFP sequence contains 846 nucleic acid base pairs (bps) and was designed upon amplification by PCR to have restriction site overhangs that would allow direct subcloning into the eGFP fluorescent marker position on the respective plasmids. Plasmids containing the promoters and tags without existing fluorescent markers were used as the restriction sites *XbaI* and *NheI* were still present. Figure 3.13 shows the plasmid vector segment with the available restriction sites.

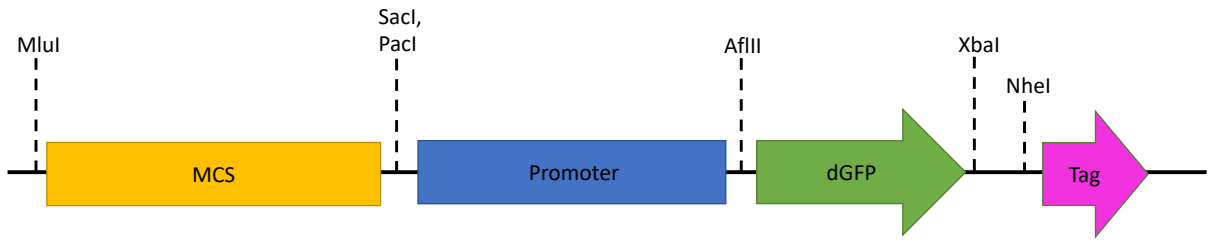


Figure 3.13 Schematic of plasmid layout with key restriction sites present. Between critical points of the DNA strand were placed restriction sites the user can utilise to insert the desired segments of DNA. The above sites are highlighted to visualize how the cloning was planned and performed in this study.

Digestion was performed with *AflII* and *XbaI* restriction enzymes for the plasmid vectors, and *AflII* with *NheI* for the dGFP inserts. Three differently tagged vectors were used to insert the sequence: a CMV containing plasmid, SV40 and a randomly selected tagged plasmid. During the insertion of the fluorescent segment, the tag itself would be removed by addition of a STOP codon to additionally create a *tag-less* plasmid. After ligation, transformation and growth of colonies, a digest screen using *AflII* and *ApaI* was performed to release the insert and assess whether it was present, noting again here the *XbaI/NheI* sites were destroyed upon ligation. The expected band release of ~850 bp was observed in a number of the colonies screened (Figure 3.19) confirming the correct insert was likely present. Subcloning of the promotional regions and dGFP result in a slightly larger band than simply the dGFP sequence only, see Figure 3.15. Those selected colonies that were successful in the digestion, were then used to generate plasmid DNA for transfection.

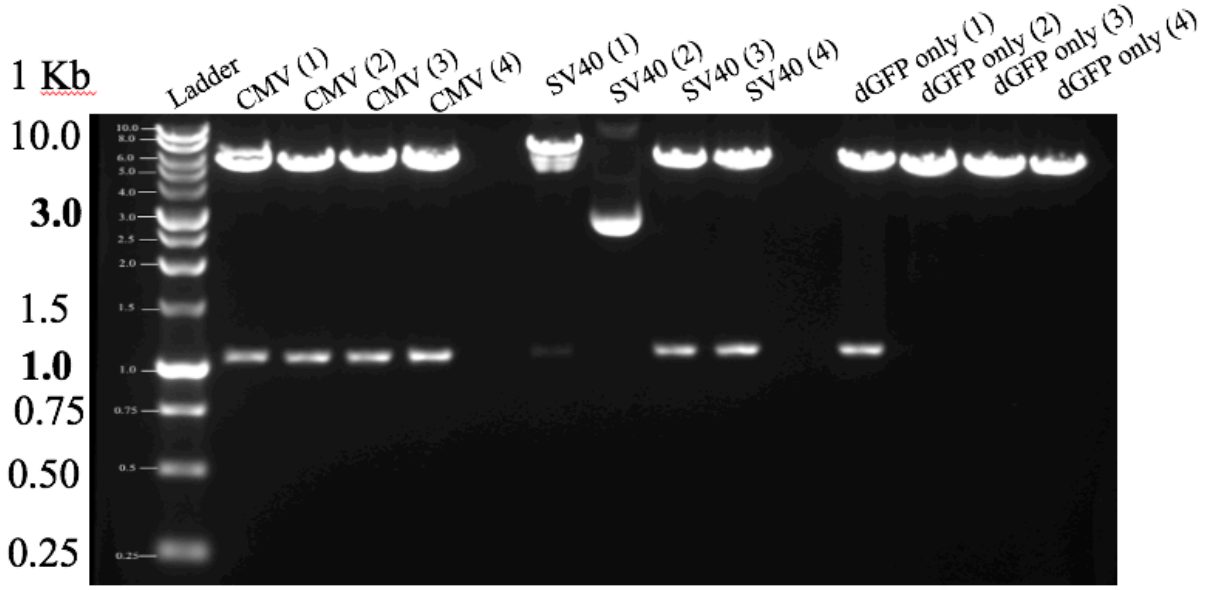


Figure 3.15 Verification of successful dGFP sub-cloning into target vectors using *AflII* and *XbaI* or *NheI* on a 1.0% agarose gel as described in the text. The band at approximately 1.1 kb indicates the presence of the required insert in all CMV + dGFP colonies screened, in several SV40 + dGFP colonies and in one dGFP only colony.

3.5.3 Transfection of dGFP Constructs into CHO-S Cells and Analysis of dGFP Expression by Flow Cytometry

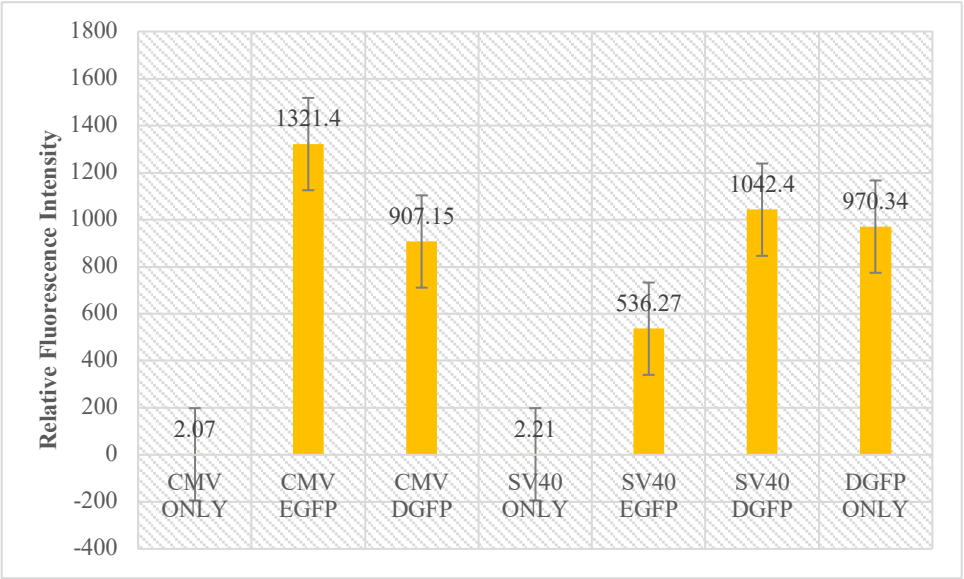
The *arithmetic mean* is the standard mean, number of events in each fluorescent event is divided by the total number of events. As fluorescence intensity increases logarithmically, the arithmetic mean becomes an inadequate measurement to generalize a set of events. The *geometric mean* is used to compensate for the log-normal behaviour of the flow cytometry data.

Transfections of the generated dGFP constructs (20 µg of DNA) were performed by electroporation into CHO-S cells and subsequent harvesting of transfected cell pools at 24 and 72 hours after transfection. Cell counts were also taken at the harvest times. The fluorescence of the cells at the harvest timepoints was then analyzed by flow cytometry. The mean and geometric mean fluorescence of the cells harvested were determined at the two time points from the analysis of 10 000 events (Figures 3.16 & 3.17). Constructs CMV only and SV40 only were used as negative controls.

At the 24-hour time point, the results showed fluorescence in cells transfected with eGFP constructs for the CMV promoter and less so with SV40 as would be expected from the relative

strength of these promoters (CMV is a stronger promoter than SV40). Fluorescence in cells transfected with dGFP constructs changed over time from constructs with both promoters at the 24-hour time point (Figure 3.16A), and more prevalently at the 72-hour time point. The geometric fluorescent mean (Figure 3.16 B) showed similar patterns to the mean values. At the 72-hour time point, the results indicate a decrease in fluorescence for dGFP expression for both CMV and SV40 promoter constructs. This presumably is due to the turnover rate being much faster in destabilized GFP, meaning that it does not accumulate to the same extent as compared to the normal GFP and thus when production is reduced or stopped, the GFP present decreases much more rapidly in dGFP samples compared to standard GFP. Therefore, less fluorescence is observed at later time points with regards to the dGFP constructs. However, fluorescence from the eGFP constructs also decreased by almost 2-fold, whilst the SV40 dGFP fluorescence was decreased by 10-fold at the 72-h time point compared to the 24 h time point. When cells with different fluorescent intensities were investigated by setting fluorescent thresholds termed M1 (lowest fluorescence), M2, and M3 (highest fluorescing cells) (Figure 3.18) the percentage of cells in the highest fluorescence group (M3) decreased over time for all transfections and the number in the lower threshold increased (M1).

A.



B.

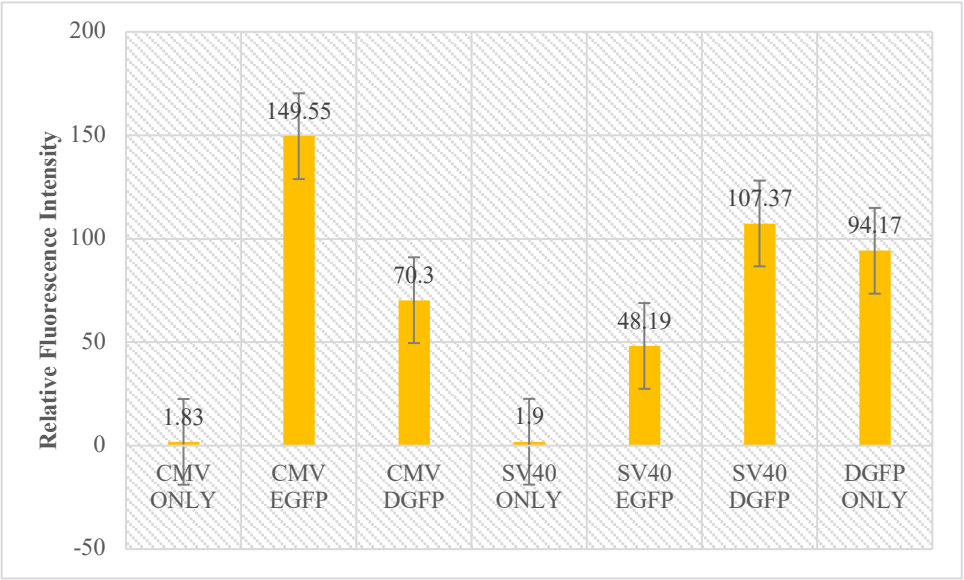
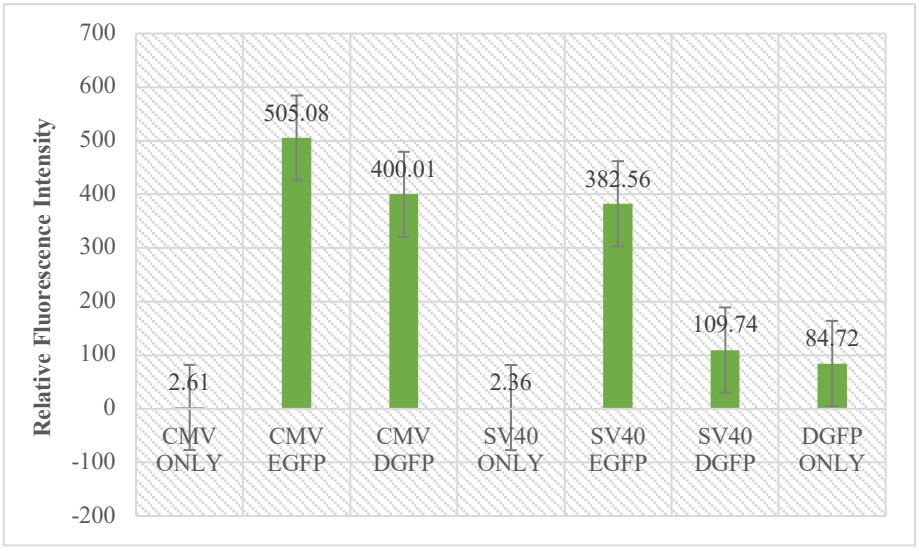


Figure 3.16 Mean (A) and Geometric Mean (B) fluorescence of 10 000 events recorded in CHO-S cells 24 hours after transfection. The relative fluorescence was recorded with the indicated constructs using a flow cytometer. Error bars of standard error are included in the triplicate data collected.

A.



B.

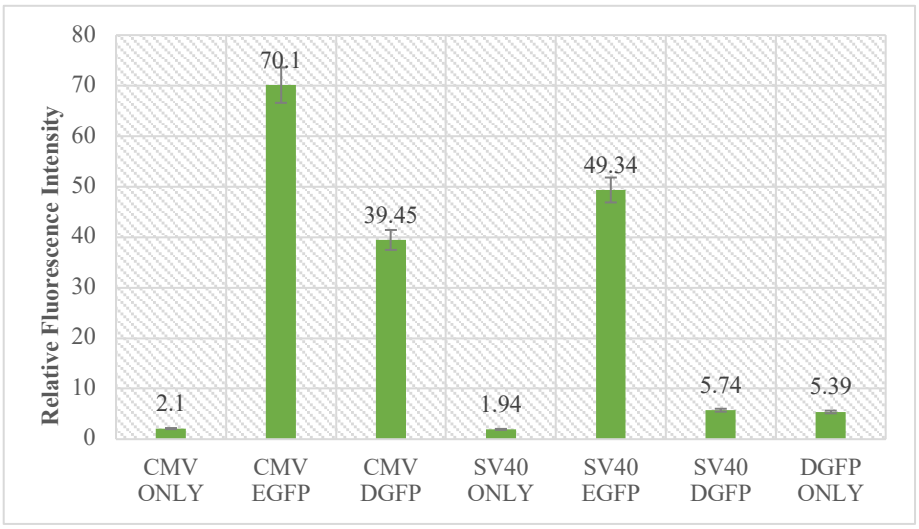
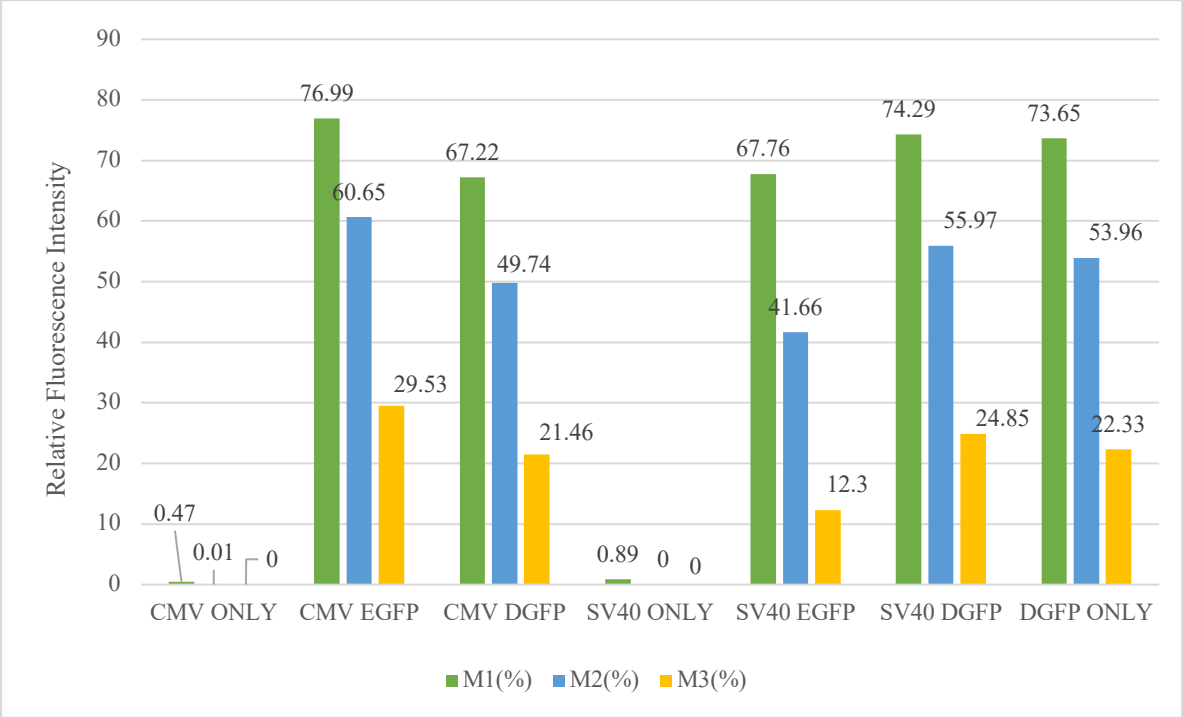


Figure 3.17 The relative Mean (A) and Geometric Mean (B) fluorescence of 10 000 cells in CHO-S cells 72 hours after transfection with the indicated constructs. Fluorescence values are more differentiated between constructs at 72 hours than observed at 24 hours post-transfection. Error bars of standard error are included in the triplicate data collected.

A.



B.

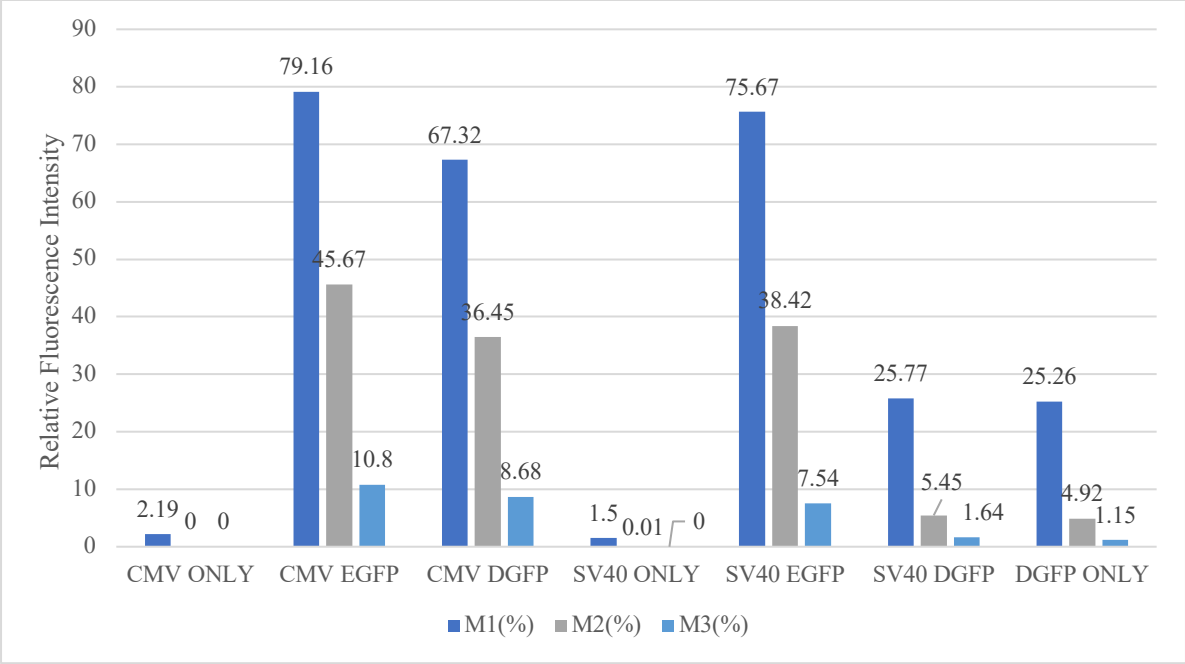


Figure 3.18 The relative fluoresence to the mean values of the constructs cloned. The percentage of cells where fluorescence exceeds a pre-determined threshold (M1, (lowest intensity),

M2 and M3 (highest intensity)) in 10 000 CHO-S cells 24 hours (A) and 72 hours (B) after transfection with the indicated constructs.

3.5.4 Sequencing of the dGFP Constructs and Further Cloning Attempts

These dGFP plasmid constructs were sent for commercial sequencing at *GeneWiz* in order to confirm the correct sequences. The sequencing obtained was correct however it was found that some dGFP-containing constructs held eGFP rather than the desired dGFP GOI. The dGFP gene sequence was cloned directly into either the SV40 or CMV containing plasmids. The differences of these two plasmids were described in previous sections. The expression levels between the promoters can be displayed in Supplementary Figure 7.4. Digestions were undertaken with either *BsiWI/KasI* or *AflII/XbaI* and *NheI*. Plasmids with the CMV and SV40 promoters were used as the "acceptor" vector for the dGFP. Following ligation and transformation, a PCR screen of colonies was utilized to identify colonies containing the dGFP insert using primers to flanking regions of the dGFP in the vector. A colony without an insert would give a PCR product of approximately 250 bp. As depicted in Figure 3.19, only one colony revealed positive results with a band of the expected size of approximately 850 bp with dGFP only whilst the remainder had no dGFP present.

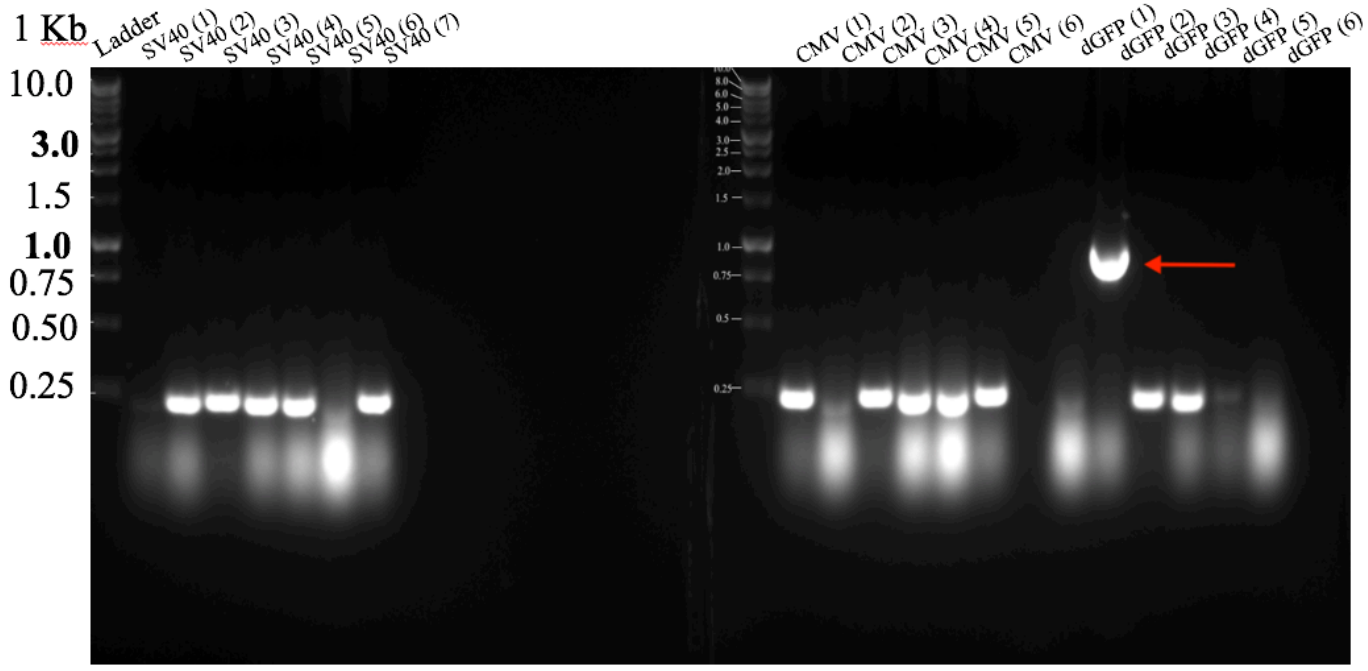


Figure 3.19 PCR screen of colonies for the presence of the dGFP insert following dGFP digestion and ligation into the SV40, CMV or vector that would contain dGFP alone**.** The lowest

band indicates no insert in the PCR products. The red arrow shows the only construct sample that contains the possible insert being ~850 bps.

3.5.5 Use of the dGFP Construct Obtained in Section 3.5.4 to Generate the Remaining Constructs

With the construct generated in *Section 3.5.4* above, that contained the dGFP sequence, next steps in the cloning were to insert the CMV and SV40 promoter sequences. The alternative method of generating the dGFP plasmids were more successful as depicted on the following digest screens in Figure's 3.20 and 3.21. These two figures show analysis of DNA from constructs generated using restriction enzymes *BsiWI* and *KasI* to insert the promoter regions into the dGFP vector plasmid isolated in Figure 3.19 (CMV in Figure 3.20 and SV40 in Figure 3.21). A band of approximately 580 bp (for CMV) or 330 bp (for SV40) on these 1.5% agarose gels indicated the presence of the correct insert size. The second approach of cloning the dGFP sequence and promoter regions together into the backbone vector was undertaken in parallel using *AflII* and *XbaI/NheI* digests for the backbone and inserts respectively. These results are shown in Figure 3.22 where it is apparent that this approach was successful only for two of the SV40-containing constructs.

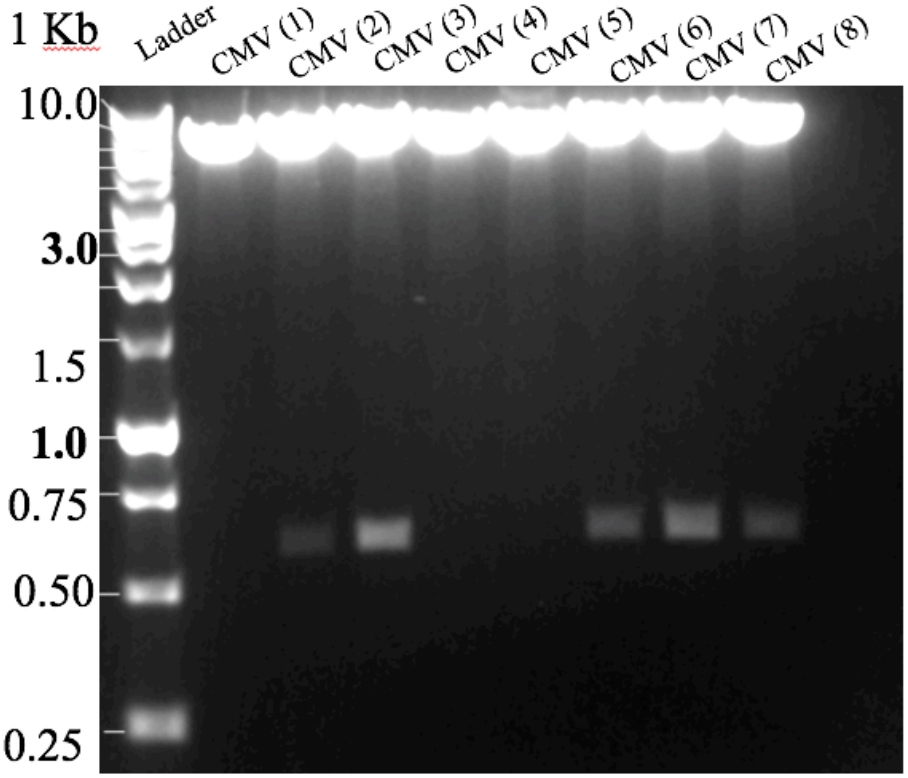


Figure 3.20 Verification restriction enzyme digests when ligating the CMV promoter sequence into the dGFP plasmid. The restriction enzymes used to screen the constructs were *BsiWI* and *KasI*, successful ligation constructs yielding a band at ~580 bp.

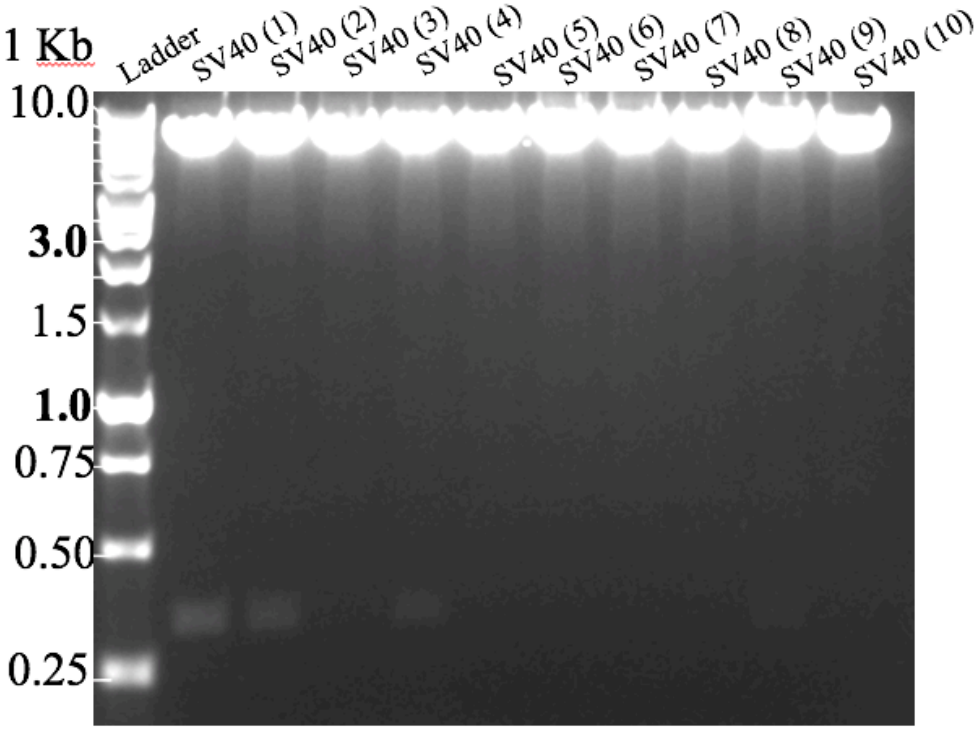


Figure 3.21 Verification restriction enzyme digests when ligating the SV40 promoter sequence into the dGFP plasmid. The restriction enzymes used to screen the constructs were *BsiWI* and *KasI*, successful ligation constructs yielding a band at ~330 bp (as seen in constructs labelled 1, 2, 4, and 9).

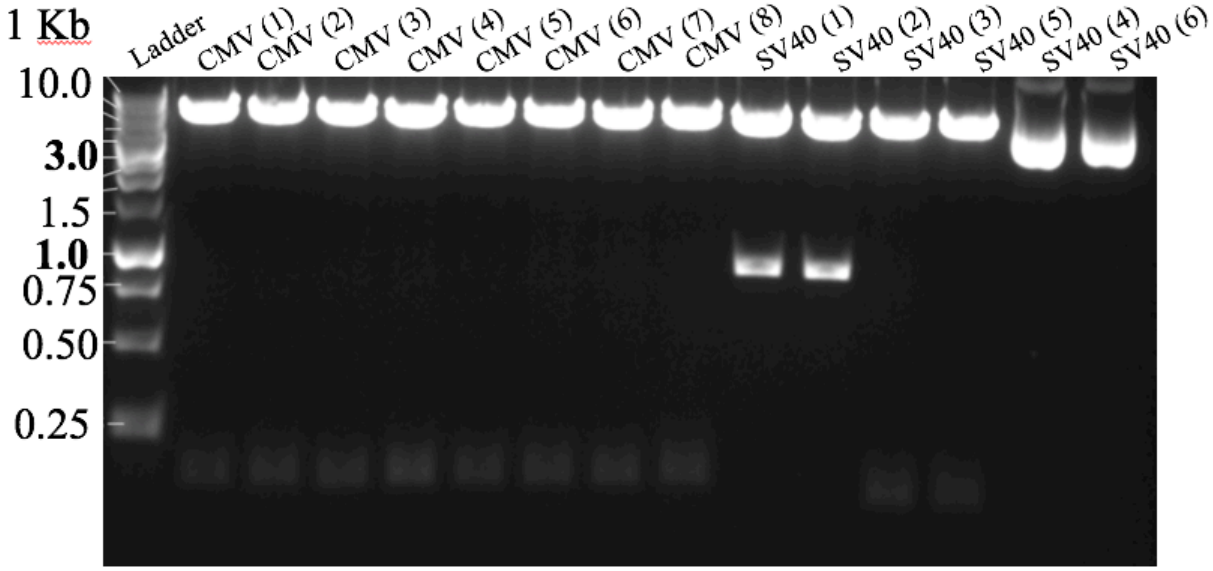


Figure 3.22 Verification restriction enzyme digests using dGFP as the insert sequence during ligation. The CMV/ SV40 plasmid constructs were used and cut with *AflII/ NheI* and *dGFP* insert with *AflII* and *XbaI*. The digest screen was run on a 1.5% agarose gel revealing only two positive constructs on the SV40 plasmid vectors as indicated by the presence of the dGFP insert band at approximately 1000 bp.

3.6 TRANSCRIPTOPN FACTOR BINDING SITE DESIGN AND CLONING

Once the basic promoter-fluorescent reporter-tag constructs were generated, it was necessary to identify and design transcription factor binding sites that could be cloned upstream of the promoters and that would bind transcription factors to be activated in response to different cellular stresses. The hypothesis was that these transcription factors would bind the complementary transcription factor binding sites and upregulate transcription of the reporter gene, ultimately resulting in enhanced reporter gene protein expression. In this way, responses to the cellular stresses would activate the transcription factors in a response that could be monitored.

3.6.1 Transcription Factor Binding Site Sequence Identification and Design

A set of transcriptional factors (TFs) and the sequences these bind to were chosen based on the regulatory roles they play during cell stress and homeostasis. Each of these is described in more detail later. Design and ordering of the binding site sequences was performed through the

use of the commercial *Thermofisher* GeneArt custom portal ordering system. Subcloning of the TF binding site sequences was performed by digesting out the insert from the donor plasmids they were provided in. The inserted TF binding sequences were located adjacent, in the forward direction (5'), to the SV40 promoter sequence. Although CMV promoter sequences were generated previously, time limitations and the fact that the SV40 promoter is weaker resulted in focussing on generation of these constructs only in SV40 constructs. The strength of the promoter would reflect more visible differentiation of the GFP expression levels with the weaker SV40 promoter allowing further increases in expression due to TFs to be observed (transcription is not maximal) whilst the strong CMV promoter would mask any impact of TFs as the expression is already considered high and maximal..

3.6.2 Transcription Factor Binding Site Sequence Design and Subcloning

Three designs were constructed for easier and comprehensive cloning with regards to each TF binding site. Each sequence was cloned three different ways into the same plasmid, containing SV40 and dGFP. To understand this visually, see Figure 3.24 for an example of the cloning involved. 'Whole sequence' cloning maintained the complete transcription factor binding site, including all promoter regions/ repeated sequences found in the various papers sourced. Also, downstream enhancers already embedded in the sequence were included in the whole sequence cloning protocol (see Figure 3.24).

Table 3.4 Design of cloning of TF binding sequences. The whole sequence represents no alterations were made during cloning. Half of the TF binding sequences, and minimal promoter regions had specific enzymes included as detailed in each section that enable digests in particular regions.

Sequence design	Enzymes associated	Labeling on gels
Whole sequence insert	MluI & PacI	A
TF without CMV/SV40	MluI & AflII	B
TF sequence without Minimal Promoter	MluI & SacI	C

Another sequence design contained the whole TF sequence *without* the SV40 core promoter. This was performed to test the SV40 strength and expressional effects on the TF motif. To test the

control sequences a designed plasmid without the minimal promoter region was also digested out. The minimal promoter region was integrated in the ordered TF sequences. The labelling on the gels and samples were noted as A-C for their respective cloning designs to simplify annotation (see Table 3.4). The latter two designs were created to contain a library for future experimental designs.

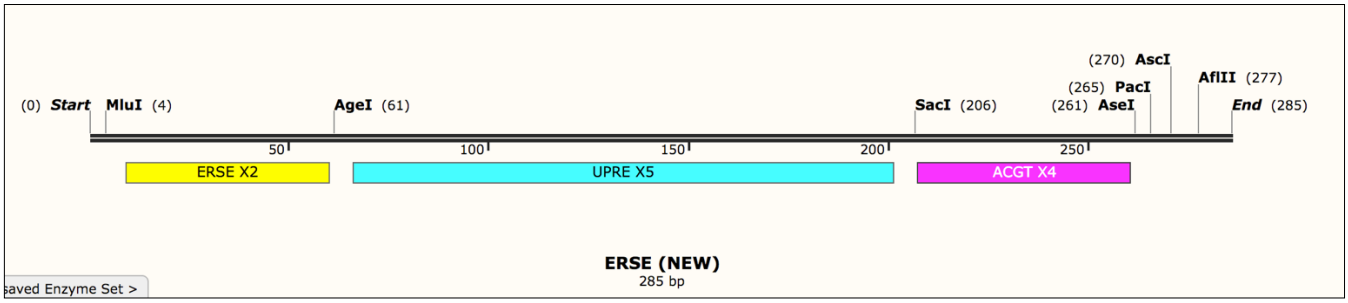


Figure 3.23 Schematic of ERSE TF insert inside a plasmid vector. This figure shows the significant restriction sites and the basis of the differential cloning. TF sequences were sourced from individual research papers described in the literature. This plasmid map was cut from SnapGene plasmid mapping tool. The whole sequence is 285 bps long.

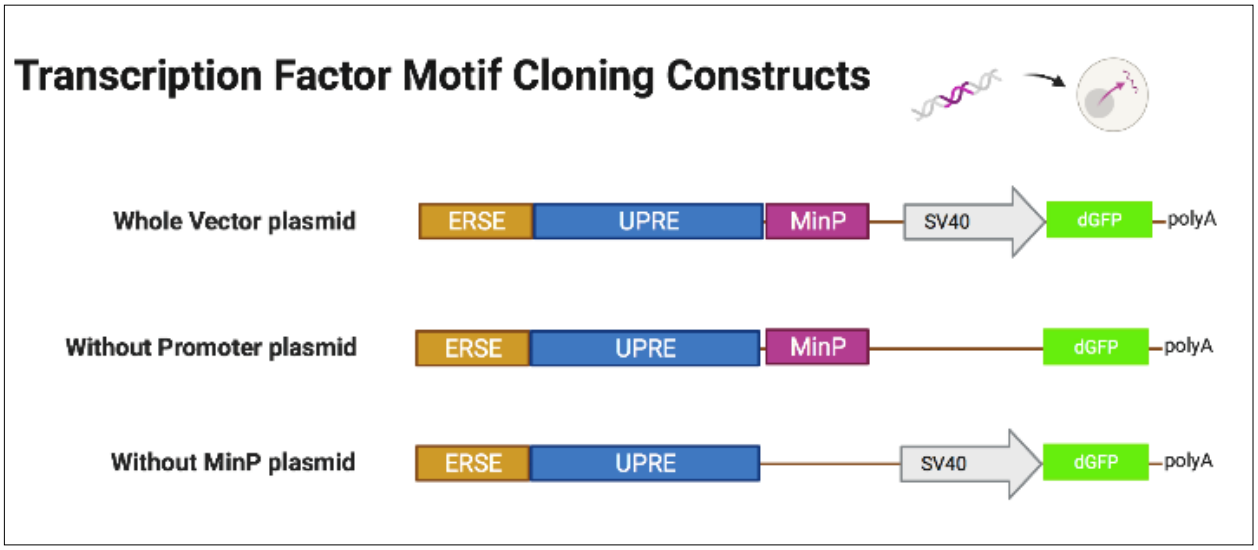


Figure 3.24 TF motif construct cloning sets for ERSE and the SV40 promoter. The first set contains the whole vector sequence of the TF, including the Minimal promoter region, which in the case of ERSE is the ACGT site. The following constructs either are without a minimal promoter site or a SV40 promoter site.

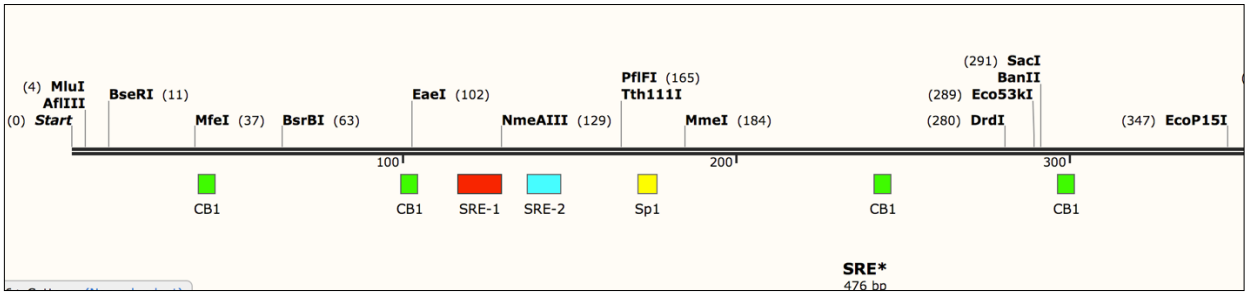


Figure 3.25 Plasmid sequence of the SRE insert. The sequence contains CB2 repeats and SRE repeats, the sequence was taken from a literature source, Inoue et al. 1998. The TATA box was integrated as the Minimal promoter site in Figure 3.26.

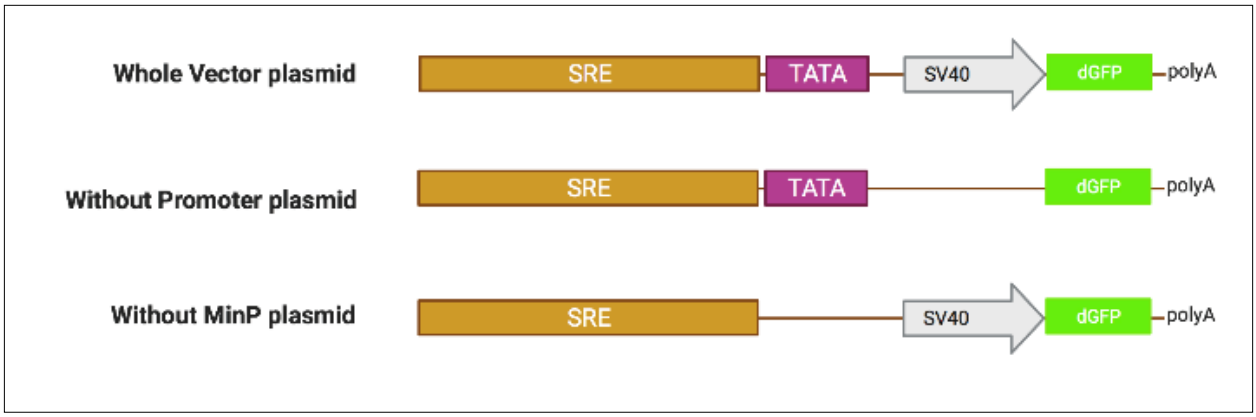


Figure 3.26 Schematic of the SRE cloning plasmid vectors. The whole sequence contains the SRE regulatory sites, a TATA box and the additional SV40, dGFP genes. Whereas, further cutting out the TATA box and SV40 promoters was commenced through cloning.

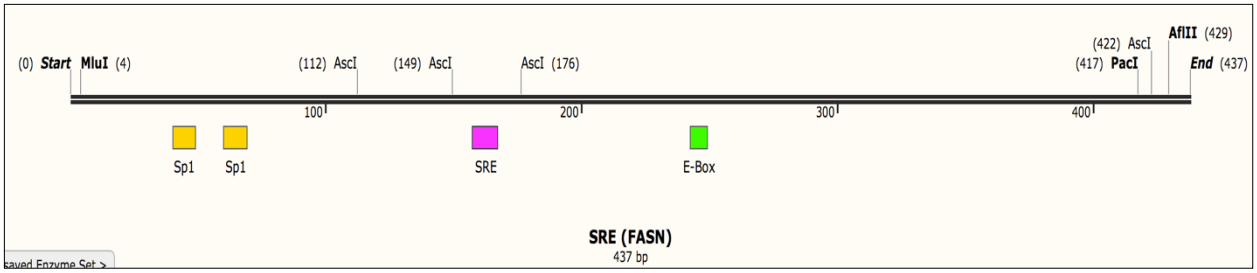


Figure 3.27 Plasmid map of the proposed FASN transcription factor motif. Sourced from Han et al. 2018 and displayed on a plasmid map through SnapGene.

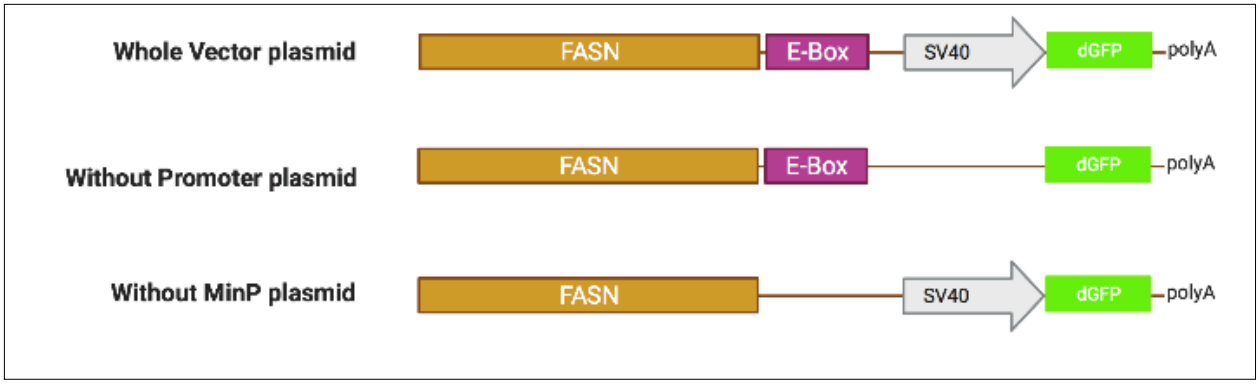
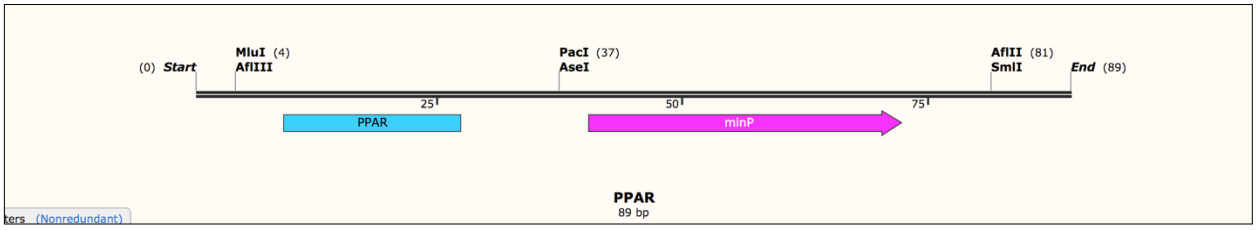
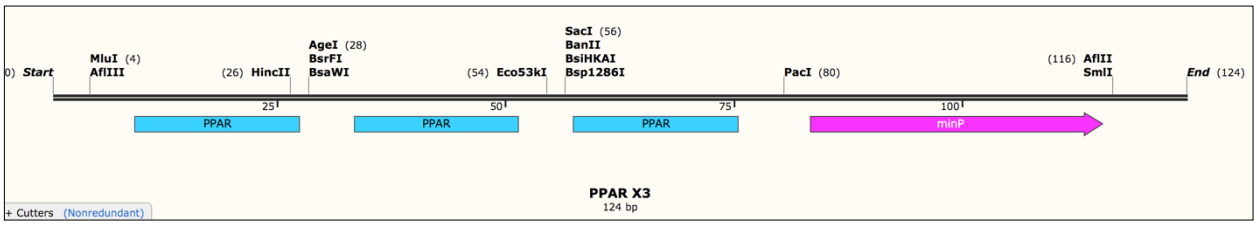


Figure 3.28 Schematic of FASN cloning plasmids. The E-Box on the FASN transcription factor motif was separated with a restriction site to allow for comparable analysis. The whole sequence contains the FASN repeats, the essential sequences, an E-Box, the SV40 promotional region and the dGFP fluorescent marker.

(A)



(B)



(C)

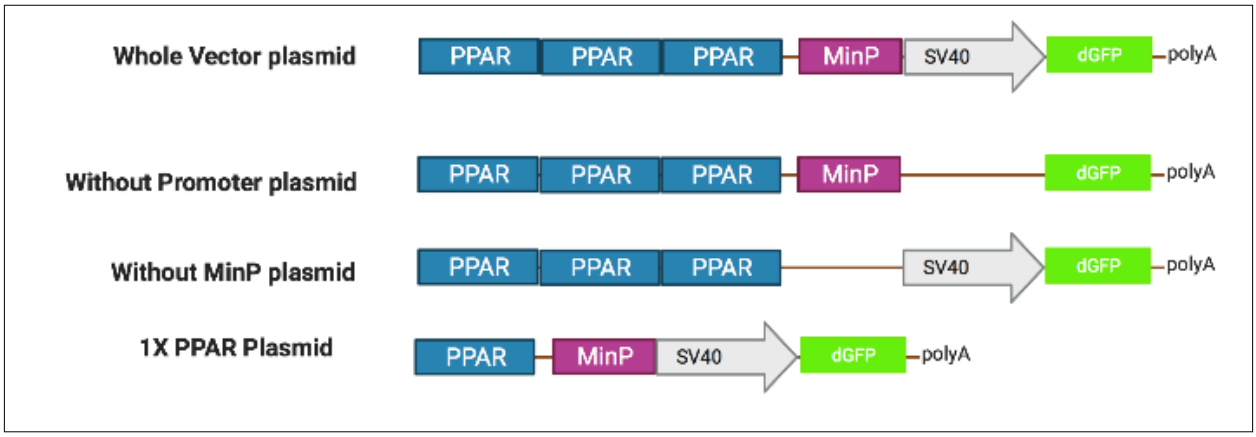
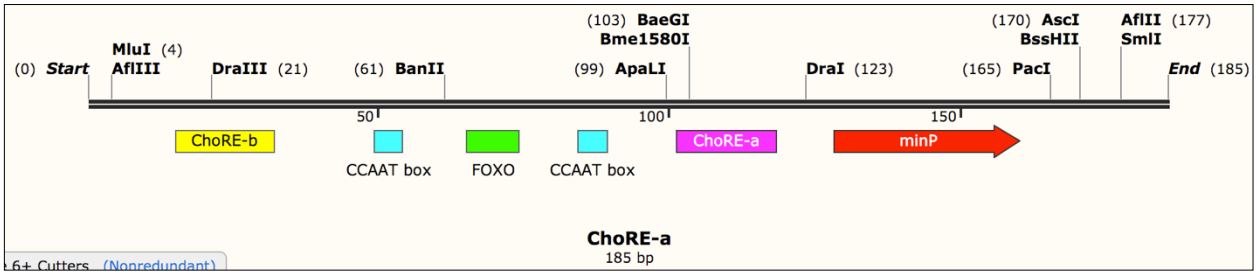
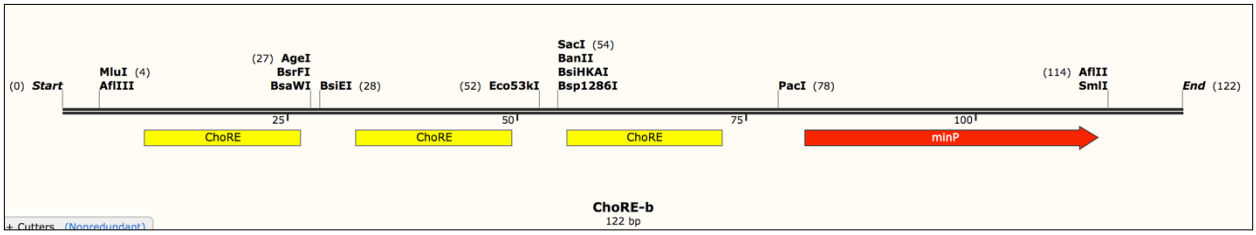


Figure 3.29 Transcription factor motif PPAR cloning schematic. (A) &(B) display the plasmid maps of the PPAR sequences in triplicate and single repeats with the implementation of restriction sites between the individual repeats allowing for flexible plasmid production. (C) The layout of the produced variations of plasmid vectors, with the possibility of comparing the expression levels of single repeats rather than the full sourced sequence from Tzeng et al. 2015.

(A)



(B)



(C)

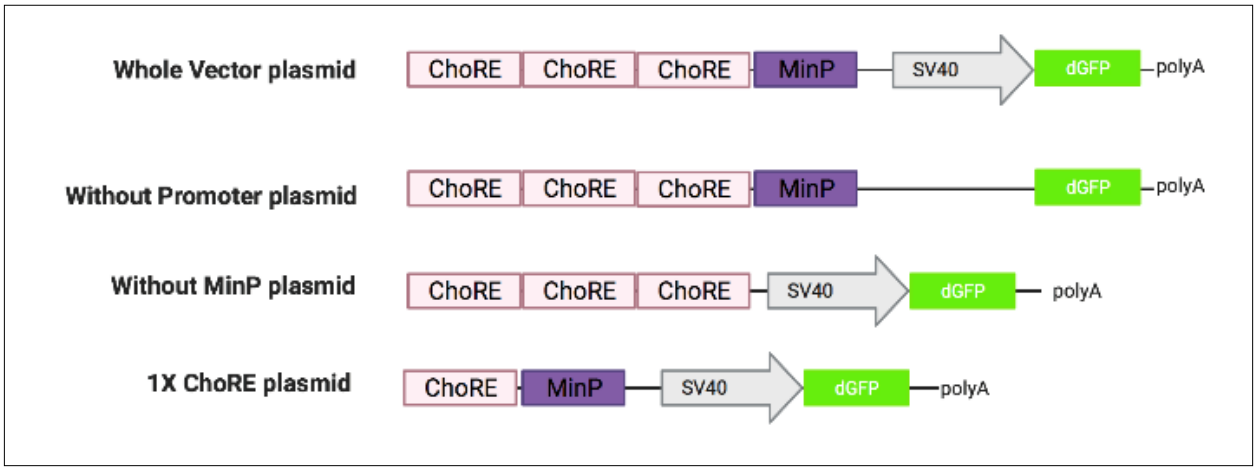


Figure 3.30 ChoRE-a &-b sequences along with the cloning layout. (A) Derived from a sourced article, containing several repeats enhancing expression levels, the ChoRE motifs and a minimal promoter region. (B) A triplicate of the ChoRE-b sequence with a minimal promoter sequences sourced from Yu F-X, Luo Y., 2009. (C) The cloning schematic of the ChoRE-b plasmids generated, along with a single repeat using the triplicate vector with integrated restriction sites allowing for this transaction.

3.6.3 ERSE & FASN TF Binding Site Sequence Cloning

To begin with, an ER Stress Response Element (ERSE) and Fatty Acid Synthase (FASN) transcription factor binding site were the first set of TF binding site sequences cloned into the SV40 + dGFP vector plasmids. To these motifs, bind transcription factors that upregulate transcription of genes in response to ER stress or fatty acid stress is necessary. Figure 3.23 reflects

the plasmid insert for ERSE, which was discovered and used by other research scientists. This same insert was cut from a donor plasmid and inserted into one of our SV40+dGFP acceptor plasmids. The different ways the cloning was performed is described in Figure 3.24. The '*whole vector plasmid'* map contains all the relevant information for the experimental analysis. The two subsequent plasmid schematics describe and show the differences between the promoters and constructs generated (Figure 3.24). The FASN TF motifs were sourced from Han et al. 2018, and are described in Figure 3.27 and Figure 3.28. Restriction digests were performed on the designed sequences in three ways, using Table 3.5 and 3.4 as guidance, organizing multiple sets of enzymes. Set A used *MluI* and *PacI* enzymes to digest out and release the whole sequence, as well as the vector plasmid dGFP + SV40 constructs. Set B used *MluI* and *AflII*, to release a smaller sequence with selected TF binding site elements from the complete sequence. Finally, Set C was used to release a further smaller portion of the full TF binding site, containing specific sequence elements. After digestion, the inserts and vectors were ligated together and cultured onto agar plates. Screening of the subsequent colonies for successful ligation of the insert are shown in Figure 3.31, where the numerical values represent the colony number taken from the growth plates, while the alphabetical value dictates which enzymes were used for the digestions.

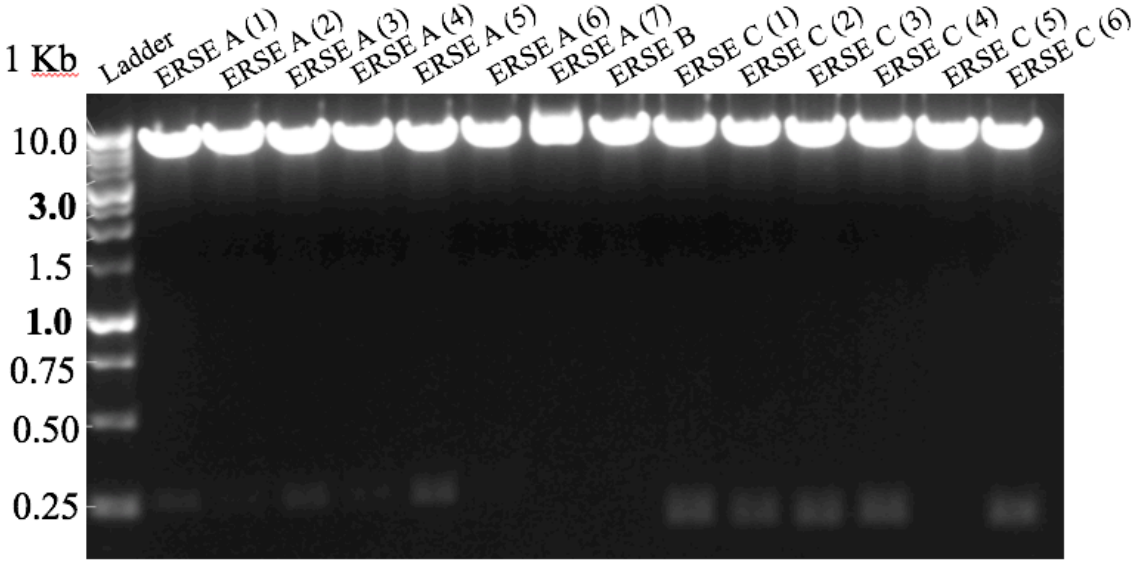
Table 3.5 Transcriptional factor binding and enhancer sequences size of expected bands in base pairs on agarose gels for verification.

Transcriptional Factor Binding Site Sequence Name	Size of band (bp)
ERSE	~285 bp
FASN	~427 bp
SRE	~470 bp
ChoRE-b	~116 bp
PPAR	~118 bp

The size of the ERSE A (whole sequence) band is ~280 bp and seen on the 2.0% agarose gel as a faint band near the 250 bp marker in multiple samples (Figure 3.31 A). Ligations with the B sequence did not reveal any bands; therefore, this was repeated, along with FASN C ligations. The ERSE C ligations were successful, as indicated by the second band at ~280 bp (Figure 3.31 A).

The C sequences should give smaller bands than A, as they are shortened when digesting. FASN A sequences were positive as shown by the band at ~430 bp (Figure 3.31 B).

A.



B.

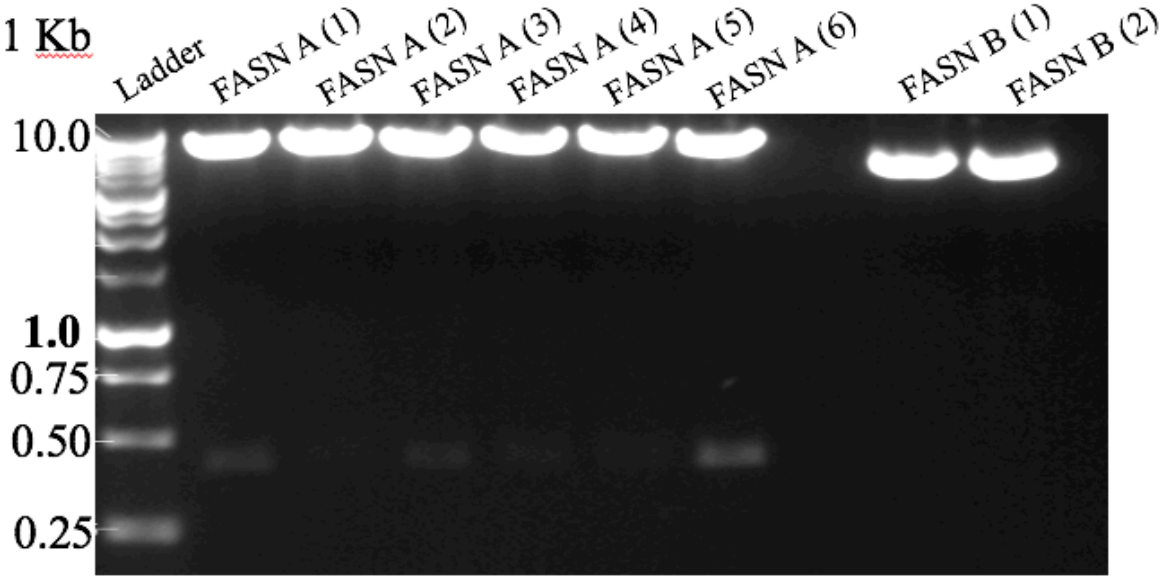
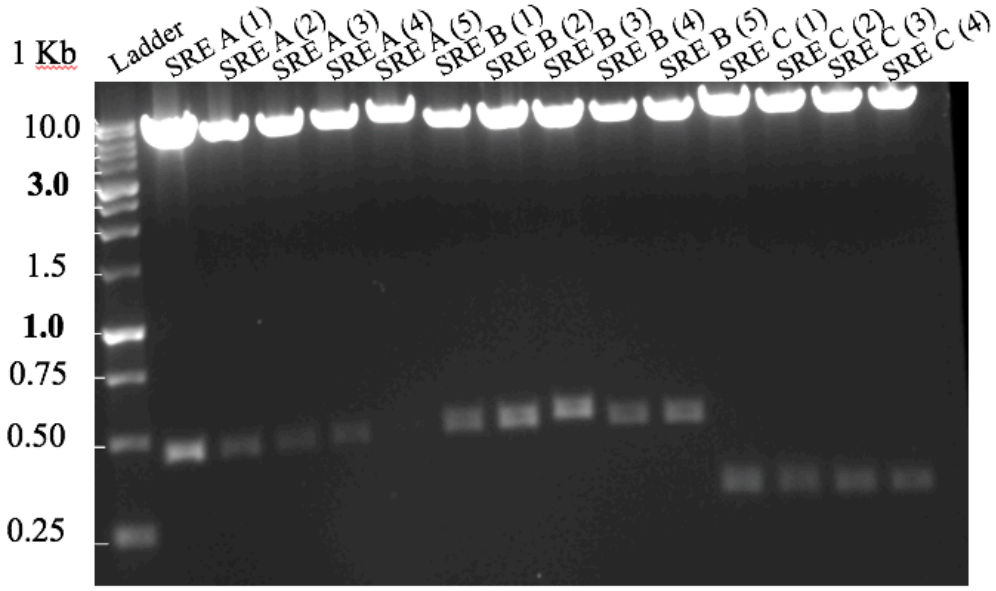


Figure 3.31 Digest screen of ERSE (A) and FASN (B) TF binding and enhance site subcloning. The numerical values indicate the number of colonies selected and letter values are associated with the type of sequence insert the plasmid contains. Successful ligation was indicated by a band at approximately 250 bp for the ERSE constructs and 430 bp for the FASN constructs. Constructs labelled *A, B, & C* refer to the cloning protocol, see Table 3.4.

3.6.3 Cloning of SRE, PPAR and ChoRE-b Transcription Factor Binding and Enhancer Site Sequences

The remaining transcription factor binding, and enhancer site sequences were cloned in a similar manner to the ERSE and FASN constructs described in *Section 3.6.2* using the A, B, C cloning arrangement (See Table 3.4). The SRE TF contains a TATA box instead of a minimal promoter sequence, as mentioned and described previously by (Inoue et al. 1998) that promotes transcription activity (see Figure 3.25). The TATA box was considered as the minimal promoter investigated (Figure 3.26) +/- the SRE TF. Figure 3.29 depicts the cloning scheme and the inserts for cloning the PPAR sequence previously described by Yu F-X, Lou Y., (2009). The PPAR sequence was cloned as a single sequence but also as a triplicate repeat sequence to investigate whether multiple repeats enhanced transcription from this sequence compared to a single stand-alone sequence. ChoRE constructs were generated analogously to the PPARconstructs as single and triple repeat sequence motifs (Figure 3.30). Figure 3.32 depicts the digest verification analysis of SRE transcription factor binding and enhancer site A, B, and C sequence constructs (see Table 3.4). The 1.5% agarose gel showed the presence of the expected band for successful cloning in nearly all colonies and lanes (lower band in all cases).

A.



B.

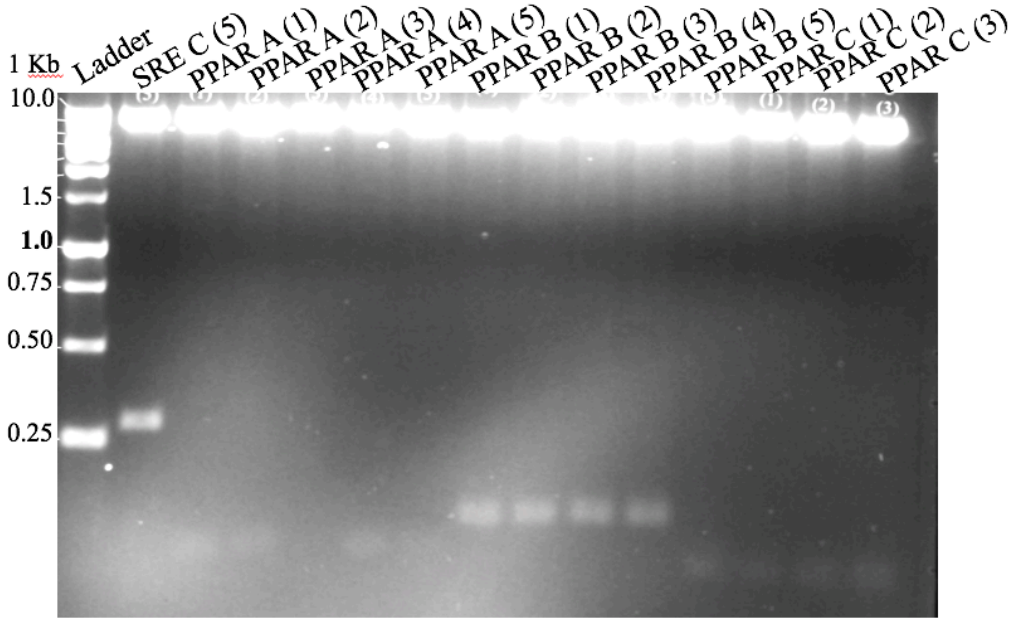


Figure 3.32 (A) Verification of SRE TF binding and enhancer site sequence cloning of sets A, B and C by restriction enzyme digestion. The band at ~470 bp is seen for constructs A, and B. Set C is observed at a lower size as expected due to the decreased size. (B) Digest screen of SRE and PPAR TF binding and enhance site sequence subcloning. The numerical values indicate the number of colonies selected and letter values are associated with the type of sequence insert the plasmid contains. Successful ligation was indicated by a band at approximately 118 bp.

One colony sample from the SRE C cloning was loaded on the next gel, due to insufficient wells in Figure 3.31 (Figure 3.32). DNA digestions from the PPAR TF binding site and enhancer sequence were loaded on a 2% agarose gel due to the small size of the band. The expected band was visible with adjusted brightness settings on the gel visualization system, somewhat compromising the resolution of the bands, however, they were clearly visible confirming the presence of the desired insert (Figure 3.32). Successful ligation was indicated by a band at approximately 118 bp for ChoRE insertions in Figure 3.33.

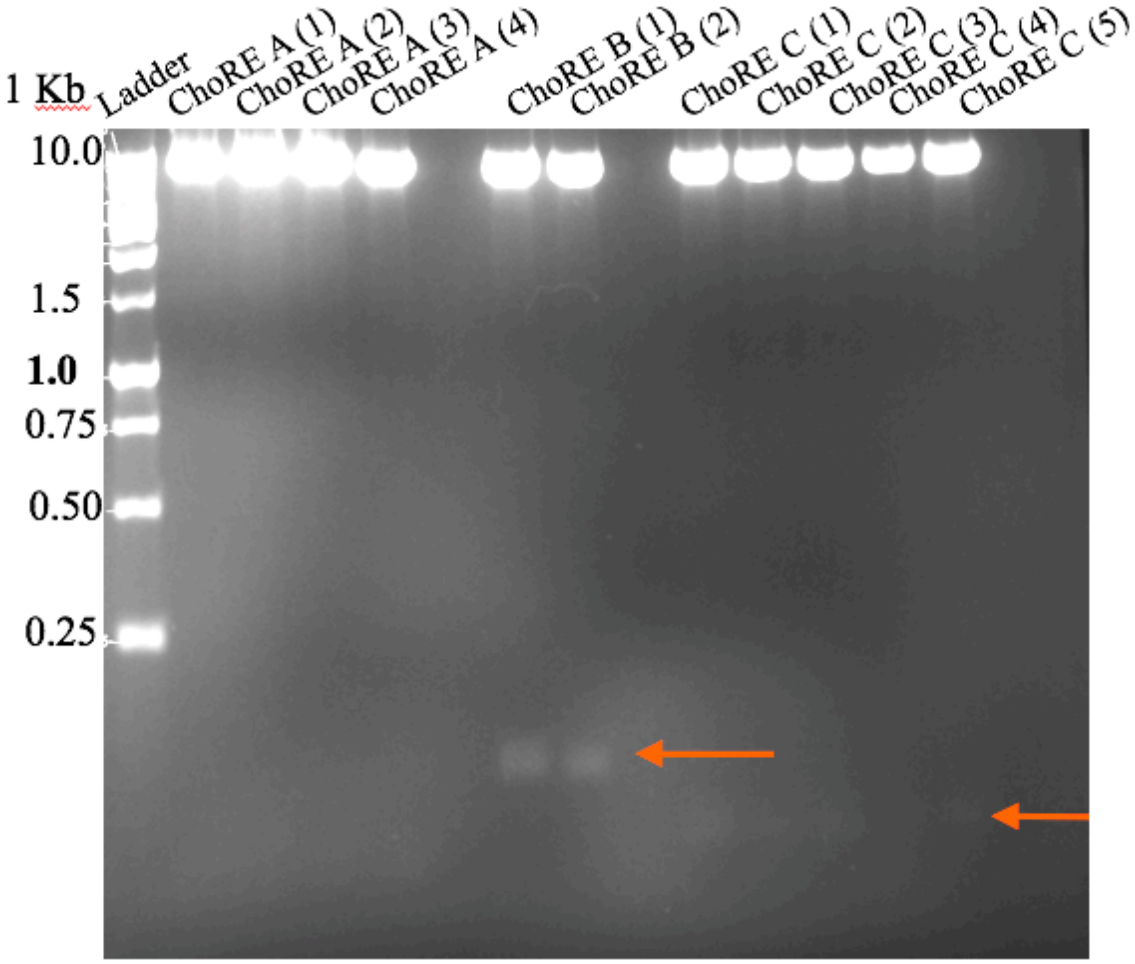


Figure 3.33 Digest screen of ChoRE A, B, C transcriptional factor binding and enhancer site sequence subcloning. The numerical values indicate the number of colonies selected and letter values are associated with the type of sequence insert the plasmid contains. Red arrows indicate where the band are present, although not noticeable under low resolution settings. Only sets B and C have evident bands at ~118 bps.

The final plasmid products are depicted in Figure's 4.2, 4.5, and 4.7. The relative sizes of the each segment is no to scale, but each part plays an important role in the expressive nature of the plasmid.

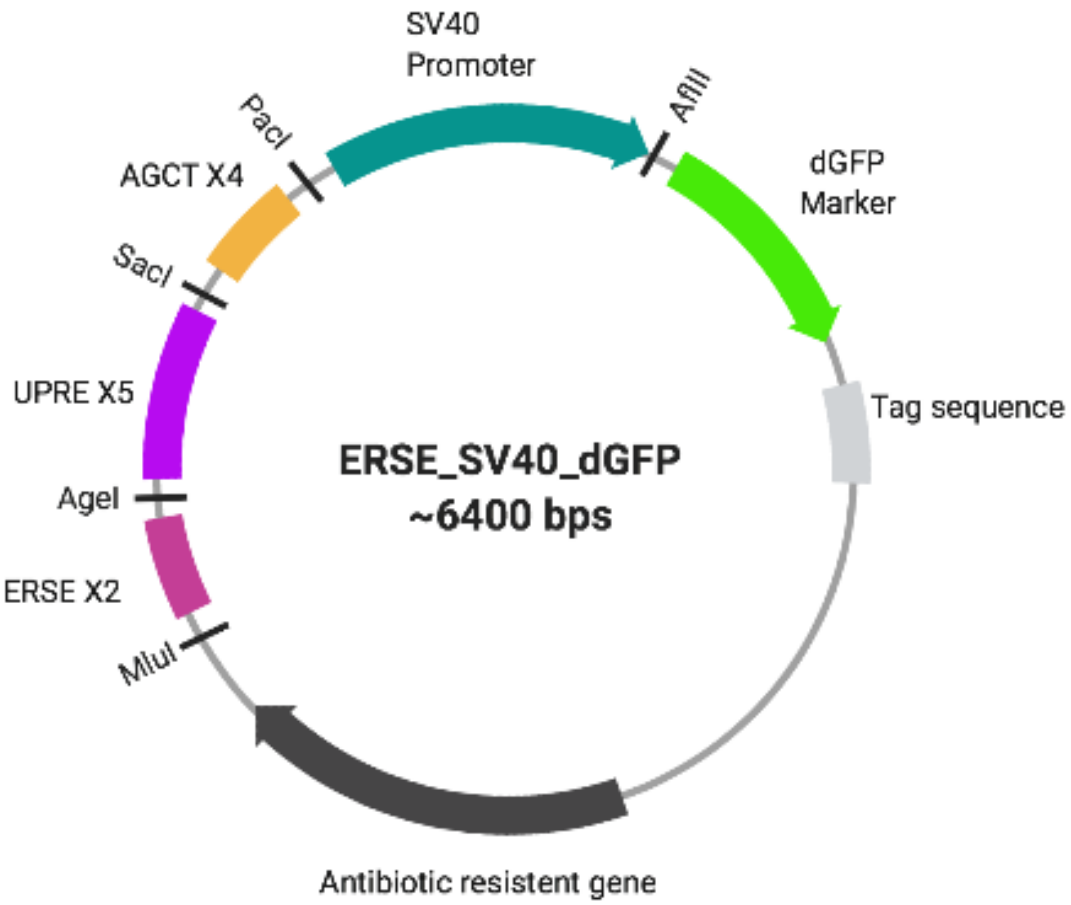


Figure 4.2 The cloning of the TF motif in one of the constructed recombinant plasmids. This plasmid contains the promoter region SV40, the selectable marker dGFP, and a tag sequence. This schematic represents the location and sequence the genes are placed within the MCS, along with critical restriction sites implemented during the cloning process. The sizes of the individual sequences are not to scale.

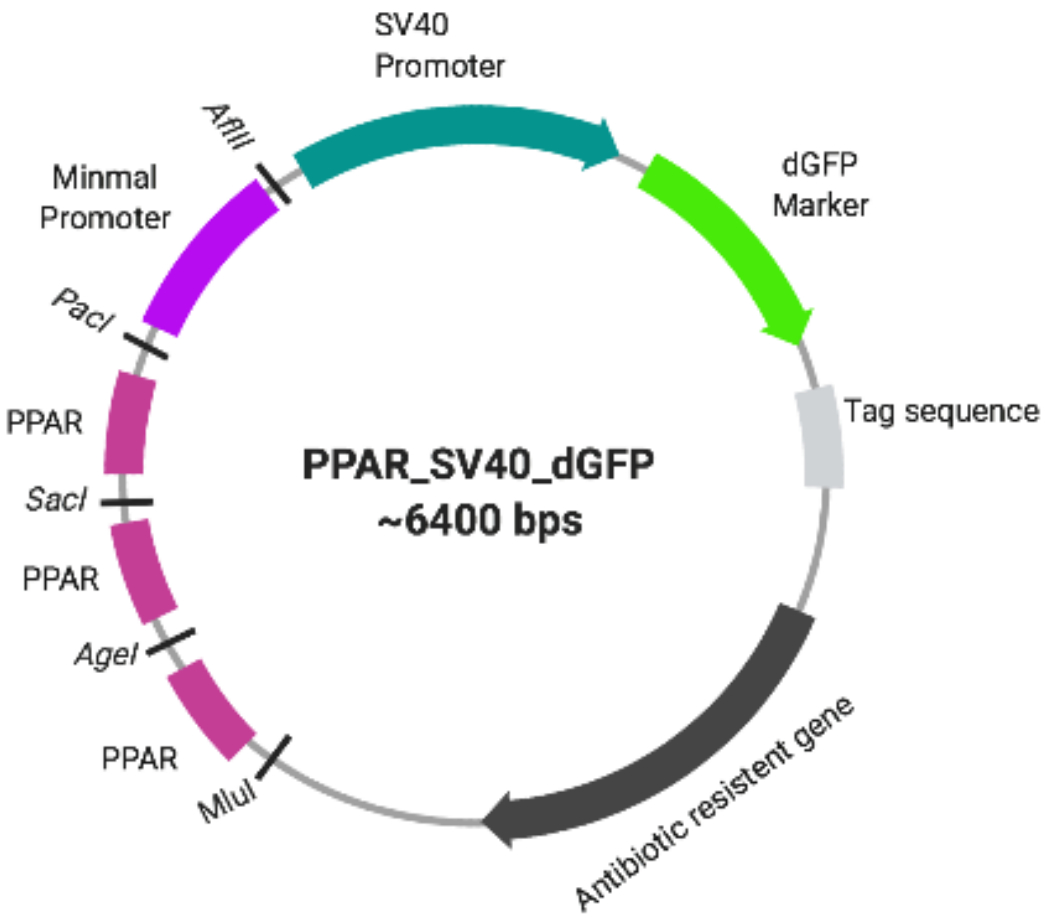


Figure 4.5 Designed Plasmid map of PPAR (X3) completed construct post-cloning. The plasmid contains the PPAR in a triplicate repetition, along with the minimal promoter sequence. The restriction sites allow flexible cloning for desired plasmid constructs.

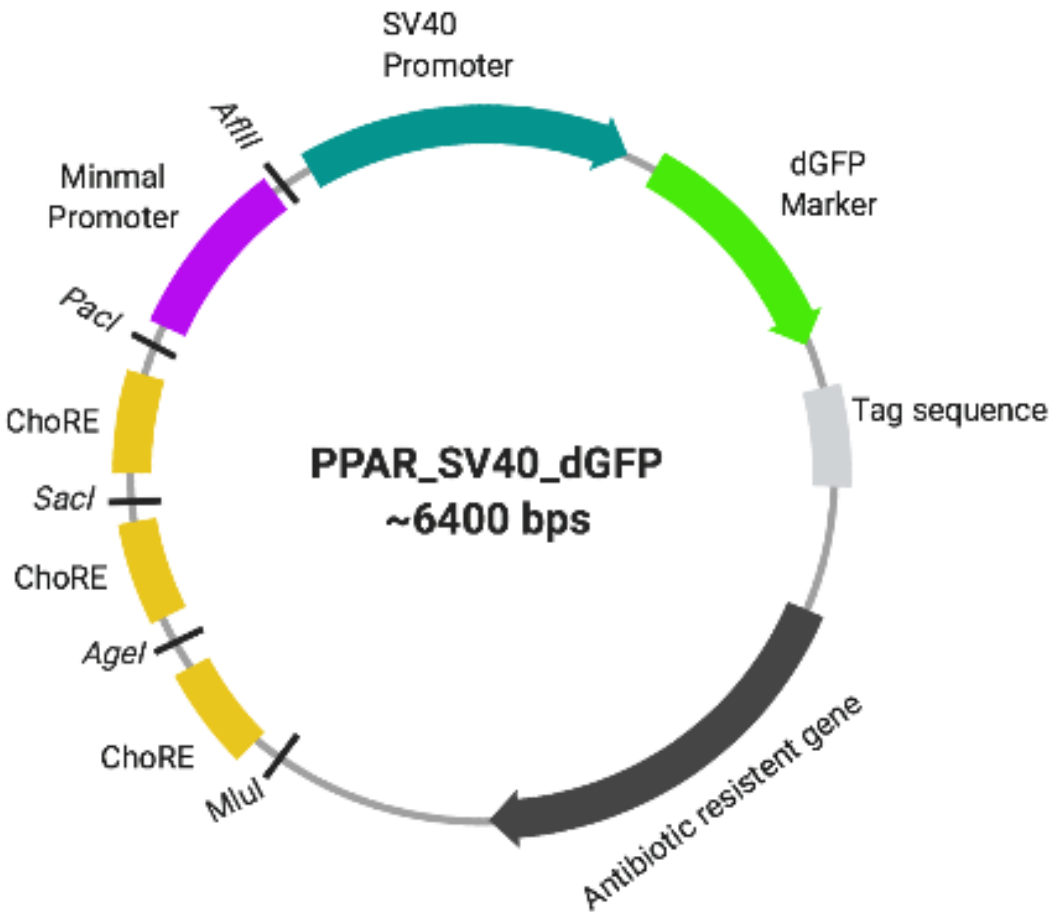


Figure 4.7 Plasmid construct map of the recombinant ChoRE TF motif generated in this study. The map contains a triple repeat of the ChoRE TF motif along with an addition of a minimal promoter identical to that of the PPAR motif sequence. The plasmid also retains the existing genetic information such as the antibiotic resistance gene (Amp), a tag sequence, SV40 and dGFP.

3.6.5 Minimal Promoter Cloning

The TF binding sequence motifs that were inserted all contained potential minimal promoter regions, embedded within the TF binding sequence itself, aside from the promoter regions that were already present. The minimal promoter (MinP) region for ChoRE and PPAR are exactly the same, therefore the same MinP was generated for both. The MinP was digested and ligated into the dGFP-only plasmid vector, using *PacI* and *ApaI*. Since the MinP is only 20 - 30 base pairs in size, it would be too small to observe on a gel separation, therefore the region was

digested with *Pac*I and *ApaI*. This region encompasses the MinP along with the dGFP, revealing a band at ~1000 bp. This was verified as shown in Figure 3.34, the second band indicates the MinP along with the dGFP construct. The band, when compared to other possible sequences, was above ~850 bp, where dGFP only would be expected. Confirmation of generation of the correct constructs was also achieved through sequence analysis commercially at *GeneWiz*.

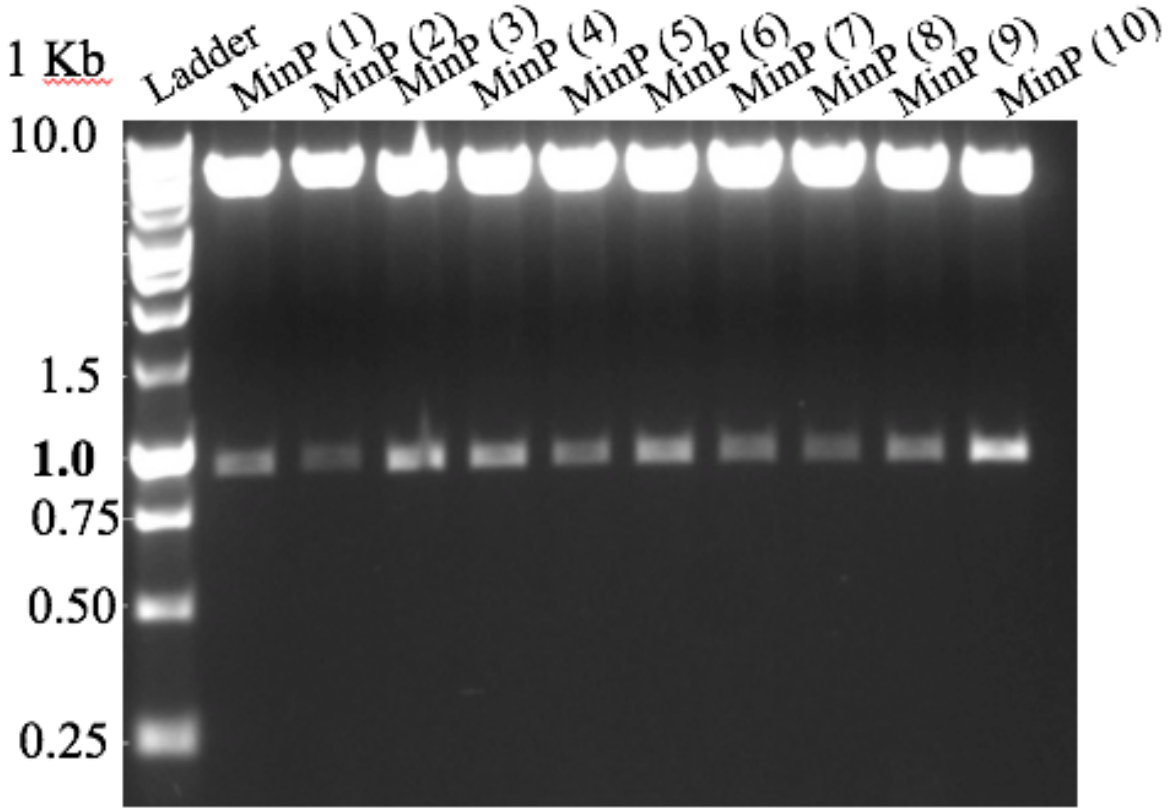


Figure 3.34 Verification of MinP ligation by restriction enzyme digest using *PacI* and *ApaI*. A band at ~1000 bp was expected, including the dGFP and MinP sequence. All 10 colonies chosen were regarded as successful ligations.

Next, the ERSE and SRE minimal promoters were selectively cloned into the dGFP plasmid vectors. The restriction enzymes *SacI* and *AflII* were used for cloning of the SRE minimal promoter from these sequences. These digests should release a band at ~177 bp. The enzymes were also used for verification after digestion and ligation with dGFP-only plasmid vector (Figure 3.35), confirming the presence of the expected insert.

The ERSE B plasmids were used to digest out the minimal promoter region, also known as the *ACGT box* (in other sources), to produce the MinP constructs, one of the designs described previously. The plasmids were digested with *SacI* and *ApaI.* Following digestions, ligations were successful as shown by a restriction enzyme digest screen (Figure 3.36). The agarose gel analysis revealed a band of a size of ~1000 bp, which corelates with the expected sizes.

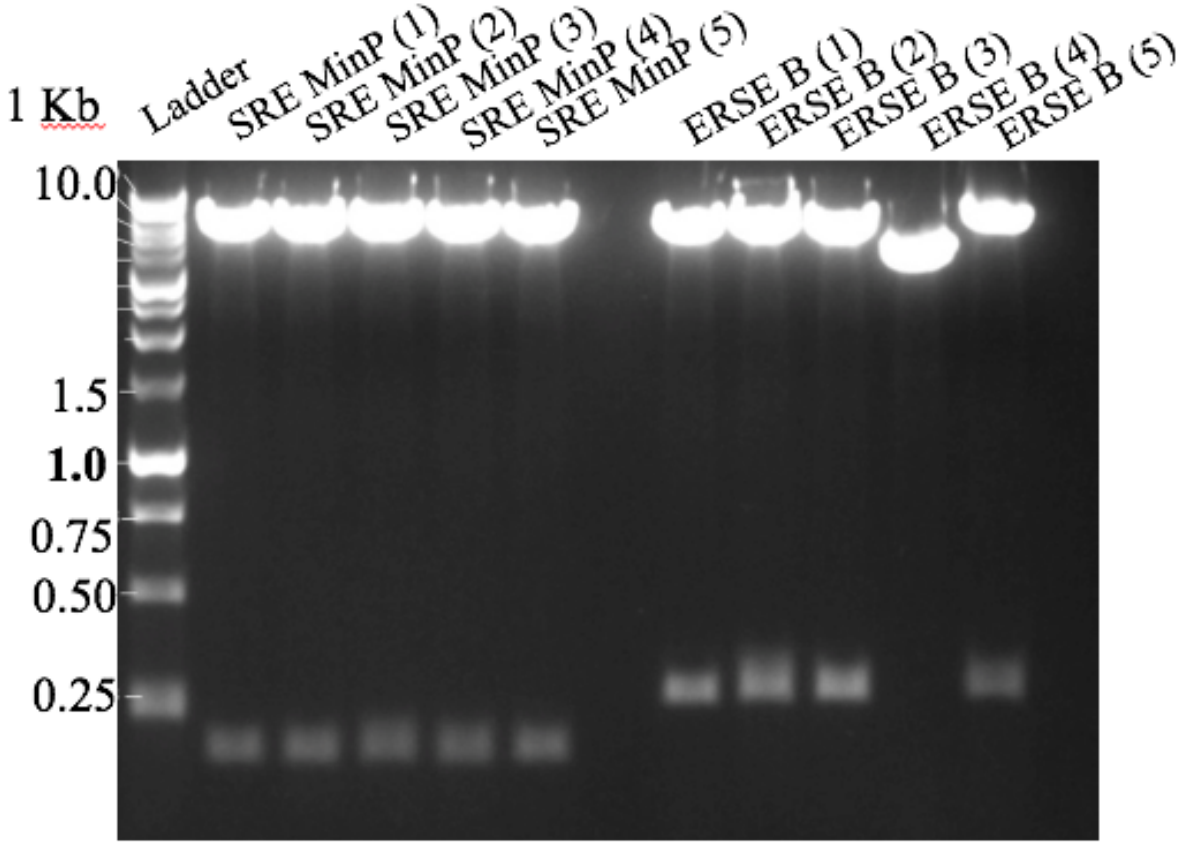


Figure 3.35 Digest screen of ERSE B and SRE MinP constructs on a 1.5% agarose gel. The SRE minimal promoter sequences were large enough to be visualized without including the dGFP region. Positive constructs were indicated by a band at 177 bp for the SRE MinP and 273 bp for the ERSE MinP.

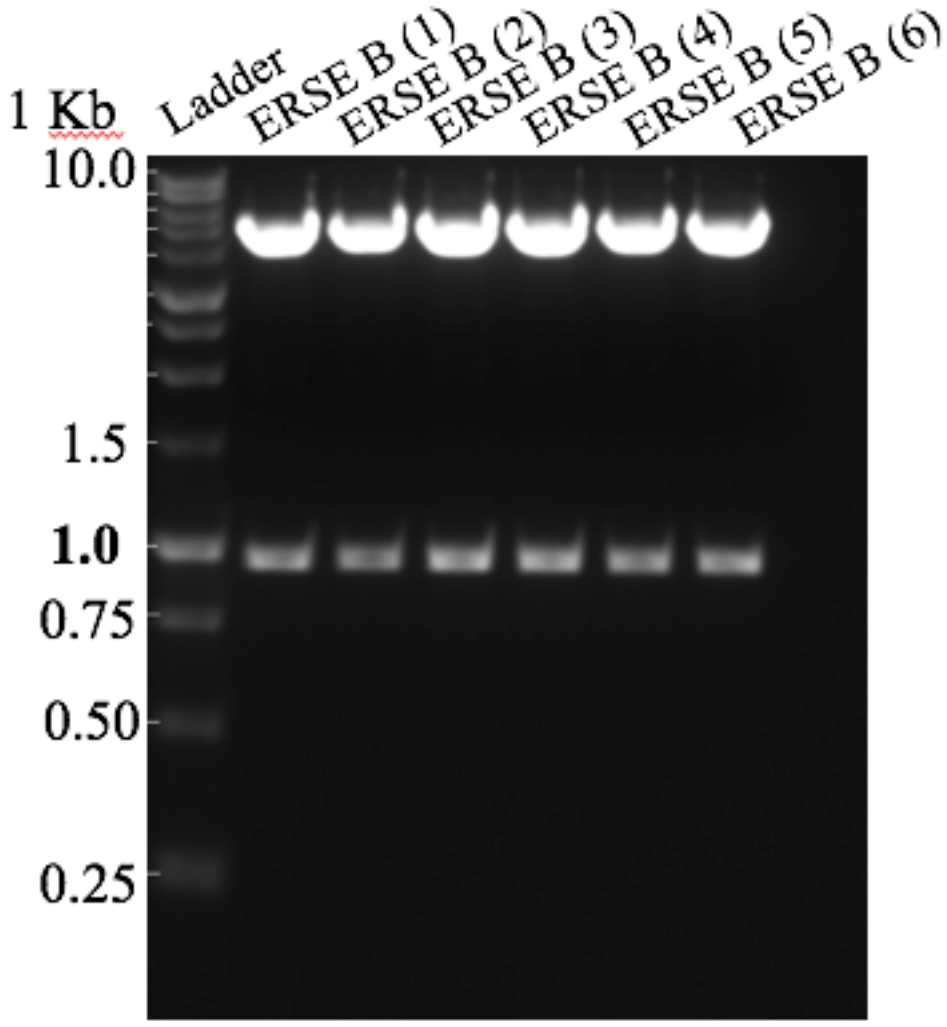


Figure 3.36 Digest screen of ERSE MinP constructs on a 1.5% agarose gel. The band reveals a ~1000 bp fragment indicating the dGFP along with the minimal promoter region in the ERSE motif.

3.6.6 Transfection of Constructs into CHO-S Cells and Analysis of Resulting Fluorescence Expression

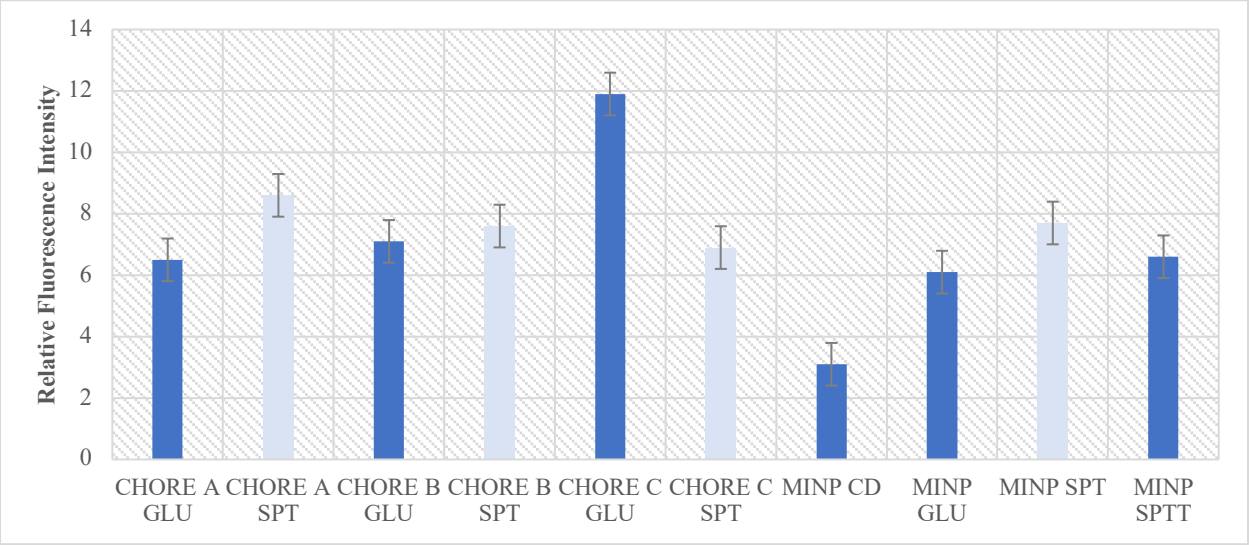
To start, all the PPAR and ChoRE-b constructs were transfected into mammalian CHO-S cells through electroporation. 20 µg of DNA was diluted in EB buffer solution and prepared for each transfection in a total volume of 200 µL. This included control samples 'SV40-only', along with the 'MinP-only' constructs. For 30 transfections, 320 x 10^6 viable cells were required and split 4 days prior to transfection. The initial cell seeding density was modified to attain the target

cell density of 10 x 10^6 cells/mL. The cells were counted on the Vi-Cell as being 13.84 viable cells/mL at a 95-99% viability and kept on pre-warmed media (37°C) for optimized cell growth conditions. Volumes of culture was 30 mL for each transfection, and in shaking conditions at 120 rpm at 8% CO_2. The cells were then counted and harvested at 24 hours after transfection. However, cell counts were extremely low, and the data is not shown.

Conditions for each construct differed due to the specific regulatory roles they play during cellular stress. Both constructs were transfected in triplicate using two different media conditions. This was to test the expression levels during stressful conditions (depleted/spent media) and during optimal growth conditions for the cells. PPAR was transfected into CHO-S cells cultured in: (1) with 10 mL of total CD-CHO media formulated with L-glutamine & anti-clump, and (2) with 50% (5 mL) of spent media, and 50% (5 mL) of CD-CHO media (included L-glutamine & anti-clump). ChoRE constructs were transfected into CHO-S cells and cultured in two differing conditions also in triplicate: (1) 100% (10 mL) of all spent media, with addition of L-glutamine at 4 mM and anti-clump solution, to match the CD-CHO media the cells grew in; (2) 100% of spent media (+L-glut & anti-clump) with addition of glucose at 12 g/L.

At the 24-hour time point, 250 µL of cells were harvested with 500 µL of 1X PBS to analyse on the flow cytometer. Data obtained from the harvested cells are shown using the mean fluorescent values and geometric means of the histogram analyses (Figure's 3.37 & 3.38). The averages of each triplicate transfection was taken and plotted to compare fluorescent expression levels from the different plasmids in CHO-S cells grown in specifically selected environmental conditions. Although expression was generally low based on grow rate and flow cytometry data, we can observe that there are expression differences from the different constructs.

A.



B.

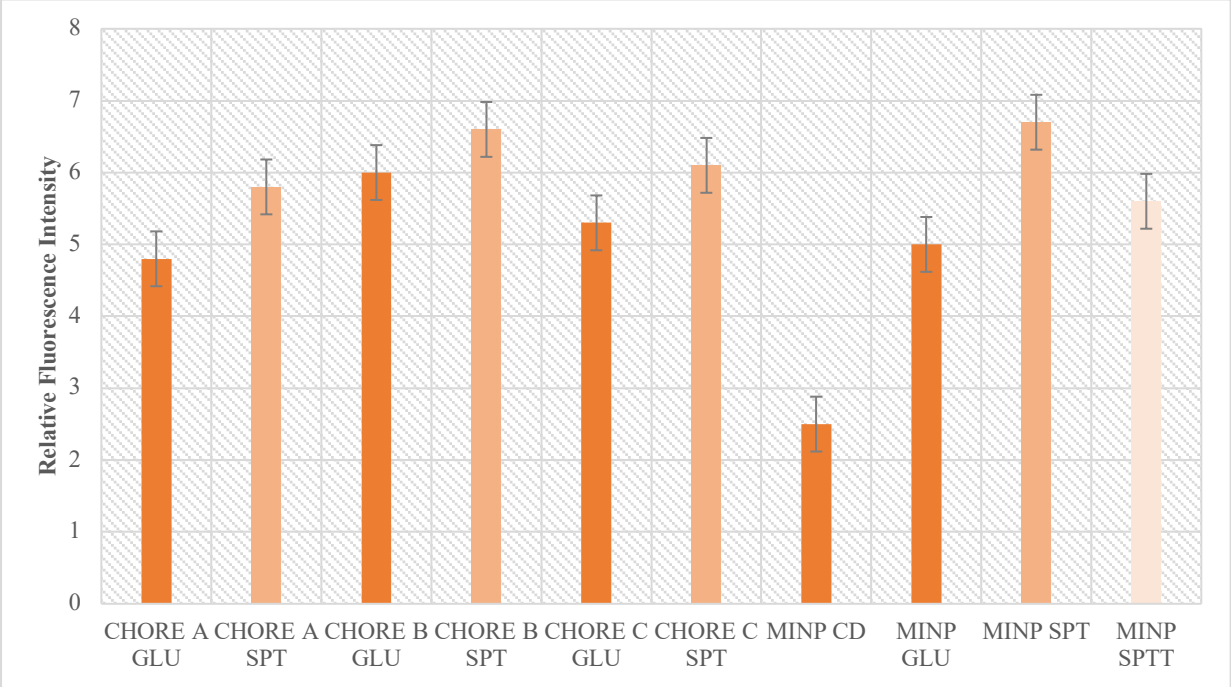
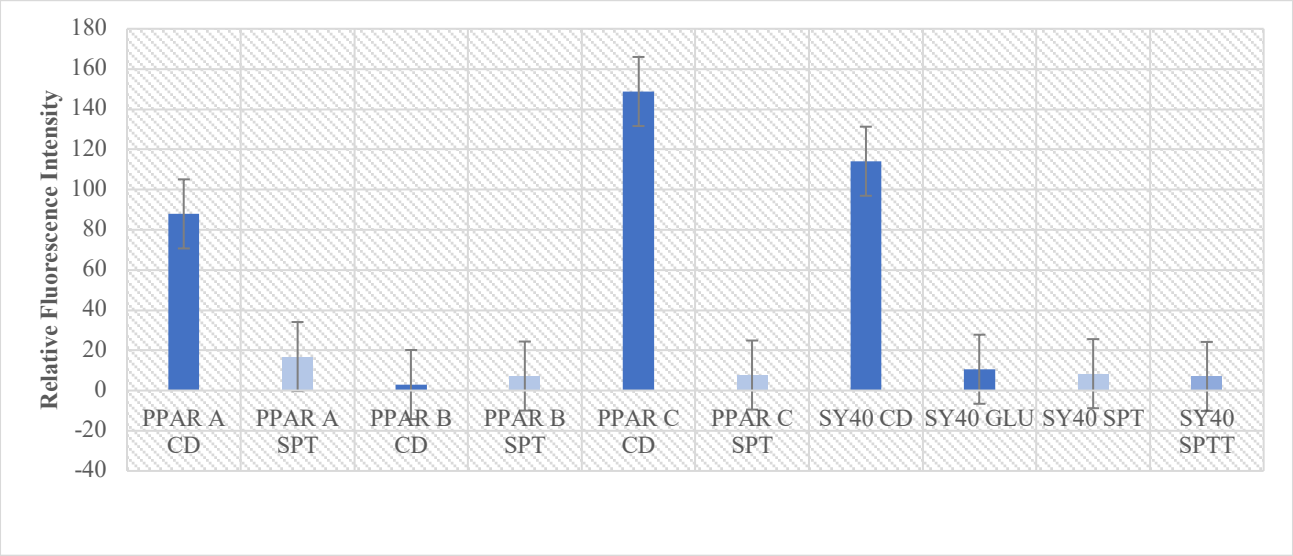


Figure 3.37 Mean (A) and geometric mean (B) values of ChoRE constructs along with the negative and positive controls. GLU indicates the addition of glucose in the media after electroporation; SPT as the spent media only, with no additional enhancements into the media. CD media was used for the control vectors to compare changes in expression. n=3.

A.



B.

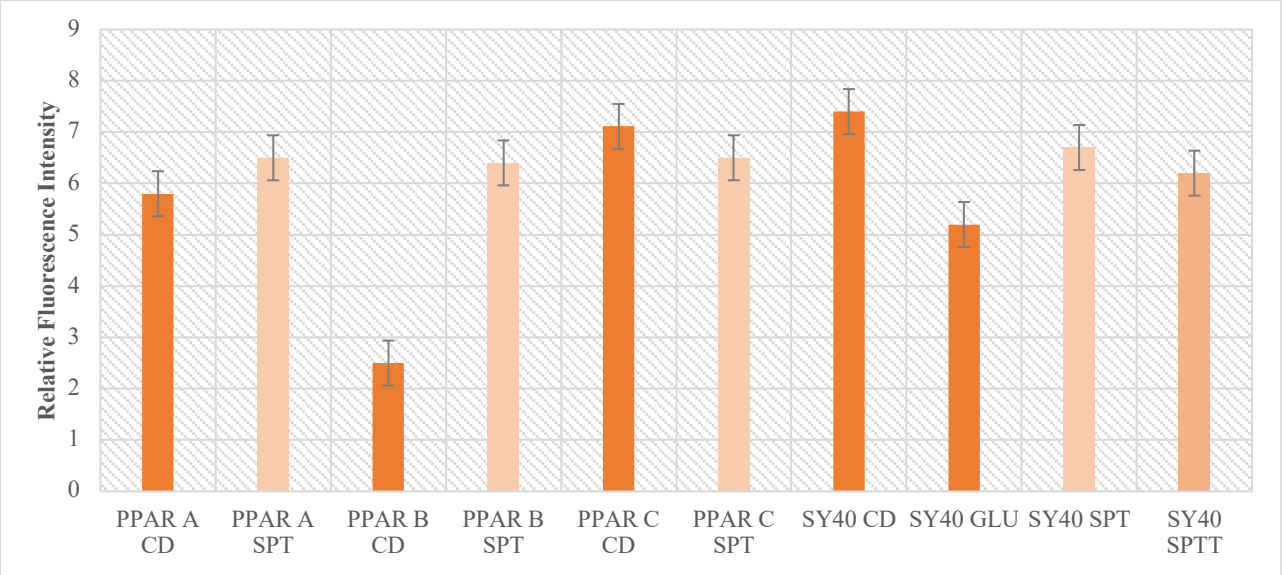
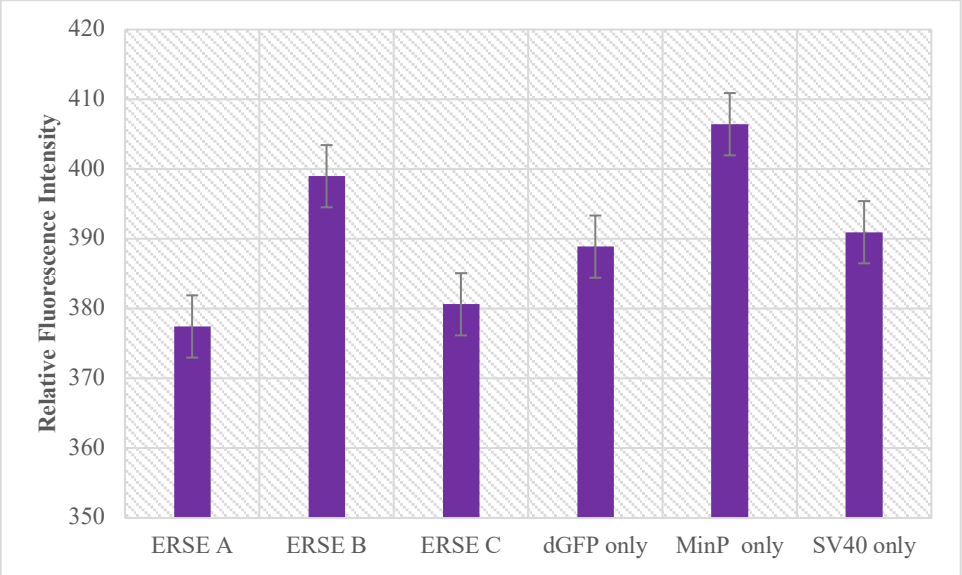


Figure 3.38 Mean (A) and geometric mean (B) values of PPAR constructs along with the negative and positive controls. The averages of the triplicate are presented here. GLU indicates the addition of glucose in the media after electroporation; SPT as the spent media only, with no additional enhancements into the media. CD media was used for the control vectors to compare changes in expression.

ERSE-cloned plasmids were transfected into CHO-S cells in CD-CHO growth media. This protocol called for no alterations of the environmental conditions. 20 µg of DNA was transfected into CHO cells, and transferred into 30 mL of total CD-CHO growth media, with additional L-

glutamine and anti-clump. Figure 3.38 depicts the mean and geometric mean values of the triplicate data from the flow cytometer. The data obtained proved that the transfections had been successful. The relative fluorescence was higher from some constructs compared to others, which warrants further investigation. It can also be seen that the constructs without the SV40 promoter region still give gene expression. This can also be seen for the control constructs, where the MinP only plasmid gave a higher fluorescence intensity than the SV40 only.

A.



B.

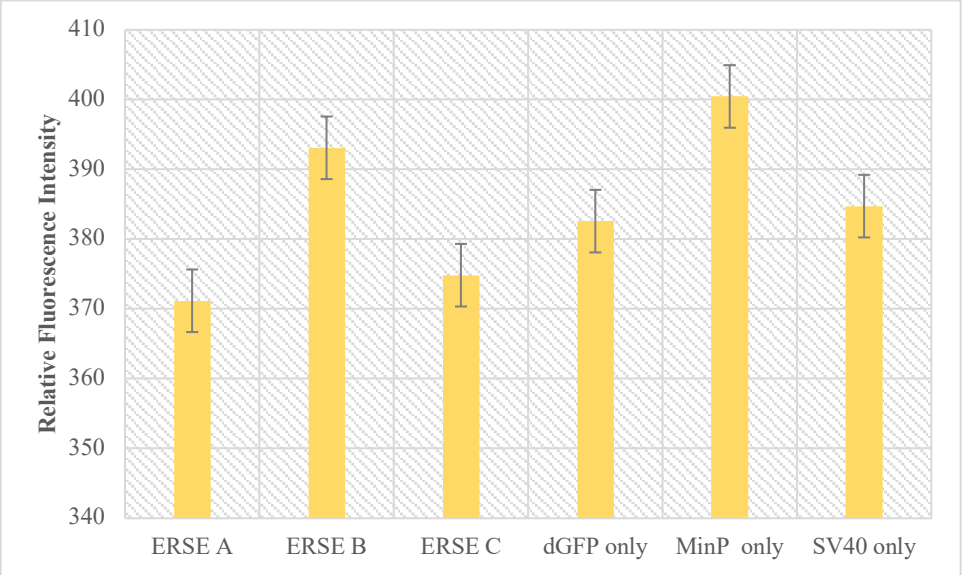


Figure 3.39 Mean (A) and geometric mean (B) values of a triplicate set of data from histograms provided by the flow cytometer. ERSE A, B C are differing lengths of the TF motif sequence within a dGFP-containing plasmid. Each analysis counted 10,000 events. Error bars show +/- SE of the mean.

4 DISCUSSION

The long-term focus of this study is to contribute experimental data to assess whether reporter genes whose expression is linked to the presence of transcriptional factor sequences upstream of the promoters driving reporter gene expression can be utilised as markers and to analyze and detect the perception of different stress responses in cellular stress/regulation in CHO cells making biotherapeutic proteins. Construction of the final plasmid vectors containing the TF motifs was completed through subcloning of key features prior to TF insertion. Firstly, the multiple cloning site (MCS) was enriched by the addition of restriction endonuclease sites which were not present in the original *Hygro3.1* donor plasmid. This allowed for multiple approaches to cloning and verification of correct inserts being presence through availability of a wider range of restriction enzyme sites.

In addition to the modification of the plasmid to aid insertion of promoters and transcription factor sites, the plasmids were designed to allow additional antigen tags to be fused on reporter or other genes inserted into the MCS. Each tag chosen was done so to provide a method of detection and purification of targeted proteins without any protein-specific probes (Liu, 2016). Choice of several tag epitopes depended upon the downstream analysis of the protein, such as western blot or immunodetection. Advantages of utilising the tag sequences was for rapid detection without the need for an antibody (and it is assumed the tag has little impact on the 'normal' protein function(s)), to allow pull down of any reporters or proteins inserted to determine if these were interacting with other proteins or nucleic acids and as another means of quantitation of reporter/protein expression. Although the use of fluorescent proteins allows in time visualisation of reporter gene expression, only active reporter protein will be observed and any incorrectly folded, inactive protein or protein that is partially degraded will not be detected via fluorescence. The tag provides another means of validating the fluorescence data and of potentially detecting total protein produced if some of the protein is indeed inactive or incorrectly folded/fragments/degraded.

The different tags utilised here (V5, Strep, His, & Flag) had additional restriction site sequences surrounding the tag so that the tag could be fused N- or C-terminally. Importantly, this was

designed such that a N-terminal start codon was included and at the C-terminal, a stop codon was at the end of the tag. Using restriction enzymes sites *MluI* and *ApaI*, the designed MSC and tag sequences were inserted into the donor plasmid. This encompassed most of the key sites that have been modified during the project. The tag fragments themselves are short in length, usually 20-50 nucleotides long. Consequently, the addition to the plasmid constructs became problematic in the visualization or detection of successful cloning on a gel by determining a change in size of excised bands due to the presence/absence of the tag, for instance on a gel (Figure 3.3). Verification of the cloning was achieved through digest screening using the same enzymes and observation of release of a correct band size. Results indicate that the tag sequences were inserted correctly (Figure 3.1 & 3.2), and subsequent sequencing confirmed this.

The required transcription factor binding site sequences, enhancer sequences and promoters were then cloned upstream of the MCS where the appropriate reporter gene was to be cloned. As described, a promoter is an arrangement of short regulatory sequences that recruit the transcription machinery and sequence-specific transcription factors that drive transcription initiation (Dalton & Barton, 2014). The viral CMV and SV40 promoters were chosen for use in this study based upon their common application in mammalian systems, and unique transcriptional activities. CMV is a very strong promoter, whereas SV40 is considered a much weaker promoter and gives much lower gene expression based upon previous studies with ectopic gene expression in mammalian cells (Qin et al. 2010). SV40 also has a sequence that is commonly used as a mammalian terminator sequence that provide both polyadenylation and termination (Gil & Proudfoot, 1987; Goodwin & Rottman, 1992; Lanoix & Acheson, 1988; Schek, Cooke, & Alwine, 1992). The terminator, a sequence-based element, defines the end of the transcript, creating a free 3' end and initiates release of the mRNA from the transcriptional machinery. This terminator SV40 region is frequently used in mammalian expression plasmids. This was again used in these studies regardless of whether the CMV or SV40 promoter sequence was used upstream of the gene of interest. Cloning of the promoters into the tagged vectors was achieved with PCR amplification with specifically designed primers to initially amplify the promoter sequences from other vectors and include the required restriction sites for cloning into the tag containing vectors (*See Section 3.2.1*). Following gel electrophoresis, the amplified fragments were purified and digested with enzymes *BsiWI* and *KasI*, inserting the promoter regions inside the MCS before the tag gene. A digest screen

provided evidence that the insertion was successful through the detection of fragmented bands on the gel (Figure 3.7). Confirmation of the bands meant that cloning of the fluorescent reporters into the vectors could then be undertaken.

For the two constitutive promoters investigated in this study, each one was assessed to drive gene expression of either the eGFP (or later, destabilized eGFP) or mCherry reporter gene. As described previously, fluorescent reporters are used as a means of following gene expression and allow for protein visualization in sub-cellular and cellular localizations which can be followed in real time in live cells. The dynamics of the two fluorescently tagged markers can be compared when observed in transfected cells under a microscope when driven by the same or different promoters in the presence of absence of transcription factor binding sites upstream of the promoters. The fluorescent marker genes were cloned into the vectors containing tag sequences plus an individual promoter, either CMV or SV40. Alongside these digests, vectors without tags, only inserts of promoter sequences and the reporter genes, were created. Doing this provides a control group of plasmids for future analyses. Using enzymes *HindIII* and *XbaI* together, the tag sequence was removed, and a fluorescent marker replaced in its position. This was done to establish if the tag sequence may have any impact on the general protein expression, and in cases where a tag is simply unnecessary for a set of experiments.

After successfully generating the vectors containing the desired promoters, reporter genes and tags, the GFP constructs were transfected into CHO-S mammalian cells and observed under a microscope for fluorescence intensity to confirm expression of the reporter genes. The observed fluorescence, and intensity, confirmed successful construction of the plasmids (data not shown). The level of fluorescence was consistently high across all the tagged constructs, with the exception of the V5 tag constructs. Plasmid constructs with a V5 tag were visibly less intense, an indication of lower GFP expression. This could reflect an impact on transfection efficiency or an impact of the V5 tag on protein synthesis or degradation rates. Other probable causes may be directly related to the protein structure and folding being impacted or the tag somehow interfering with the ability of the protein to fluoresce (McLean et al., 2001). The reason for the lower fluorescence from the V5 tagged constructs remains unknown but the aspects described could be robustly investigated in the future.

Destabilized GFP (or dGFP) is an alternative to the eGFP fluorescent reporter system. Its half-life of ~2 hours allows for rapid detection in cells but also with a high turnover rate, which was constructed using eGFP fusion with a peptide signal sequence (Lui et al., 1998 and Rogers et al., 1986). The dGFP has been implemented to investigate protein degradation and production in real time when the cell perceives stressful environments. This advantage of using the short-lived fluorescent reporter in this study as it pertains to stress which is time-dependant, is that it can allow changes in gene expression to be more readily detected. eGFP, with a longer half-life, will build up over time and therefore to observe increases and decreases in expression requires a longer period of time than with the dGFP. However, some disadvantages include significant loss of fluorescence intensity or signal in the destabilized version of GFP. For the observations of this experimental design, the advantages were considered to outweigh the disadvantages in using the dGFP. The dGFP molecule was therefore also cloned into all the promoter, tag and TF sequence plasmids, replacing the existing eGFP reporter. The cloning of the dGFP into the vectors proved more challenging than anticipated, hence the numerous gel analyses for the selected constructs presented in the *Results* section. Eventually the dGFP was cloned downstream of both the CMV and SV40 promoters to compare protein expression intensity (Figure 3.16) from these different promoters.

The initial work looked to validate the generation of the constructs and confirm these gave suitable expression. All constructs were transfected into CHO-S cells and harvested at 24-hour and 72-hours after transfection. Following harvest, the cells were processed through a flow cytometer to measure fluorescence intensities of protein products. At 24 hours, the intensity of eGFP and dGFP was overwhelming any differences in the promoters. As previous reported, CMV is a much stronger promoter than SV40, but this was not apparent in the harvested cells at this time. However, at 72-hours, the cells projected more expected intensities (Figure 3.16) for the expected promoter activity level.

The flow cytometry data can be presented in several different ways and the mean fluorescence intensity (MFI) can be presented in multiple ways also. The arithmetic mean is the standard mean, number of events in each fluorescent event is divided by the total number of events. As

fluorescence intensity increases logarithmically, the arithmetic mean becomes an inadequate measurement to generalize a set of events. To combat this, the geometric mean is used to compensate for the log-normal behaviour of the flow cytometry data (Figure 3.16 B & 3.17 B). The fluorescence intensity from eGFP and dGFP construct transfection at both 24- and 72- hour time points were as expected when considering the reported promoter strengthens and the half-life of the two GFPs. Due to the lower half-life of the destabilized GFP protein, it should exhibit lower expression levels over time in comparison to the enhanced GFP. The dGFP inherently has a faster turnover rate, where it will degrade faster than its counterpart eGFP inside the cell. At the 72-hour time point, the geometric mean fluorescence for the CMV-dGFP showed an intensity geometric mean around 40 units whilst for the SV40-dGFP there was a lower intensity of ~6 units. This indicates that the different promoters did give different expressions as expected from their reported strengthens with CMV being a stronger promoter than the SV40. Similarly, comparing the eGFP constructs with dGFP, the dGFP fluorescence was lower from the respective plasmids suggesting that the turnover characteristic of the destabilized protein results in lower fluorescence than observed from the eGFP.

In the flow cytometry analysis 10,000 events (or individual cells) were measured for each sample and the fluorescence intensity was plotted as a histogram. The values were noted and graphed for all the constructs analyzed (Figure 3.18), where M1, M2 & M3 values denote the percentage of cells whose fluorescence was above the three different thresholds set values from low (M1) to high (M3), respectively. This data confirmed that the dGFP fluorescence was reduced with respect to the eGFP at 72 hours (Figure 3.18). These observations led us to select the SV40 promoter sequence as that to use to investigate the impact of TF binding sequences on gene expression. As the CMV expression was exhibiting much higher expression levels, masking any TF activity, thus observing activation with the presence of the TF binding site sequence would be much more difficult than from the lower baseline expression of the SV40 promoter. Therefore, only SV40 constructs with TF binding sequence 5’ to the promoter sequence were further developed, generated and investigated.

Several transcription factor sequence motifs were selected for investigation that were known to bind to transcription factor binding proteins that control response to stresses that CHO cells are

likely to experience during production of biotherapeutic proteins. The motifs selected for expression include the ERSE, SRE, FASN, PPAR, & ChoRE sequences that are described in the *Introduction* section of this thesis. ERSE, the ER Stress Element, affects biosynthetic pathways in the ER, where malfunctions in the cell can cause serious damage to proteins. The ERSE binds TFs that signal and activate a respond to ER stress and play a role in responses to ER calcium depletion and protein glycosylation in the ER. Du et al. (2013) described three important sequences of DNA that all enhance gene expression in response to ER stress (Figure 4.1). These are termed the ERSE, URPE & ACGT core element. The work presented in this thesis describes attempts to use the described sequences (Figure 4.1) as a whole, and as individual elements (ERSE, UPRE and ACGT on their own) to modulate reporter gene expression when placed upstream of the SV40 promoter. Another experimental design was to investigate if there was any transcriptional activity from these sequences without the SV40 promoter region. As reported by Du et al., the TF motif contains several repeated regions, which became evident that they play a critical role in the transcription of the downstream gene sequence during ER stress.

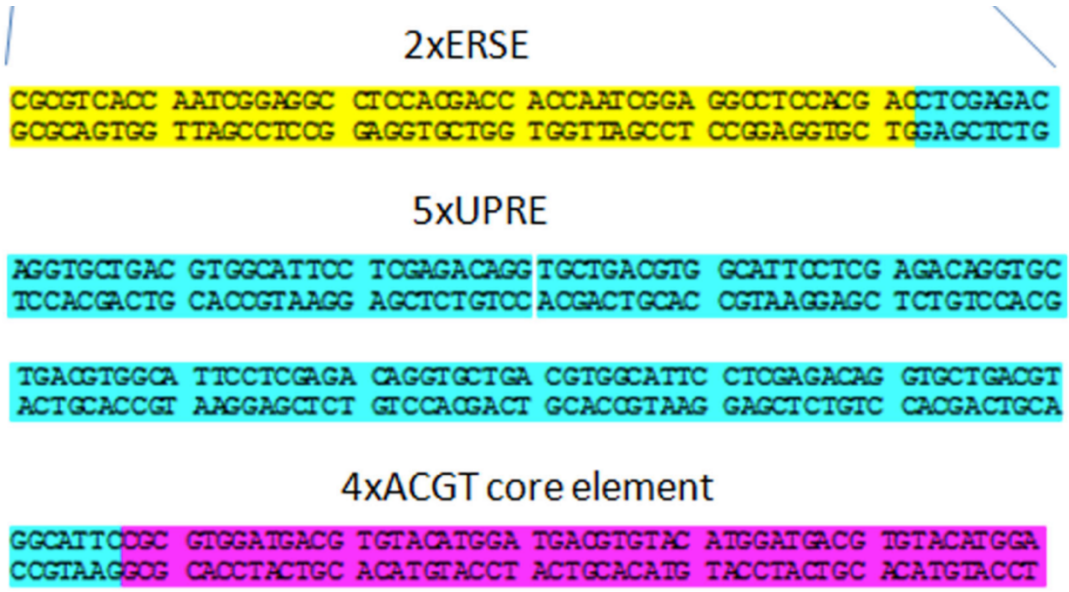


Figure 4.1: Construction of the TF motif used in Du et al. (2013) studies reporting induction of gene expression during ER stress. The UPR system contains three transcription elements ERSE, UPRE and a core element, that comprise of multiple repeats of the consensus sequences of the binding motifs. Image derived from Du et al, 2013.

The cloning protocol for introduction of the regulatory TF binding sites was designed to integrate restriction enzyme sites between the individual elements as shown in Figure 4.2. In this way, restriction digests can be implemented to generate any final desired construct. The TF motif was cloned into the desired plasmid construct using three different restriction site methods to produce variations of possible gene expressions. Figure 4.2 depicts one of the ERSE plasmid constructs, containing all repeated sequences from Du et al. (2013), with integrated selected restriction sites. The authors also inferred that the spacing between these segments of translational DNA has some elasticity, making the choice of additional nucleic acid addition possible and non-consequential.

The SRE sequence is known to help regulate the expression of those genes involved in cholesterol biosynthesis and lipid homeostasis. Numerous research articles provided only general sequence variations of spacers, CCAAT boxes and SRE (sterol regulatory element) sequences for the respective TF motif. Consequently, the design of the SRE construct used in the studies here originated from researchers who discovered a conservative sequence to hamster genes. Inoue et al. (1998) identified repeated sequences that potentially affected the initiation of transcription in the regulatory elements. The general pattern of the motif discovered proceeds as such:

SRE-1 ------- spacer ------- **CCAAT** ------ spacer ------ **SRE-2** ------ spacer ------ **CCAAT**

Irregularities in sterol regulation causes a stress pathway to be activated and genes to be initiated for transcription. The SRE1/2, CB1, and Sp1 were the potential regulatory elements influencing transcription. The hamster sequence described above and in Figure 4.3 was used directly to clone the TF motif pattern and place this upstream of the SV40 promoter sequence.

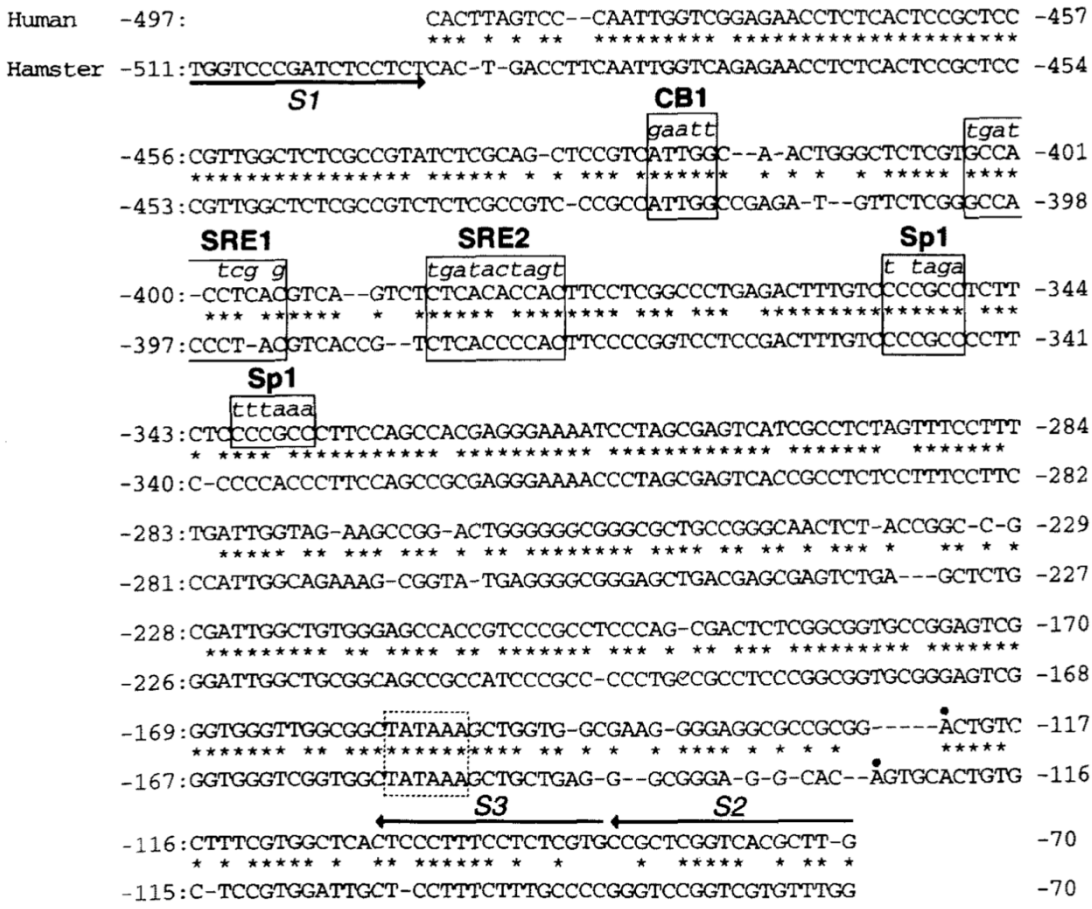


Figure 4.3 Human and hamster conserved transcription factor binding sequences found to influence the sterol stress regulatory pathway. The boxed sequences represent the essential regulatory regions for transcription initiation. Figure derived from Inoue et al. 1998.

The FASN transcription factor sequence motifs also follows a similar pattern to the SRE motif, containing multiple SRE repeats and pattern of regulation. FASN is known to regulate fatty acid synthesis. The transcription initiation sites are designated in Figure 4.4, highlighting the SRE binding sites and boxes critical for control of gene expression. This whole segment of DNA information was derived from Han et al. (2018) was used and cloned. The designed DNA sequences contained the following restriction sites:

ACGCGT (MluI)

ACCGGT (AgeI)

GAGCTC (SacI)

TTAATTAA (PacI)

CTTAAG (AflII)

GGCGCGCC (AscI)

The sites were embedded within the sequence from Figure 4.4 with consideration that spacers have an effect on the expression of transcription, therefore the sites were dispersed individually between the elemental regulatory sites. This insertion of the restriction enzymes sites allows flexibility of cloning any desirable sequence. For instance, a future examination can correlate the effects of number of repeats of the sequence elements on the gain/loss of function of cells under fatty acid deprivation or stressful conditions.

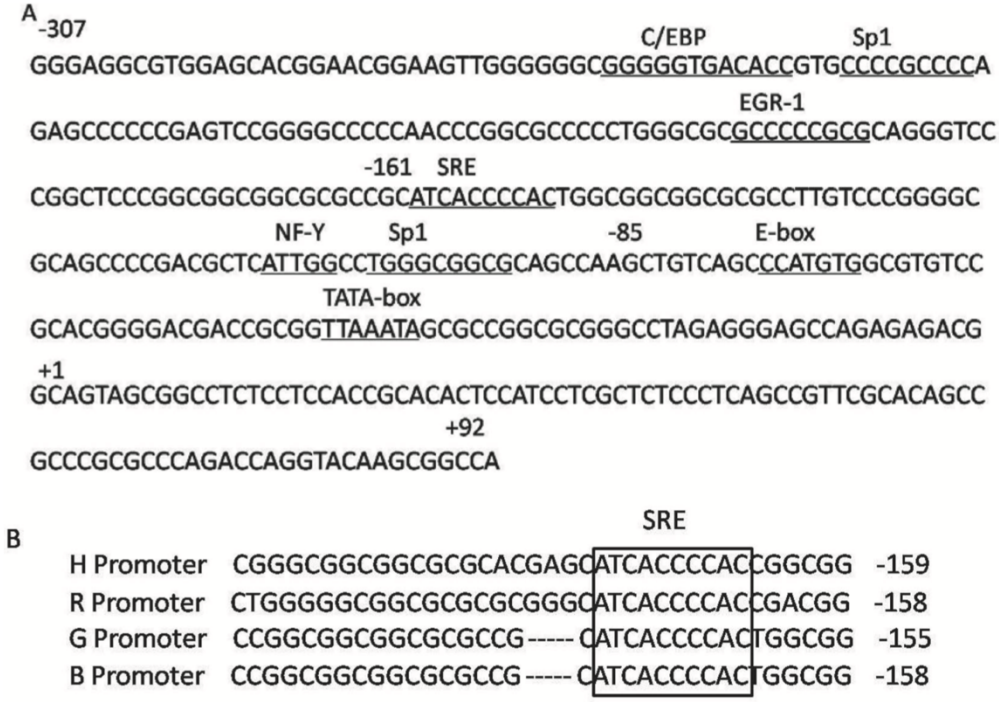


Figure 4.4 Nucleic acid transcription binding sequences used to develop reporters of FASN motif activity. (A) The bovine bioinformatics revealed the FASN sequence contains binding sites for the TF. (B) Alignment of SRE sites of different species, indicating similar across species. Derived from Han et al. 2018.

Fatty acid metabolism has another regulator of gene expression which was found to be done through a nuclear receptor, PPAR. Tzeng and colleagues found a PPAR responsive sequence for

gene expression containing two AGGTCA-like sequences (Tzeng et al. 2015). A minimal promoter region was inserted by the authors to enhance the expression. Along with a minimal promoter sequence obtained from the article, the establishment of the TF motif construct was completed and is shown in Figure 4.5. For experimental purposes, the restriction sites were inserted between each repeated PPAR segment to enable cutting between the sites for analysis of potential intensity comparisons.

Similarly, with ChoRE transcription factor motifs, several repeats were included. The sequences associated with the generation of ChoRE sequence repeats of an inverted CCAAT box and the ChoRE specific regions, as indicated from Figure 4.6. Again, the ChoRE sequence was manipulated to have restriction sites within the sequences. Figure 4.6 shows the nucleic sequence for the ChoRE TF motif that was used for cloning in this study. Upon further investigation however, a *MluI* site was identified within the sequence, therefore a single 'A' nucleic acid was added to counteract the formation of another restriction site. Figure 4.7 shows the enlarged TF motif with the selected sites imbedded between each segment of the ChoRE consensus sequences. Many scholars have noted that spacers between critical transcriptional sites can play an enhancement role in the expression of downstream genes (Hook-Barnard, I. G., & Hinton, D. M. (2009), and Typas, A. and Hengge, R. (2006)) in ChoRE. This provides some reassurance that the several additions of nucleic acids will not affect the ability of the TF to bind the appropriate proteins and activate transcription, but it cannot be ruled out that these additions negatively impact the motifs activity or ability to activate transcription of downstream genes. In addition to the triple ChoRE sequences, a minimal promoter was inserted adjoining the ChoRE. The minimal promoter was taken from Tzeng et al. 2015, the PPAR algorithm.

Figure 4.6 TF motif sequence required for mediating the stimulatory effect of ChoRE. The components of the designed motif incorporated this pattern of nucleic acids for the production and insertion into one of our existing plasmid maps.

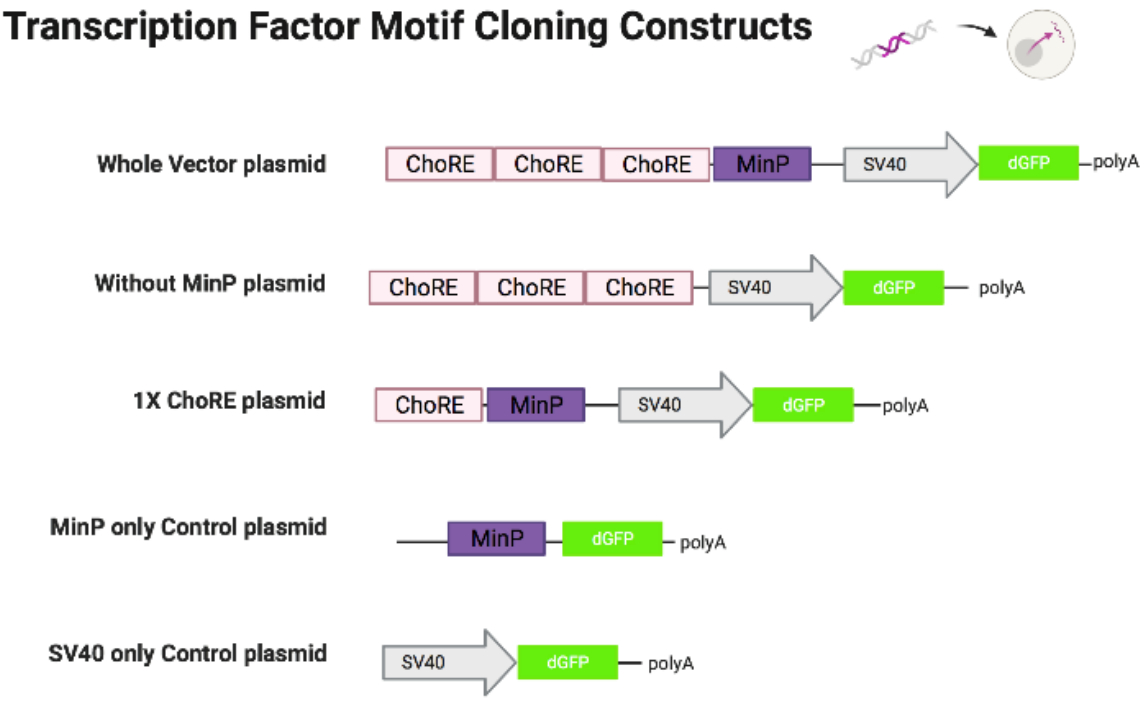


Figure 4.8 Schematic of the plasmid variations of each TF motif insertion. With the use of restriction enzyme placement, the above fragments within the MCS were achieved for the transfection experimental procedure. Along with two control sequences.

To reiterate, the focus of the thesis was to generate constructs that respond to cellular stress through networks of visualizable markers, such as the GFP, that report on the degree of stress observed and timing of these events. Theoretically, cells transfected into spent media will respond to a lack of nutrients and presence of toxic metabolites by exhibiting slowed growth and the reporters should also respond by showing increased GFP expression. After 24 hours the cultures were analysed for cell number, culture viability and GFP fluorescence by flow cytometry. Unfortunately, the results showed low viable cell numbers suggesting that the spent media or transfection process had impacted the cells to an apoptotic degree. When the fluorescence was measured on the flow

cytometer, a minor but observable difference in fluorescence was detected between the different constructs at the 24-hour time point. Expression in those constructs containing the complete transcription factor motif sequences was higher than the partial TF motif constructs. This preliminary data confirms that the complete sequence is required to give maximal transcriptional activity and should be utilised in future studies. Feasible causes for the low cell count and low fluorescence are the sudden shift from high nutrient to nutrient depletion conditions in the spent media. Changes in the experimental design are therefore likely to be necessary to give higher expression levels at the initial stages of the transfection. Longer incubation time and time points beyond 24 h, or indeed shorter time periods, should also be investigated alongside introducing more gradual changes in stress levels.

Subsequently, and in the remaining time available for the project, the ERSE plasmid constructs (Figure 3.24) along with the appropriate negative and positive controls were transfected into CHO-S cells, without altering any media environments to impart a stress. The transfected cells were then analyzed by flow cytometry to determine the fluorescence from each of the constructs. Based on the data 24-hours post-transfection, it was apparent that the ERSE whole sequence constructs gave greater reporter gene expression than those with specific elements or sequences alone from the ERSE motif (see Table 3.3). In addition, those constructs with the minimal promoter activity (AGCT box) showed higher expression levels compared to those where this was not present. However, the minimal promoter region alone did not support gene expression anywhere near that from the ERSE TF motif (ERSE constructs “A”). With only a 24-hour time point, the data should be carefully interpreted, and provides evidence that the different elements of the TF motifs can influence subsequent reporter gene activity. However, there was insufficient time in the project to fully evaluate these reporter constructs, particularly whether they were responsive to stresses imposed upon the cell. Therefore, future steps would be necessary to induce appropriate stresses, to different degrees, in cells that would activate the ERSE/UPR system, with dGFP as the reporter and determine if a response was observed and if the response was proportional to the stress induced. Such stresses can be artificially induced by the addition of chemical agents known to induce such stresses.

5 CONCLUSIONS AND FUTURE WORK

Despite the variety of mammalian cell lines available for protein production, CHO cells remain the workhorse of biopharmaceutical production since the approval of the first CHO derived recombinant protein in 1986. CHO cells are involved in the production of over 70% of therapeutic recombinant proteins, most of which are monoclonal antibodies (Jayapal, Wlaschin, Hu, & Yap, 2007). In 2017, the global revenue from monoclonal antibodies was over US $90 billion, providing a strong incentive for the continued development of cell line engineering technologies. The work presented in this thesis describes attempts to generate reporter systems that could report on the stresses perceived by CHO cells during biopharmaceutical production. The project was significantly impacted by the COVID-19 pandemic and access to the laboratory. Nevertheless, new TF reporter constructs were designed and generated, and some initial validation work undertaken. Further work would initially focus on establishing that the reporter constructs did respond to the stresses to which they are designed to respond to and that these were able subsequently to reflect the levels of stress observed in CHO cells during bioproduction. Once this was established, it would be appropriate to develop CHO cell lines that stably expressed the reporter systems into which target biotherapeutic gene vectors could be transfected to determine if different proteins and variants imposed different stresses on the CHO cell factory. The learnings from these studies can then be used to design processes and cell engineering strategies to limit or overcome such stresses and limitations. A limitation of this approach is that only those stresses for which the reporter systems are programmed to respond by inclusion of appropriate TF sequences can be monitored. Further, the sensitivity of such systems is untested and therefore whether such reporter systems are sensitive, and respond rapidly, enough to highlight and identify perceived stress remains to be answered. Regardless, the application of such reporter systems is likely to shed new light on those stresses that CHO cell systems experience, and when, during biopharmaceutical production which can feed into development of new strategies to avoid these.

6 BIBLIOGRAPHY

Aggarwal R. S. (2014). What's fueling the biotech engine-2012 to 2013. *Nature biotechnology*, *32*(1), 32–39. https://doi.org/10.1038/nbt.2794

Alberts B, Johnson A, Lewis J, et al. (2002). Isolating, Cloning, and Sequencing DNA. *Molecular Biology of the Cell. 4th edition. New York: Garland Science.* Available from: https://www.ncbi.nlm.nih.gov/books/NBK26837/

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *Journal of bacteriology*, *62*(3), 293–300. https://doi.org/10.1128/JB.62.3.293-300.1951

Borkotoky, S., & Murali, A. (2018). The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *International journal of biological macromolecules*, *118*(Pt A), 49–56. https://doi.org/10.1016/j.ijbiomac.2018.05.198

Borukhov, S., & Nudler, E. (2008). RNA polymerase: the vehicle of transcription. *Trends in microbiology*, *16*(3), 126–134. https://doi.org/10.1016/j.tim.2007.12.006

Budge, J. D., Knight, T. J., Povey, J., Roobol, J., Brown, I. R., Singh, G., Dean, A., Turner, S., Jaques, C. M., Young, R. J., Racher, A. J., & Smales, C. M. (2020). Engineering of Chinese hamster ovary cell lipid metabolism results in an expanded ER and enhanced recombinant biotherapeutic protein production. *Metabolic engineering*, *57*, 203–216. https://doi.org/10.1016/j.ymben.2019.11.007

Büssow, Konrad. (2015). Stable Mammalian Producer Cell Lines for Structural Biology. *Current Opinion in Structural Biology*, *vol. 32, pp. 81–90.* doi: 10.1016/j.sbi.2015.03.002.

Christofk, H. R., Vander Heiden, M. G., Harris, M. H., Ramanathan, A., Gerszten, R. E., Wei, R., Fleming, M. D., Schreiber, S. L., & Cantley, L. C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumor growth. *Nature*, *452*(7184), 230–233. https://doi.org/10.1038/nature06734

Clancy, Suzanne. (2008). DNA Transcription | Learn Science at Scitable. *Nature*. www.nature.com/scitable/topicpage/dna-transcription-426/#

Da Fonseca, Anna Carolina Carvalho, et al. (2021). The Multiple Functions of the Co-Chaperone Stress Inducible Protein 1. *Cytokine & Growth Factor Reviews*, *vol. 57, pp. 73–84*, 10.1016/j.cytogfr.2020.06.003.

Dumont, J., Euwart, D., Mei, B., Estes, S., & Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical reviews in biotechnology*, *36*(6), 1110–1122. https://doi.org/10.3109/07388551.2015.1084266

Fujimoto, S., & Ike, Y. (2001). pAM401-based shuttle vectors that enable overexpression of promoter-less genes and one-step purification of tag fusion proteins directly from Enterococcus faecalis. *Applied and environmental microbiology*, *67*(3), 1262–1267. https://doi.org/10.1128/AEM.67.3.1262-1267.2001

Estes, S., & Melville, M. (2014). Mammalian cell line developments in speed and efficiency. *Advances in biochemical engineering/biotechnology*, *139*, 11–33. https://doi.org/10.1007/10_2013_260

González-Gualda, E., Baker, A. G., Fruk, L., & Muñoz-Espín, D. (2021). A guide to assessing cellular senescence in vitro and in vivo. *The FEBS journal*, *288*(1), 56–80. https://doi.org/10.1111/febs.15570

Guéroult-Bellone, M., Nitta, K. R., Kari, W., Jacox, E., Beulé Dauzat, R., Vincentelli, R., Lemaire, P. (2017). *Spacer sequences separating transcription factor binding motifs set enhancer quality and strength*. https://doi.org/10.1101/098830

He, L., Binari, R., Huang, J., Falo-Sanjuan, J., & Perrimon, N. (2019). In vivo study of gene expression with an enhanced dual-color fluorescent transcriptional timer. *ELife*, *8*, e46181. https://doi.org/10.7554/eLife.46181

Hoffmann, F., & Rinas, U. (2004). Stress induced by recombinant protein production in Escherichia coli. *Advances in biochemical engineering/biotechnology*, *89*, 73–92. https://doi.org/10.1007/b93994

Hook-Barnard, I. G., & Hinton, D. M. (2009). The promoter spacer influences transcription initiation via sigma70 region 1.1 of Escherichia coli RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(3), 737–742. https://doi.org/10.1073/pnas.0808133106

Hunter, M., Yuan, P., Vavilala, D., & Fox, M. (2019). Optimization of protein expression in mammalian cells. *Current Protocols in Protein Science*, *95,* e77. doi: 10.1002/cpps.77

Hussain, H., Patel, T., Ozanne, A., Vito, D., Ellis, M., Hinchliffe, M., Humphreys, D. P., Stephens, P. E., Sweeney, B., White, J., Dickson, A. J., & Smales, C. M. (2021). A comparative analysis of recombinant Fab and full-length antibody production in Chinese hamster ovary cells. *Biotechnology and bioengineering*, *118*(12), 4815–4828. https://doi.org/10.1002/bit.27944

Iacobini, C., Vitale, M., Pugliese, G., & Menini, S. (2021). Normalizing HIF-1α Signaling Improves Cellular Glucose Metabolism and Blocks the Pathological Pathways of Hyperglycemic Damage. *Biomedicines*, *9*(9), 1139. https://doi.org/10.3390/biomedicines9091139

Jayapal, K. P., Wlaschin, K. F., Hu, W. S., & Yap, M. G. S. (2007). Recombinant protein therapeutics from CHO Cells - 20 years and counting. *Chemical Engineering Progress*, *103*(10), 40-47.

Kamiyama, D., Sekine, S., Barsi-Rhyne, B. *et al.* (2016). Versatile protein tagging in cells with split fluorescent protein. *Nat Commun* **7,** *11046*. https://doi.org/10.1038/ncomms11046.

Kazilek. (2010). Cutting DNA | Ask a Biologist. *Asu.edu*. Derived from: askabiologist.asu.edu/restriction-enzymes.

Kilpinen, H., Waszak, S. M., Gschwind, A. R., Raghav, S. K., Witwicki, R. M., Orioli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N. I., Yurovsky, A., Lappalainen, T., Romano-Palumbo, L., Planchon, A., Bielser, D., Bryois, J., Padioleau, I., Udin, G., Thurnheer, S., Hacker, D., … Dermitzakis, E. T. (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science (New York, N.Y.)*, *342*(6159), 744–747. https://doi.org/10.1126/science.1242463

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. https://doi.org/10.1007/s00253-011-3758-5

Kokame, K., Kato, H. and Miyata, T. (2001). Identification of ERSE-II, a New Cis-Acting Element Responsible for the ATF6-dependent Mammalian Unfolded Protein Response *. *Journal of Biological Chemistry*, 276(12), pp.9199–9205.

Kostylev, M., Otwell, A. E., Richardson, R. E., & Suzuki, Y. (2015). Cloning Should Be Simple: Escherichia coli DH5α-Mediated Assembly of Multiple DNA Fragments with Short End Homologies. *PloS one*, *10*(9), e0137466. https://doi.org/10.1371/journal.pone.0137466

Liu, C., Dalby, B., Chen, W., Kilzer, J. M., & Chiou, H. C. (2008). Transient transfection factors for high-level recombinant protein production in suspension cultured mammalian cells. *Molecular biotechnology*, *39*(2), 141–153. https://doi.org/10.1007/s12033-008-9051-x

Liu, J., Huang, J., Fan, K. W., Jiang, Y., Zhong, Y., Sun, Z., & Chen, F. (2010). Production potential of Chlorella zofingienesis as a feedstock for biodiesel. *Bioresource technology*, *101*(22), 8658–8663. https://doi.org/10.1016/j.biortech.2010.05.082

Liu, L., Sommermeyer, D., Cabanov, A., Kosasih, P., Hill, T., & Riddell, S. R. (2016). Inclusion of Strep-tag II in design of antigen receptors for T-cell immunotherapy. *Nature biotechnology*, *34*(4), 430–434. https://doi.org/10.1038/nbt.3461

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C. C., & Kain, S. R. (1998). Generation of destabilized green fluorescent protein as a transcription reporter. *The Journal of biological chemistry*, *273*(52), 34970–34975. https://doi.org/10.1074/jbc.273.52.34970

Maldonado-Agurto, R., & Dickson, A. J. (2018). Multiplexed Digital mRNA Expression Analysis Profiles System-Wide Changes in mRNA Abundance and Responsiveness of UPR-Specific Gene Expression Changes During Batch Culture of Recombinant Chinese Hamster Ovary Cells. *Biotechnology journal*, *13*(3), e1700429. https://doi.org/10.1002/biot.201700429

McLean, Pamela & Kawamata, Hibiki & Hyman, Bradley. (2001). Alpha-synuclein-enhanced green fluorescent protein fusion proteins form proteasome sensitive inclusions in primary neurons. Neuroscience. 104. 901-12. 10.1016/S0306-4522(01)00113-0.

Orellana, C. A., Marcellin, E., Munro, T., Gary, P. P., & Nielsen, L. K. (2015). Multi-omics approach for comparative studies of monoclonal antibody producing CHO cells. *BMC Proceedings*, *9*(Suppl 9), O8. https://doi.org/10.1186/1753-6561-9-S9-O8

PcDNA™3.1/Hygro(+) Mammalian Expression Vector. *Thermofisher.* www.thermofisher.com/order/catalog/product/V87020#/V87020.

Pete Corish, Chris Tyler-Smith. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Engineering, Design and Selection,* Volume 12, Issue 12, Pages 1035–1040. https://doi.org/10.1093/protein/12.12.1035

Phillips, Theresa. (2008). Noncoding RNA and Gene Expression | Learn Science at Scitable. *Nature.* Accessed through: www.nature.com/scitable/topicpage/regulation-of-transcription-and-gene-expression-in-1086/

Qin JY, Zhang L, Clift KL, Hulur I, Xiang AP, Ren B-Z, et al. (2010) Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. PLoS ONE 5(5): e10611. https://doi.org/10.1371/journal.pone.0010611

Reid, D. W., Chen, Q., Tay, A. S., Shenolikar, S., & Nicchitta, C. V. (2014). The unfolded protein response triggers selective mRNA release from the endoplasmic reticulum. *Cell*, *158*(6), 1362–1374. https://doi.org/10.1016/j.cell.2014.08.012

Remenyi, A., Scholer, H., & Wilmanns, M. (2004). Combinatorial control of gene expression. *Nature Structural and Molecular Biology* **11**, 812–815. doi:10.1038/nsmb820

Rogers, S., Wells, R., & Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science (New York, N.Y.)*, *234*(4774), 364–368. https://doi.org/10.1126/science.2876518

Roy, G., Zhang, S., Li, L., Higham, E., Wu, H., Marelli, M., & Bowen, M. A. (2017). Development of a fluorescent reporter system for monitoring ER stress in Chinese hamster ovary cells and its application for therapeutic protein production. *PloS one*, *12*(8), e0183694. https://doi.org/10.1371/journal.pone.0183694

Swiech, K., Picanço-Castro, V., & Covas, D. T. (2012). Human cells: new platform for recombinant therapeutic protein production. *Protein expression and purification*, *84*(1), 147–153. https://doi.org/10.1016/j.pep.2012.04.023

Tumanov, Sergey & Bulusu, Vinay & Kamphorst, Jurre. (2015). Analysis of Fatty Acid Metabolism Using Stable Isotope Tracers and Mass Spectrometry. Methods in Enzymology. 561. 10.1016/bs.mie.2015.05.017

Typas, A. and Hengge, R. (2006). Role of the spacer between the −35 and −10 regions in σ^s promoter selectivity in *Escherichia coli*. Molecular Microbiology, 59: 1037-1051. https://doi.org/10.1111/j.1365-2958.2005.04998.x

Vander Heiden, M. G., Locasale, J. W., Swanson, K. D., Sharfi, H., Heffron, G. J., Amador-Noguez, D., Christofk, H. R., Wagner, G., Rabinowitz, J. D., Asara, J. M., & Cantley, L. C. (2010). Evidence for an alternative glycolytic pathway in rapidly proliferating cells. *Science (New York, N.Y.)*, *329*(5998), 1492–1499. https://doi.org/10.1126/science.1188015

Vcelar, S., Jadhav, V., Melcher, M., Auer, N., Hrdina, A., Sagmeister, R., Heffner, K., Puklowski, A., Betenbaugh, M., Wenger, T., Leisch, F., Baumann, M., & Borth, N. (2018). Karyotype variation of CHO host cell lines over time in culture characterized by chromosome counting and chromosome painting. *Biotechnology and bioengineering*, *115*(1), 165–173. https://doi.org/10.1002/bit.26453

Vcelar, S., Melcher, M., Auer, N., Hrdina, A., Puklowski, A., Leisch, F., Jadhav, V., Wenger, T., Baumann, M., & Borth, N. (2018). Changes in Chromosome Counts and Patterns in CHO Cell Lines upon Generation of Recombinant Cell Lines and Subcloning. *Biotechnology journal*, *13*(3), e1700495. https://doi.org/10.1002/biot.201700495

V5 Tag Antibodies. (n.d.). Retrieved from https://www.thermofisher.com/uk/en/home/life-science/antibodies/primary-antibodies/epitope-tag-antibodies/v5-tag-antibodies.html

Wang, W., Jia, Yl., Li, Yc. *et al.* (2017). Impact of different promoters, promoter mutation, and an enhancer on recombinant protein expression in CHO cells. *Sci Rep* **7,** 10416 https://doi.org/10.1038/s41598-017-10966-y

Wang, XY., Zhang, JH., Zhang, X. *et al.* (2016). Impact of Different Promoters on Episomal Vectors Harbouring Characteristic Motifs of Matrix Attachment Regions. *Sci Rep* **6,** 26446. https://doi.org/10.1038/srep26446

Wildeman A. G. (1988). Regulation of SV40 early gene expression. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, *66*(6), 567–577. https://doi.org/10.1139/o88-067

Yu F-X, Luo Y (2009). Tandem ChoRE and CCAAT Motifs and Associated Factors Regulate Txnip Expression in Response to Glucose or Adenosine- Containing Molecules. PLoS ONE 4(12): e8397. doi:10.1371/journal.pone.0008397

7 SUPPLEMENTARY DATA

Table 7.1 Cell count at 72-hour time point harvest. This table shows the viable cells/ mL and percent viability of each construct in CHO-S cell line.

Construct:	Viable cells/ mL (x10^6)	Percentage viability (%)
CMV dGFP	2.48	89.4
SV40 dGFP	3.49	89.6
dGFP only	2.70	91.9
CMV only	3.60	93.2
SV40 only	4.19	91.7
eGFP CMV	2.70	88.7
eGFP SV40	4.45	91.8

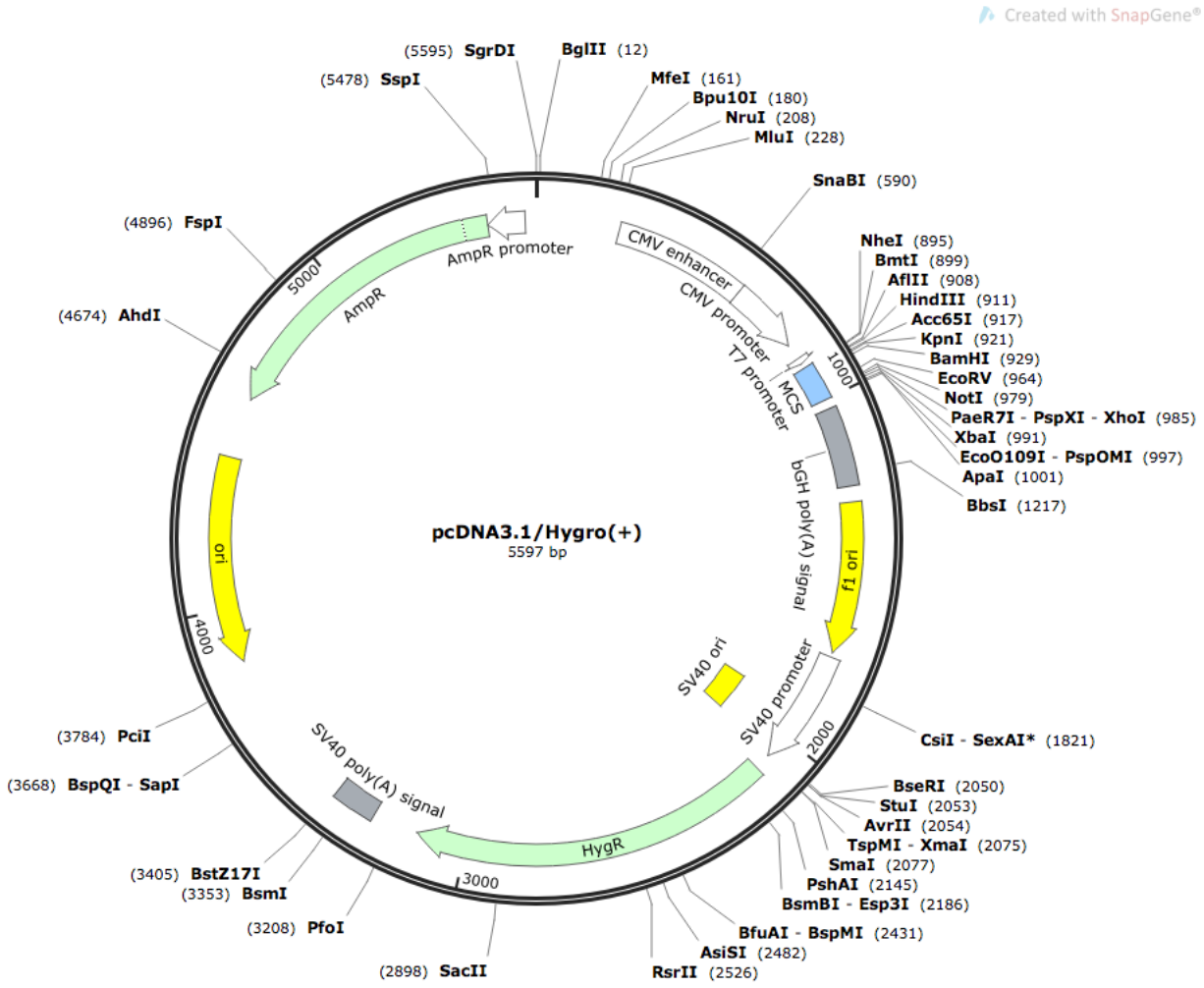


Figure 7.1 *pcDNA™3.1/Hygro(+)* plasmid derived from Addgene database

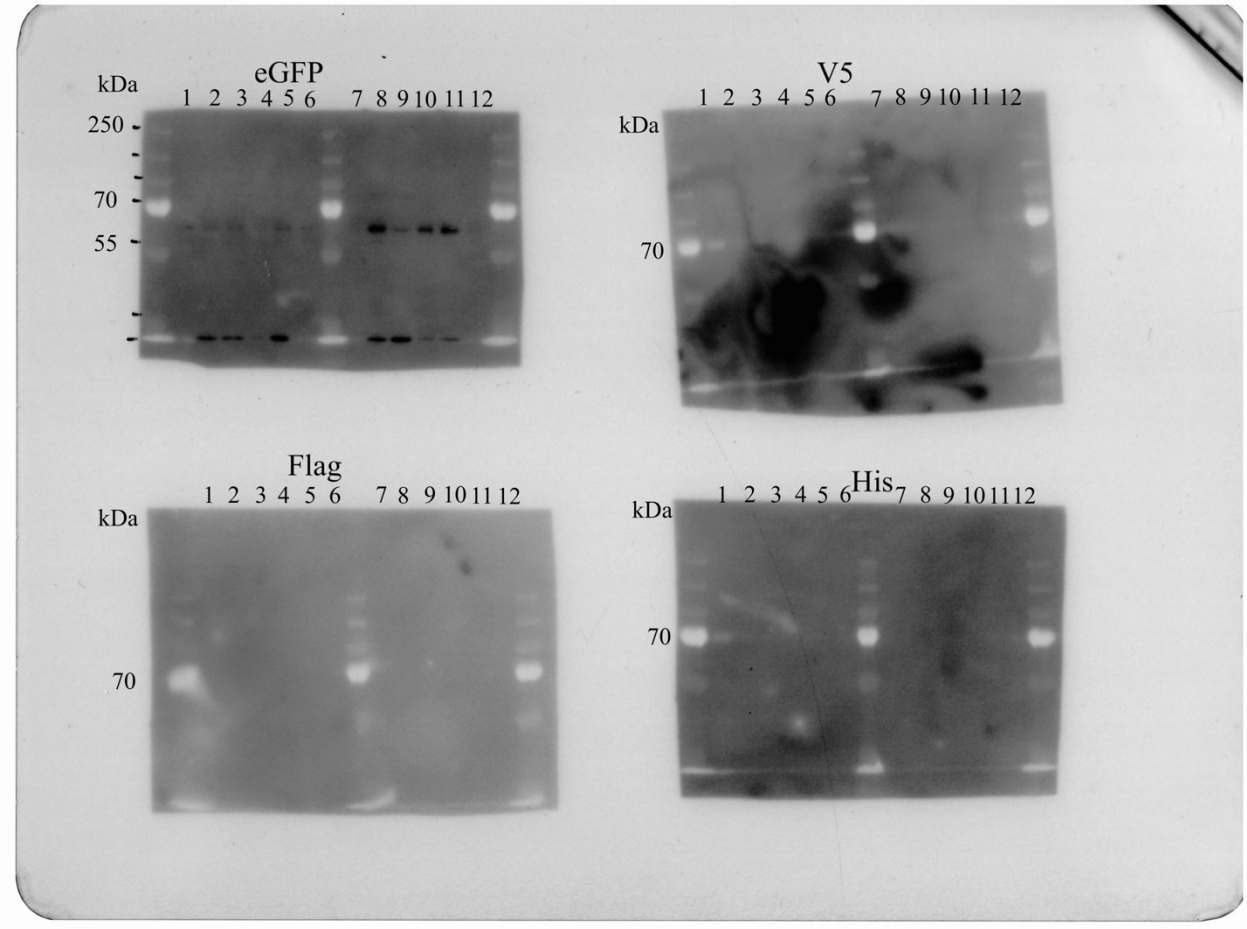


Figure 7.2 Western blot analysis for the detection of eGFP protein in V5, Strep, and His constructs. eGFP only shows proteins detected using a direct eGFP antibody for detection whereas anti-tag antibodies were used in the V5, Strep and His samples.

Legend for Figure's 3.13 and 3.14:

Legend	
1	CMV + V5
2	CMV + Flag
3	CMV + Strep
4	CMV + His
5	CMV (-)
6	CMV only
7	SV40 + V5
8	SV40 + Flag

9	SV40 + Strep
10	SV40 + His
11	SV40 + (-)
12	SV40 only

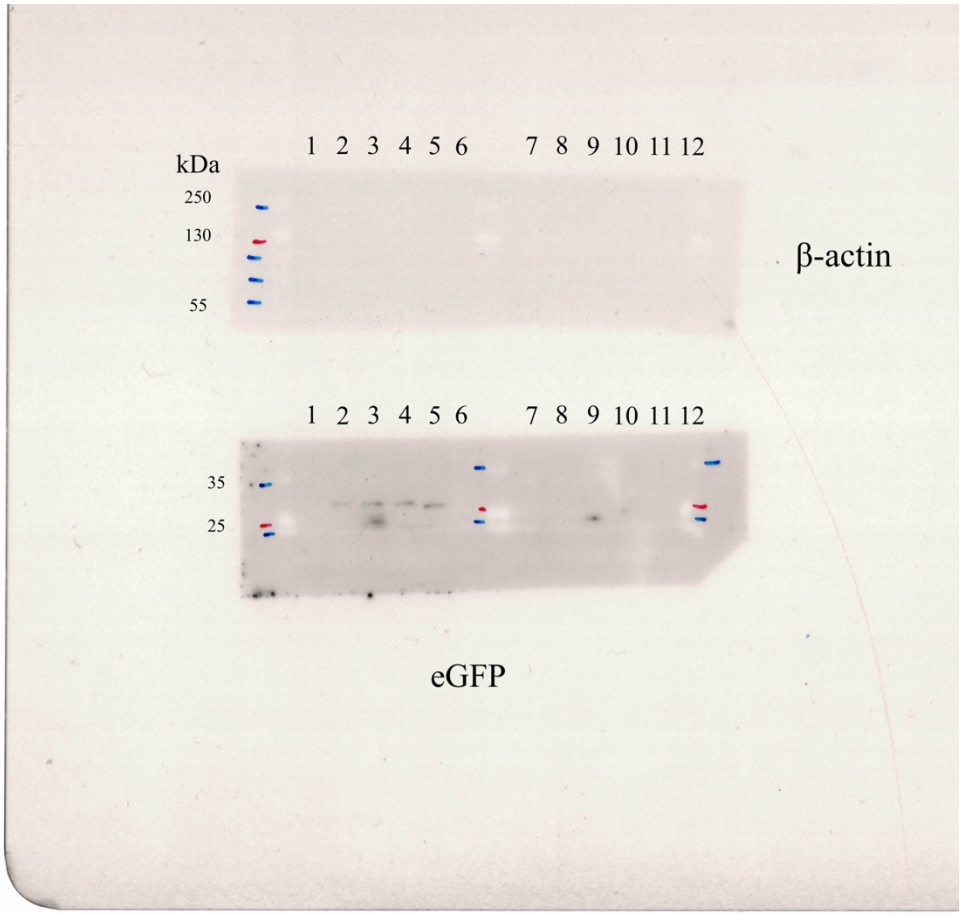


Figure 7.3 Repeat of blots described in *Section 3.12* using fresh buffer and antibody preparations.

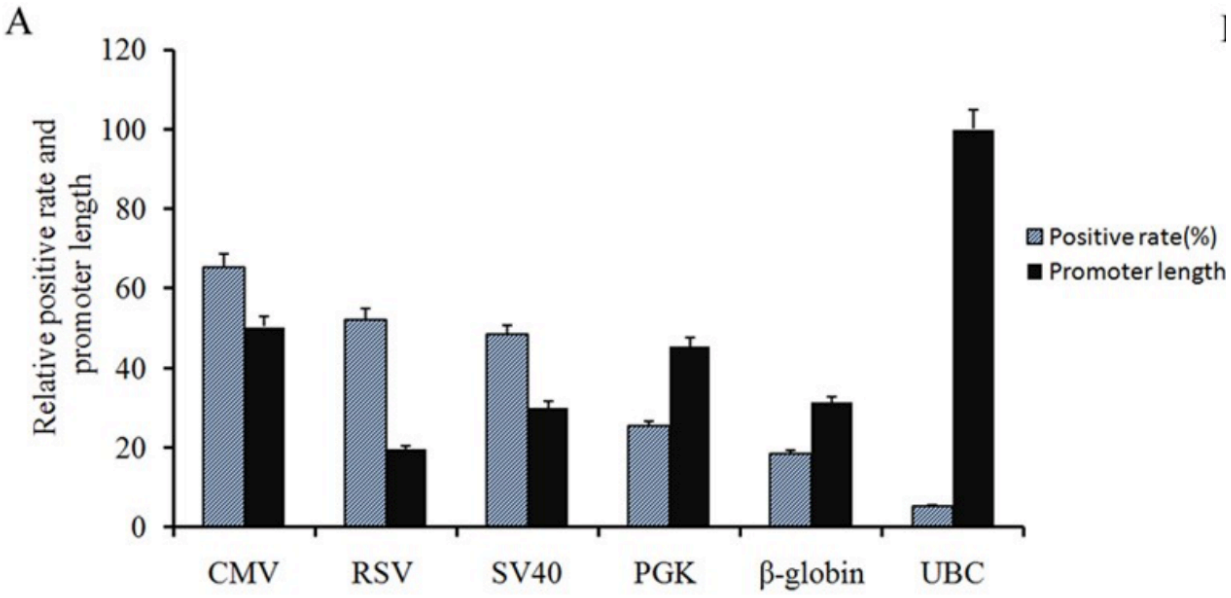


Figure 7.4 Positive recombinant protein expression rates, promoter strength and eGFP expression using CHO cells. Derived from Wang, XY., Zhang, JH., Zhang, X. *et al.* 2016.

Sequencing Data

>ERSE_A_forward-Construct_Forward_B02.ab1
NNNNNNNNNNNNNNNNATTGCATGAAGAANCTGCTTAGGNTTAGGCGTTTTGCGCTGCTTCGCGATGT
ACGGGCCAGATA
TACGCGTCGCGTCACCAATCGGAGGCCTCCACGACCACCAATCGGAGGCCTCCACGACACCGGTCTCG
AGACAGGTGCTG
ACGTGGCATTCCTCGAGACAGGTGCTGACGTGGCATTCCTCGAGACAGGTGCTGACGTGGCATTCCTC
GAGACAGGTGCT
GACGTGGCATTCCTCGAGACAGGTGCTGACGTGGCATTCGAGCTCCGCGTGGATGACGTGTACATGGA
TGACGTGTACAT
GGATGACGTGTACATGGATTAATTAAACTAGTCCAGTGGCGCGCCATCAGCCTCGTACGGTGTGTCAG
TTANGGTGTGGA
AAGTCCCCAGGCTCCCCNGNAGGNATAAGTATGCANAGCATGCATCTCAATTAGTCAGCAACCAGGTG
TGGAAAGTCCCC
ANGCTCCCCAGCANGNNNAAGTATGCAAAGNNTGCATCTCANTTAGTCAGCAANNNTANTCCCGCCC
CTAACTCCGCNNA
NNNCGNCGCTAACTCCNCCNAGTTCCGNNCATTCTNNGNNNNATGNNTGANNANNNTTTTTTTATTTN
TGNANANGNCGA
GGNCNCCTCTGCCTCTGANCTATTCCNNANNNANNGAGGAGGCTTTTTTGGAGGCNTNNGNNTNTGCN
NNNNNNNNTCTG
GNNAACTANNNAANNCACTGCTTACNNNNNNACNNNNNTNANNNNACTCANNNNNGNAGANCNANG
NNGNNTAANNNNNN
NNNAGANANGNNNANGNNNNNNNNNNTGGNNNNNNNNNNNNNTNNNNNNNCTGGNNNNNANNNAN
NNNNCNNNNNNTNNN
NNGNGCCNNNNANNNNNNNNNNNNNCNNNCTNNNNNNNNNNANCTNNNANNNNNNNNNNANCAN
CGNCNNNNNNNNN

>ERSE_B_forward-Construct_Forward_F02
NANGCTGGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACG
GGCCAGATATACGCGTCGCGTCACCAATCGGAGGCCTCCACGACCACCAATCGGAGGCCTCCACGACA
CCGGTCTCGAGACAGGTGCTGACGTGGCATTCCTCGAGACAGGTGCTGACGTGGCATTCCTCGAGACA
GGTGCTGACGTGGCATTCCTCGAGACAGGTGCTGACGTGGCATTCCTCGAGACAGGTGCTGACGTGGC
ATTCGAGCTCCGCGTGGATGACGTGTACATGGATGACGTGTACATGGATGACGTGTACATGGATTAAT
TAAGGCGCGCCCTTAAGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG
ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTT
CAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTAT
ATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC
GACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCCNTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGC
CATGGCTTCCCGCCGGAGTGGAAGAACAGGATGATGGCACGCTGCCATGTCTTGTGCCCAGGAAGACG
GGATGGACGTCACCCTGCACCCGTGCTTCGCTAGGATCAAGGGGCTAAATGAAAAGGGCCGTTTAAAC
CCCGGNTCAGCCCGAAGGGGCCTTTTGTTTCACCNNNTTTTTTTTCCCCCCCCCCCCGGCCTTCCTTACCC
CGAGGGGGCCCCCCCCCCTGCCCTTTAAAAAAAAAAAAAAATTTTATCTTTTTTTTAGGGTTTTTTNTTT
GGGGGGGGGGGGGGGGGGANNAGGGGGGAGATGGATAAAAAAAATGTGGGTGGGTGGGGNNTTA

>ERSE_C_FORWARD-Construct_Forward_J02
NGNNNGGGCTTGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTCGCGTCACCAATCGGAGGCCTCCACGACCACCAATCGGAGGCCTCCAC
GACACCGGTCTCGAGACAGGTGCTGACGTGGCATTCCTCGAGACAGGTGCTGACGTGGCATTCCTCGA
GACAGGTGCTGACGTGGCATTCCTCGAGACAGGTGCTGACGTGGCATTCCTCGAGACAGGTGCTGACG
TGGCATTCGAGCTCTTAATTAAACTAGTCCAGTGGCGCGCCATCAGCCTCGTACGGTGTGTCAGTTAGG
GTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAA
CCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCA
GCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCG
CCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAG
AAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGGCGCCCTCTGGCTAACTAGAGAACC
CACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGTTTAAACTTAA
GATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTAC
GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC
CGAAGGCTACGTCCAGGAACGCACCATCTTCTTCAAGGACGAAGGCAACTACAAGACCGCGCCGAAG
GTGAAGTTCGAGGGCGAAACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAGGAGGACGG
CAACATCCTGGGGCCAAGCTGGAATAAAATACACAGCCCAACGTCTATTCNNGGCCCAAAAACAAAA
AAAGGGGATCAGGGGAAATTTTAAAATTNGCCCAAAAATTGGGGAAAGGGGGGGGGGGGNTNNCAA
AAAATAAAAAAAAAACCCCCTGGGAAGGGCCCGGGGGGGGGCAAAAAACANNNNNAAAACCCCNNN
CCCCCCCAAAAAAAACAACAAAAAAAAAAAGAGAAAAATGTTNNNTNNNTGTTTTGGGCCCCCGCGG
GTATTCCCNNNGGGGGAAAAAAAAAAAAAAAANAAAATGNNNTCCCCC

>ERSE_MINP_FORWARD-Construct_Forward_N02
NNNNNGGCNGGACGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTTAAACCCGACCGGTGAGCTCCGCGTGGATGACGTGTACATGGATGAC
GTGTACATGGATGACGTGTACATGGATTAATTAAGGCGCGCCCTTAAGATGGTGAGCAAGGGCGAGG
AGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGC
GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCT
ACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGC
ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCT

GGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG
GAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGA
ACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC
CCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCA
AAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATG
GCACGCTGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCT
AGGATCAATGTGGCTAGATGATAAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCCTCCACTGTCCT
TTCCAATAAAATGAGGAAATTGCATCGCATTGTCTGAATAGGGTCTTCTATNNGGGGGGGTGGGGNGG
GCAGGAACCAAGGGGGAGGATTGGAAAAAAATACAGGCTGCTGGGGATGGGGGGGCTTTTGGTTTCG
AGGGGAAAAAACCNNNGGGNCTTAGGGGGATCCCCCCCCCCCCTGGGAGGCGCCNATTNGCNNGGGGG
GGGGGGGGGTTAACCCCAAGGGGACCCATAACNNTTGNNNCGCCCCAAAAGCCCCNTTTTTTTCTTTT
TCTTTNTTTCTCCCCGTCTCCCGTNNCCCCCCAAACTTAAAAGGGGNNNTTTTGTNCAATNNNNNNTTN
NC

>SRE_A_FORWARD-Construct_Forward_B04
NNNGCTTGGCCGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTAC
GGGCCAGATATACGCGTTGGTCCCGATCTCCTCTCACTGACCTTCAATTGGTCAGAGAACCTCTCACTC
CGCTCCCGTTGGCTCTCGCCGTCTCTCGCCGTCCCGCCATTGGCCGAGATGTTCTCGGGCCACCCTACG
TCACCGTCTCACCCCACTTCCCCGGTCCTCCGACTTTGTCCCCGCCCCTTCCCCCACCCTTCCAGCCGCG
AGGGAAAACCCTAGCGAGTCACCGCCTCTCCTTTCCTTCCCATTGGCAGAAAGCGGTATGAGGGGCGG
GAGCTGACGAGCGAGTCTGAGCTCTGGGATTGGCTGCGGCAGCCGCCATCCCGCCCCCTGCCGCCTCC
CGGCGGTGCGGGAGTCGGGTGGGTCGGTGGCTATAAAGCTGCTGAGGGCGGGAGGCACAGTGCACTG
TGCTCCGTGGATTGCTCCTTTCTTTGCCCCGGGTCCGGTCGTGTTTGGTTAATTAAACTAGTCCAGTGGC
GCGCCATCAGCCTCGTACGGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAA
GTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGC
AGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCC
GCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAG
GCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT
TGCAAAGGCGCCCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCA
CTATAGGGAGACCCAAGCTGGTTTAAACTTAAGATGGTGAGCAAGGGCGAGGAACTGTTCACCGGGG
TGGTGCCCATCCTGGTCAAACTGGACGGCAACTTAAACGGCCACAAGTTCAGCGGGTCCGGGAAGGG
CAAAGGCAATGCCACCTACGGCAAGCTGACCCTGAATTTATCTGCCCACCGGGAAGCTGCCGGGGCCT
GGCCACCCCCGGGACNNCCTGAACAAAGGGGGGGGAGGGTTTACCCCCTACCCCAACCNAAAAAAAA
AATTTTTTTAAATCCCCCCGGGCCAAAGNGTTTTCCGGAAGGCCCCCTTTTTTTTAGGGAAAAAGGAAA
ATAAAANNNNCCCCCCCAGGGGAAAATTGGGGGGCCCCCCGGGGAACNNCCCCGGAAGAAGGNNCTT
CTTTGGGGGGGAAGCACCCTCCGGGGGGGAGGGGGGTAAATAACCCCCCCCCCCTTTTTTTTGGGGGA
AAAAAAAAAAGGTGGGGGGTTTTTTTTCCCCANANGGGGGGGGGGGCCCCCACAAAAAAAACCCCGG
GGNNGGGGGGGG

>SRE_B_FORWARD-Construct_Forward_F04
NNNNNGNGGGGTTGGACCGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGC
GATGTACGGGCCAGATATACGCGTTGGTCCCGATCTCCTCTCACTGACCTTCAATTGGTCAGAGAACCT
CTCACTCCGCTCCCGTTGGCTCTCGCCGTCTCTCGCCGTCCCGCCATTGGCCGAGATGTTCTCGGGCCA
CCCTACGTCACCGTCTCACCCCACTTCCCCGGTCCTCCGACTTTGTCCCCGCCCCTTCCCCCACCCTTCC
AGCCGCGAGGGAAAACCCTAGCGAGTCACCGCCTCTCCTTTCCTTCCCATTGGCAGAAAGCGGTATGA
GGGGCGGGAGCTGACGAGCGAGTCTGAGCTCTGGGATTGGCTGCGGCAGCCGCCATCCCGCCCCCTGC
CGCCTCCCGGCGGTGCGGGAGTCGGGTGGGTCGGTGGCTATAAAGCTGCTGAGGGCGGGAGGCACAG
TGCACTGTGCTCCGTGGATTGCTCCTTTCTTTGCCCCGGGTCCGGTCGTGTTTGGTTAATTAAGGCGCG
CCCTTAAGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC
GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGC
TGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTG
ACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC
CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGA

GGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCC
GACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCACAACATCGAGGACGGCAGCGTGCA
GCTCGCCAACANTACAGCAGAAAACCCCCATCGGGGAAGGGCCCGGGGTGGTGCCGAAAACCCTACT
GAACCCCAGTCCGCCTGAACAAGACCCCACGAGAAGCCGATCCCTGGTCCTGCTGGAATTCGGACCGC
GCCGGATCCTTTCGGATGGACAAATGTAAAAAAATTAACCTGGTTTCCCCCCGAAAGGGGAAAAAAA
GAAAAAGGCCCTCCCCTTTTTTGGCCCCAAAAAANNAAGACCCCCCCCCCCCCTGTTTCNCTTAAAAA
TGNNNGAAAAAAAAAAGGCCCTTTAACCCCCTAAAACCCCCCGGCCCTTNNCCCCCCCCCTTGTTGTCC
CCCCCCCCNTCCCCCCCCCGGGGGGGCCCCCCCCCNNNTTTAAAAAAAAAAAAATTTTTTTNATTTTTTT
TTTTGGGGGGGGGGGGGGNNAGAGAAAAAAAAAAAAGTGGGGGGG

>SRE_C_FORWARD-Construct_Forward_J04
NNNNGGGGCTTGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTGGTCCCGATCTCCTCTCACTGACCTTCAATTGGTCAGAGAACCTCTCA
CTCCGCTCCCGTTGGCTCTCGCCGTCTCTCGCCGTCCCGCCATTGGCCGAGATGTTCTCGGGCCACCCT
ACGTCACCGTCTCACCCCACTTCCCCGGTCCTCCGACTTTGTCCCCGCCCCTTCCCCCACCCTTCCAGCC
GCGAGGGAAAACCCTAGCGAGTCACCGCCTCTCCTTTCCTTCCCATTGGCAGAAAGCGGTATGAGGGG
CGGGAGCTGACGAGCGAGTCTGAGCTCTTAATTAAACTAGTCCAGTGGCGCGCCATCAGCCTCGTACG
GTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCT
CAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATG
CATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGT
TCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCT
CTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGGCGCCCTCTGG
CTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGC
TGGTTTAAACTTAAGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC
GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCCNATGAAGCAGCACGACTTCTTCAA
GTCCGCCTGCCCGAAGGCTACGTCCAGGAACGCACATCTTCTTCAAGGACGACGGAACTACAGACCGG
GCCAAGTGAATTCGAGGGGAACCCTGGTGAACCCATCGACTGAAGGGATCGACTTCAGGAGGANGGC
AAATCCGGGGGCAAACTGGAANAATTCAAANCCCCCCTTTTATTTGGCCCNAANNAAAAAAGGGCTT
AGGGGATTTAAATCCCCCCAATTNANGGGGGGGGGGTTCCCAAAATAAAAAAAAAACCCCCTGGGGG
GGCCCGGGNNNCAAAAAAAATAAAAAACCTCCCCCGAAAAAAAAAAAAAAANNNAAAATTGTTTGGG
TGGTGNCCCGGNNTTTTGGGAAAAAAANNNTTTTTTCCCCGGGGGGAAAAANACCCTTTTTTCNNNAG
ACCCCCCCCTTTTTTATAATAAAA

>SRE_MINP_FORWARD-Construct_Forward_N04
NNGGAGGCTGGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGT
ACGGGCCAGATATACGCGTTTAAACCCGACCGGTGAGCTCTGGGATTGGCTGCGGCAGCCGCCATCCC
GCCCCCTGCCGCCTCCCGGCGGTGCGGGAGTCGGGTGGGTCGGTGGCTATAAAGCTGCTGAGGGCGGG
AGGCACAGTGCACTGTGCTCCGTGGATTGCTCCTTTCTTTGCCCCGGGTCCGGTCGTGTTTGGTTAATT
AAGGCGCGCCCTTAAGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG
ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTT
CAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTAT
ATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC
GACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGC
CATGGCTTCCCGCCGGAAGTGGAGGAGCAGGATGATGGCACGCTGCCCATGTCTTGTGCCCAGAAGAG
CGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCATGTGGCTAGATGATAAGGGCCCGT
TTAACCCGCTGATCACCTCGACTGGGCCTTCAATTGCCAGCCATCTGTTGTTTGCCCTTCCCNGGCCTTC
TTGACCNGGAAGGGCCCNCCCCTGTCCTTTCCAATAAATNGAGAAATGCCCCCATTGTTGGAAAGGGT
TTTTTTTTTGGGGGGGGGGGGGGGGGGGANNNAAGGGGGNNNAATGGAAAAAAAAAAAAGGGTTTG

GGGGNGGGGGGGGTTTTTNNNTTTTGGGAAAAAAAAAAACGGGGGGTTTGTNGGNTCCCCCCCCCCCC
CGNNNNCGCCTTAATNNNNNGGGGGNNNGGTCCCCGGGNCCCTTTTTCGNNNCCCCCCCCCCTNNTTT
TTTTTTTTTTTTTTTNCCNAGCGGGGTTTCNNNACAANNNANNGGCC

>FASN_A_FORWARD-Construct_Forward_B06.ab1
NNNNNNNNNNNNNGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGT
ACGGGCCAGATA
TACGCGTGGGAGGCGTGGAGCACGGAACGGAAGTTGGGGGGCGGGGGTGACACCGTGCCCCGCCCCA
GAGCCCCCCGAGT
CCGGGGCCCCCAACCCGGCGCCCCCTGGGCGCGCCCCCGCGCAGGGTCCCGGCTCCCGGCGGCGGCGC
GCCGCATCACCC
CACTGGCGGCGGCGCGCCTTGTCCCGGGGCGCAGCCCCGACGCTCATTGGCCTGNGCGGCGCANCCNA
NCTGTNNNCCCA
TGTGGNGTGTCCGCACGGNGACNACCGCGGTTAAATAGCGCCGGNGCGNGNCTANAGGGAGCCAGAG
AGACNGNNGTANC
GGCCTCTCNTCNNNNGCACACTCNNNNCNNNCTCTCCCTCANNCGTTCGCACNGCCGCCCGCGCCCNG
ACCAGGTACAAG
CGGNCNTTAATTAAACTAGTCCAGNGGCGCGCCATCAGCCTCNTACGGTGTGTCAGNTNNGGTGTGGA
ANGTCCNTANGC
TCCCCAGNANGNANAANTATGCANNGNNTGCATCTNNNNTANNNNNNNNACCNNGNGNGNAAAGNNN
CNNNNNNNCCCNGN
ANGNANANNTATGCNNAGCANGNNTCTCNNNNNNNNANCAACNNTANTCNCNNNNCNAANNNNNNNN
NNNNNNNNNNNNNN
NTCNNNNNNNNNNNNNNNNNNTNNNNNNNNNNNNCTGANTNANTNTNNTTNANNNATGNNNANNNN
NNNNCCNNNNCNGC
NTCTNANCTATNNNNNAANNAGNGANNAAGNCNNNNNNTNGNAGGCCNTNNNNCNTTTGCAAAGGN
NCCNNNNNGNNNTA
NNTANNNNNNNCNNNNNNNNNANGGNNTTNNNNNNAANNNNNNNNNNCCCCNNTATNNGGNNNGA
NNNCNNNNNNGNNTN
AAANTNNNNNNNNGTNNNNCNNNNNNNNAGNANCNNNTTNNNNGGNNGGGNNNNCCCNNCNTGGAN
NNGANNTGGACTGN
NNNNNNNANGGGNNNNNAANTNNNNNNGGGNNNNNNGNNNN

>FASN_B_FORWARD-Construct_Forward_F06
NNGGAGGCTTGACCGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTTAAACCCGACCGGTGAGCTCTTAATTAAACTAGTCCAGTGGCGCGCC
ATCAGCCTCGTACGGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATG
CAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAA
GTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCC
TAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCG
AGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCA
AAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGATGAAAAAGCCTGAACTC
ACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTC
GGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATA
GCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTC
CGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGT
GTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATGGA
TGCGATCGCTGCGGCCGATCTTAGCCAGAAGAACGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTC
AATACACTACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCTGTGTATCACTGGCAAACTGTGAT
GGACGAAACCGTCAGGGCGTCCGTCGCGCAGGCTTCGATGAACTGATGCTTTGGGCCAAGGACTGCCC
CAAAGTCCGGCACCTCGGGCANGCGGATTTCGGCTCANNCAAGGCCCGACGGAAAAGGGCCCCAAAA
ACCGGTCTTTGACGGGGAGGGGNNNGTGTTCGGGAATTCCCAAAAACAAGGNCCCCACACTTTTTTTT
TGGGGGGGGGGNNNGGGGTTGTNNGGGGAAAAAAAANCCCCCCTTNNTAAGGGGAGAGGTTNNGGA
GTGTTGAGGAACACCCGCGCGCTCCNGGGGGTTATGCCCTCCTGTGTGGTTNNTAACAATCTTAAAAG
TGGTNGAAAAAT

>PPAR_A_FORWARD-Construct_Forward_J06
TTTGACGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCC
AGATATACGCGTAATGTAGGTAAAGGGTCAACCGGTAATGTAGGTAAAGGGTCAGAGCTCAATGTAG
GTAAAGGGTCATTAATTAAACTAGTCCAGTGGCGCGCCATCAGCCTCGTACGGTGTGTCAGTTAGGGT
GTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACC
AGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGC
AACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCC
CCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAA
GTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGGCGCCCTCTGGCTAACTAGAGAACCCA
CTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGTTTAAACTTAAGA
TGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGT
AAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG
AAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGG
CGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG
AAGGCTACGTCCAGGAGCGCACCCTCCTCTTCAAGGACGAGGCAACTACAAGACCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCCAAGCTGGAGTACACTACAACAGCCAAACGGCTATTTATGGCCGACAACCAAAGAAC
GNATCAGGGGAATTCAAATCCGCCAAAATCAGGAAGGAGGGGGGGTTGCCGAACCATACAAAAAAA
ACACCCATCGGAAAGGCCCAGGGGGGGGCGAAAACCATATNGGAACCCGGCGCCCGGAGAAAAACC
CAAAAAAAAGNATAAAGGTTTGGGGGTGTTGGGGCCCCCGGTATTCCTCTGTTGGAAAAATTTAAAAA
AATAAAACCTTTCCCC

>PPAR_B_FORWARD-Construct_Forward_N06
TTTGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTAATGTAGGTAAAGGGTCAACCGGTAATGTAGGTAAAGGGTCAGAGCTCAATGTAGG
TAAAGGGTCATTAATTAATAGAGGGTATATAATGGAAGCTCGACTTCCAGCTTAAGATGGTGAGCAAG
GGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACA
AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTT
CAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCC
AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGG
CGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGG
CACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCA
TCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCA
GCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCG
CCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGC
AGGATGATGGCACGCTGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGT
GCTTCTGCTAGGATCAATGTGGCTAGATGATAAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGT
GCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACT
CCACTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCNGANTAGGNGCATNNTTTTGGGGG
GGGGGGGGGTGGGGGAAGAAAACAAGGGGGGAGGATTGGGAAAAAAAAAAAGGGTTTGGGGGNNN
TGGGGGGGGCCTTTGGTTTTTAGAGGGAAAAAACCCCTGGGGGTTTNAGGGGNNNCCCCCCCCCCCTT
TCCCCTNNTAGGGGGGGGGTGGNGGTTTCCCCCCGACCCCTCTTTTTCGCCCCCCTACCCCCCTTTTTTT
TTTTTTTTTTTTTTNNCCCGGGGGGGTTTCCCAAAATATAAAGGGGCCCTTTTGGTCTTTTTNNTTTTCC
CCCCCCAAAAAAATNGGGGGTGGGNCGGGGNNNNCCCCCCAAAANGTTTTTTTTCTGTGGTCNCTTTT
TTGGGTTNNCC

>PPAR_C_FORWARD-Construct_Forward_B08
NANGGGGGCTTGGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGAT
GTACGGGCCAGATATACGCGTAATGTAGGTAAAGGGTCAACCGGTAATGTAGGTAAAGGGTCAGAGC
TCTTAATTAAACTAGTCCAGTGGCGCGCCATCAGCCTCGTACGGTGTGTCAGTTAGGGTGTGGAAAGT
CCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGA
AAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGT
CCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTG
ACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGG

AGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGGCGCCCTCTGGCTAACTAGAGAACCCACTGCTTACTG
GCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGTTTAAACTTAAGATGGTGAGCAA
GGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCAC
AAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTG
CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCT
TCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GNNCAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAAAAGAACGGC
ATCAAGGTGAACTTCAAGATCCGCCCCAACATCGAGGACGGGAGCGGGCAGCTCGCCGAACCATACA
AAGAAAAACCCCCATTGGGGAAGGGCCCGGGGGTGGTGCCNAAACCCTTTCCGGACACCCCATCCCC
CCCTGGAAAAAAAAACCCCAAAAAAAAAAACCGGATACCATGGTCTGNNGGGATTTTTGGACCCCCC
CGGGGATCCCTCNCGGGGGGAAAAANNNTAAAAAAAAAATTTCCTTGGTTTTCCCCCCCGGAGGNAG
AAAANAGAAAAGGGGCCCCNNCCCTCTTTTTTGNCCCAAAAAAAAGGGGAGACCCCCCCCCNNNNNC
CGTNNNTNNNTNNNNNNNGANAGGGGAAAAAAANNGCCCATAACCCCCCCAACCCCCCCTGGGCCTN
NTTACCCCCCATGTTTTTTNNCCCNCCNNCTCCTTNATCANNAGGGACCCCCNCNCCCTTCTATTAAAA
AAAGNAAACCCNNGGNNGTATNTTTNNNNNGGGGGGGGGGGGNAAAGAAGAAANAAAAAAAANNT
GG

>CHORE_A_FORWARD-Construct_Forward_F08
NNNNNNGGCTTGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTTAAACCCGACCGGTGAGCTCTTAATTAAACTAGTCCAGTGGCGCGCC
ATCAGCCTCGTACGGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATG
CAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAA
GTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCC
TAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCG
AGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCA
AAGGCGCCCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTAT
AGGGAGACCCAAGCTGGTTTAAACTTAAGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT
GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGC
AGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAG
CCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCNA
ACATCGAGGACGGCAGCGGGCAGCTCGCNAACANTACAGAAGAAACCCCCATNNGGGAAGGGCCNG
GGGTGGTGCCGAAAACCCTACTGAGACCCAGTCCGCCTGAACAAGACCCAACAGAAACCGATCCATG
GTCTGGTGGATTCGGACCCCGCCGGGATCCTCCCGGCNGGAAAANNNTAAAAAAATTACCCTGGTTCC
CCCCCNAAGGGGGAAAAAAAAAAAGGGCCCCTCCCCTTTTTTGCCCCAAAANNGGGAGGACCCCCCC
CCCCCCNNGGNNTTCGNGAAAAAGGGGAAAAAAAAGGCCCCNTNNACCCAAAACCCCANGGGTTTTT
TTTCCCCANTNTTTTCCCCCCCCCCCTCCCCCGAGGNNNCCCCCCCCTCNCAAAAAAAAANNNCTNNG
AAAATGTTTTTTTGGGGGGGGGGGGGAAAAGAAAAAAAAAAAAGGTGGGGNNNTTTAAAAAAAAGTG
TCCCCCCCCCGGGGGGGGGGGGGGAGCCCCCCTTTTTNCGAAAAA

>CHORE_B_FORWARD-Construct_Forward_J08
TTGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTCACACCGTGTCCACGCGACCGGTCACACCGTGTCCACGCGGAGCTCCACACCGTGTC
CACGCGATTAATTAATAGAGGGTATATAATGGAAGCTCGACTTCCAGCTTAAGATGGTGAGCAAGGGC
GAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT
CAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA
CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGC
CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACA
AGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA

GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAG
AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCT
GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
ACTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGG
ATGATGGCACGCTGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCT
TCTGCTAGGATCAATGTGGCTAGATGATAAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCCTCCANT
GTCCTTTCCTAANAAATGAGGAAATTGCATCGCATTGTCTGAATAGGGGCATTCATTCGGGGGGGTGG
GGGGGGGGAAGAAAACAAGGGGGAGAGTTGGAAAAAAAAANNAGGTGGGTGGGAANNGGGGGTCTT
TGTTNNAGGGAAAAAACCGGGGGGTTTTGGGGGTTCCCCCCCCCCCCGGGAGCGCCATTNTAGGGGG
GGGGGGGGGGGGTTCCCCGAGGGCGCCCNNCNNNGTCCCCCCCCNNNNNCCCCCTTTTTTTNTTCTC
TTTTTTCNNNCCGGGGTTTCCCCCCCAAAAAAAAAGGGGGCCCTTTTGGGTGTATTGTTTTTCCCCCCC
CCCCAAANTNTGNNAGGGCNNNGGGGCCCCCCCACNANTTCCTCTTGGAAACCTTTTTATANNTTAAC
AAAACCCCCCCTCTTTTTTTTTTTTTTTCTTTGTTAAAAAAAAAATATTNNNTGGGGGNNGGGCCCCCC
CNNCNNNNNNNNNC

>CHORE_C_FORWARD-Construct_Forward_N08
NNNGNGAGGGTTGACGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGA
TGTACGGGCCAGATATACGCGTCACACCGTGTCCACGCGACCGGTCACACCGTGTCCACGCGGAGCTC
TTAATTAAACTAGTCCAGTGGCGCGCCATCAGCCTCGTACGGTGTGTCAGTTAGGGTGTGGAAAGTCC
CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAA
AGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTC
CCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGA
CTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGA
GGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGGCGCCCTCTGGCTAACTAGAGAACCCACTGCTTACTGG
CTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGTTTAAACTTAAGATGGTGAGCAAG
GGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACA
AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
ACCACCGGCAAGCTGCGCGTGCCCTGGCCCACCCTCGTGACCACTCTGACCTACGGCGTGCAGTGCTT
CAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCC
AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAATTCGAGGG
CGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAGGAGGACGGCAACATCCTGGGGC
ANAAGCTGGAGTACAACTACAACAGCCACAACGTCNATTTCATGGCCGACAAGCAAAGAACGGCATC
AAGTGAACTTCAGATCGGCCACACATCAAGGACGGGAGCGGGAGGTCGCCGAANACTACAAAGAAAA
CCCCNATGGGGAANGGCCCGGGNNGTGGCCGAAACCATAACCGGAACCCGTCCGCCTGAAAAAAAAC
CCAAGAAAAACGGTTAAGGGTCTGGGGNNNGGAAC

>MINPROMOTER_FORWARD-Construct_Forward_B10
NNNGGAGGCTTGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
TACGGGCCAGATATACGCGTTTAAACCCGACCGGTGAGCTCTTAATTAATAGAGGGTATATAATGGAA
GCTCGACTTCCAGCTTAAGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCG
TGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC
TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCT
ATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGC
CCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACAT
GGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGAAGCTTA
GCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCACGCTGCCCATGTCTTGTGCCCAGGAG
AGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGGCTAGATGATAAGGGCC
CGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCC
GTGCCTTCCTTGACCCTGGAAGGGGCCACTCCACTGTCCTTTCCAAATAAATGAAGAAATTGCATCCCA
TTGTCTGAATAGGGTCATTCATTCGGGGGGGGGGGGGGGAAGAAACCAAGGGGAAGAATGGAAAAA

AAAANCAGGCNGGTGGGGANNNGGGGCTTTGGTTTTGAGGAAAAAAAANNNNGGTTTTGGGTTCCCC
CCCCCGGGGGGGTTTAGCNGGGGGGGGGTTCCCCGGGCNNNNTTTCCGCCCAACCCCTTTTTTTTTTT
TTTTTCCAGTNCGGTTCCAAATTAAAGGGGCCTTTGGGTNATTTGTTTGCCCCCCCAAAANNAAAGGG
GGGCAGGGGCCCCAAAATTTTCTTTTGACCTTTTATGTNNNCAAAACCCCCTTTTTTTTATGTTTCCCA

>SV40_DGFP_FORWARD-Construct_Forward_F10
NNNAGGCTGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTAC
GGGCCAGATATACGCGTTTAAACCCGACCGGTGAGCTCTTAATTAAACTAGTCCAGTGGCGCGCCATC
AGCCTCGTACGGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAA
AGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTAT
GCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAA
CTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGG
CCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAG
GCGCCCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGG
GAGACCCAAGCTGGTTTAAACTTAAGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCC
CATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACACTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCC
CACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACGCCGACCACATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACCCCTGGNGAACCGCATCGAGCTGAA
GGGCATCGACTTCAGGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCAC
AACGTCTATATCTGGCCGACAAGCAGAAAAACGGCATCAAGGGGAACTTCAAGAACCGCCCCACCAT
CAAGGACGGGAGGTGCNGTTCGCCGAACCTAACAAAAAAAAACCCCAATGGGGAAAGGGCCCGGGG
GGGGGGCCAAAAACTCCANNGGAGGACCCCTCCTCCCCAAAAAAAACCCAAAAAAAAACNNNACTTG
TCTTGGAAATTTTGGGGCCTNCNGGGATCCCTCCGGTGGGAATAAATTTTAAAAAACTATAACTTGGTT
NCCCCGGGAAGGAGGAAAAGAGAAA

>DGFP_FORWARD-Construct_Forward_J10
TTTGACGACATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTTTAAACCCGACCGGTGAGCTCTTAATTAAACTAGTCCAGTGGCGCGCCATCAGCCTC
GTACGGTGTGGTGCCTGCAGGTGCATATGGCGCCCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGTTTAAACTTAAGATGGTGAGCAAGG
GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA
GTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCA
CCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTC
AGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA
GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGC
ACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCAT
CAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGC
CCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGG
ATCACTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTTCCCGCCGGAAGTGGAAGAACA
GGATGATGGCACGCTGCCCATGTCTTGTGCCCAGAAGANCGGGATGGACCGTCACCCTGCAGCCTGTG
CTTCTGCTAGGATCAATGTGGCTAGATGATAAGGGCCCGTTTAACCCGCTGATCAGCTCGACTGGGCC
TTCTATTGCCAGCATTTGTTGTTTGCCCTTCCCCNGGCCTTCTTGACCTGGAAGGGCCCTCCCCTGTCTT
TCCAAAAAATGAAGAAATGGATCCATTGTTGAAAAGGGCTTTTTTTTGGGGGGGGGGGGGGGGGGAG
GGAAACAGGGGGGGAAAATGAAAAAAANNAAAGGNNTGTGGGGGAGGGGTGGNNNTTGNNNTTAGG
GGGAAAAAAAANNTGGGGTTTTGGGGGTTCCCCCCCCCCCCCGGGGCCATTATCGGGGGGGGGGGGNN
GCCGCAGNNNNNCCCATTTGTGNNNCCCACCCCNNCTTTTTTTTNNTTTTTTNNCCCCNCCCGCTNCCC
CNNAANNNAAGGNNCCNNNTTNNTTTATTTTTCTNNCCCNNCNNNAAAAGNGGGTGGGGNCNNCCCC
CNNGTGTNNTTTTCTGNTCCTTCNNNTNNNTTCTCNCCACNNCCCNNTNTTTTTTTNNNNNTTTTANTA
AAAANNAATATGTGGNNNCCCC