Tools for Next-Generation Transcriptional Control in Chinese Hamster Ovary Cell Factories

Adam John Brown

The University Of Sheffield.

Submitted for the degree of Doctor of Philosophy

May 2014

Department of Chemical and Biological Engineering

University of Sheffield

Declaration of Originality

In accordance with the University regulations, I hereby declare that:

1. This thesis has been composed solely by myself
2. It is entirely my own work
3. It has not been submitted in part or whole for any other degree or personal qualification

Based on the work within this thesis the following articles have been published (shown in Appendices C and D):

Chapter Two:

Brown AJ, Sweeney B, Mainwaring DO, James DC. 2014. Synthetic promoters for CHO cell engineering. Biotechnol. Bioeng. doi: 10.1002/bit.25227

Chapter Three:

Brown AJ, Mainwaring DO, Sweeney B, James DC. 2013. Block decoys: transcription factor decoys designed for in vitro gene regulation studies. Anal. Biochem. 443(2): 205-210

Based on the work within this thesis the following patent applications have been filed:

Chapter Two:

Brown AJ, James DC. UK patent application number GB 1321109.9 (29 November 2013) Synthetic promoters for CHO cells

Chapter Three:

Brown AJ, James DC.UK patent application number GB1310853.5 (18 June 2013) Transcription factor block-decoys

Acknowledgements

This research would not have been possible without sponsorship from UCB and the Engineering and Physical Sciences Research Council.

I owe a huge thanks to Professor David James for being supportive of all research ideas, for providing the wisdom and guidance to make them a reality and, most importantly, for responding with a joke every time they went wrong. I am also grateful to my industrial supervisors Dr Bernie Sweeney and Dr David Mainwaring for their feedback, help and advice.

I would also like to thank all colleagues of the James research group for making it such a fantastic, supportive and frequently hilarious place to complete a PhD. In particular to Dr Sarah Davies for providing unofficial supervision on the way to the train station every day.

Of course, most thanks to my partner and best friend, Kim. As with everything else in life, doing this PhD would have been half as fun and twice as difficult without her.

Abstract

Recombinant gene transcription in Chinese hamster ovary (CHO) cells, the dominant cell factory utilised for biopharmaceutical production, is still routinely regulated with a limited set of functionally ill-defined and uncontrollable genetic elements. This study presents novel transcription control technologies that facilitate development of next-generation biopharmaceutical manufacturing systems.

Firstly, synthetic promoters designed specifically to harness the pre-existing transcriptional activation machinery of CHO cell factories have been constructed. Transcription factor regulatory element (TFRE) function was screened in CHO cells and active elements were utilised to create synthetic promoter libraries exhibiting 140 discrete activites, operating over two orders of magnitude, where the most active promoters significantly exceeded that of the human cytomegalovirus immediate early 1 (hCMV-IE1) promoter. Through precise control of recombinant gene expression in CHO host cells over a broad dynamic range this technology could be utilised to both maximise transcription of easy-to-express proteins and provide optimised protein-specific transcription levels (synchronised with polypeptide-specific folding kinetics) of difficult-to-express proteins. Further, it will enable construction of bespoke, synthetic cell factories that require the expression of several genes to be stoichiometrically balanced.

Secondly, a novel method of transcription factor (TF) decoy (synthetic oligodeoxynucleotides that specifically sequester cognate TFs) formation has been developed, where blocks containing discrete TF binding sites are combined into circular molecules. Unlike currently available methods block-decoys allow rapid construction of chimeric decoys targeting multiple TFREs. Moreover, they enable fine tuning of binding site copy ratios within chimeras, facilitating sophisticated control of the cellular transcriptional landscape. It was demonstrated that a bespoke block-decoy chimera was able to inhibit expression from multiple target elements simultaneously in CHO cells. Block-decoys can be utilised to investigate any multi-TF mediated cell function or phenotype and represent a valuable new tool for characterising and controlling CHO cell transcription.

Finally, the mechanistic functionality of the promoter most commonly utilised to drive transgene expression in CHO cells, hCMV-IE1, has been analysed. It was found that hCMV-IE1 promoter activity in CHO cells is predominantly mediated via just two TFREs (CRE and NFkB), where physical prevention of TF-TFRE interactions at these sites, either by intracellular TF sequestration or TFRE deletion, reduced activity by >75%. This mechanistic understanding of hCMV-IE1s functional regulation in CHO cells facilitates strategies to predictably control or improve its activity by engineering the promoter's TFRE composition or the cell factory's TF abundances. This will likely be most useful for optimising transient gene expression systems where hCMV-IE1 is the current promoter of choice. Cumulatively, the tools developed in this thesis enable sophisticated, next-generation transcriptional control in CHO cell factories.

Contents

List of Tables

List of Figures

Acronyms and Abbreviations

AP1	Activator protein 1
BRE	TFIIB recognition elements
C/EBPα	CCAAT-enhancer binding protein alpha
CArG	$CC(A/T)_6GG$ element
CDP	CCAAT displacement protein
CHEF-1α	Chinese hamster elongation factor-1α
CHO	Chinese hamster ovary
cMyb	cellular myeloblastosis
CPRE	core promoter regulatory element
CRE	cyclic adenosine monophosphate-RE
CRM	cis-regulatory module
DCE	downstream core element
DHFR	Dihydrofolate reductase
DPE	downstream promoter element
DTE	difficult-to-express
E-box	Enhancer box
E2F	Elongation factor 2
EGR1	Early growth response protein 1
ERRE	Estrogen-related receptor alpha-RE
FACS	fluorescence-activated cell sorting
Gfi	Growth factor independence
GFP	Green fluorescent protein
GRE	glucocorticoid-RE
GS	Glutamine synthetase
hCMV-IE1	human cytomegalovirus immediate early one
HDAC	histone deacetylase
HNF	hepatocyte nuclear factor
HRE	helios-RE
Inr	initiator element
IPF	insulin promoter factor
ISRE	interferon-stimulated-RE

IVCD	integral of viable cell density
KO	knockout
LTR	long terminal repeat
mAb	monoclonal antibody
MAR	matrix attachment region
MCS	multiple cloning site
MEF	myocyte enhancer factor
MPRA	massively parallel reporter assay
MSX	msx homeobox
MTE	motif 10 element
NBRE	nerve growth factor-induced gene-B-RE
NF1	nuclear factor 1
NFAT	nuclear factor of activated T cells
NFκB	nuclear factor kappa B
OCT	octamer motif
ODN	oligodeoxynucleotide
ORF	open reading frame
PIC	preinitiation complex
PTM	post translational modification
RARE	retinoic acid-RE
RE	regulatory element
RMCE	recombination mediated cassette exchange
SEAP	Secreted alkaline phosphatase
siRNA	short interfering RNA
SOI	site of integration
SV40E	simian virus 40 early promoter and enhancer
TESS	Transcription Element Search System
TF	transcription factor
TFRE	transcription factor regulatory element
TGE	transient gene expression
TRAP	Transcription Affinity Prediction tool
TSS	transcriptional start site
UR	unique region
UTR	untranslated region

XCPE	X core promoter element
YY1	Yin yang 1

Chapter 1: Introduction

This study aims to develop novel tools in order to facilitate next-generation transcriptional control in Chinese hamster ovary (CHO) cell factories. This introductory chapter contextualises the subsequent work within this thesis. An overview of biopharmaceutical production processes is provided, outlining the pathway from potential drug candidate gene sequence to purified therapeutic protein. Subsequently, the need to produce better drugs, cheaper and faster is highlighted and a set of critically required biomanufacturing developments are presented. The restrictive limitations of currently available recombinant gene expression control elements are then discussed. Finally, the chapter comes full circle to outline how sophisticated transcriptional control technologies could enable disruptive innovation in biomanufacturing, detailing the opportunities available at each discrete production process step.

1.1. Chinese hamster ovary cell factories for biopharmaceutical production

1.1.1. Biopharmaceuticals

Biopharmaceuticals, pharmaceuticals produced using biotechnology techniques, provide effective therapies to conditions such as cancer, multiple sclerosis and rheumatoid arthritis. There are currently over 400 biologics approved for use and thousands more in the developmental pipeline. Worldwide sales total in excess of $100 billion and the market is projected to maintain its 15% annual growth rate (Langer, 2013). Amongst the nine major classes of biologics (monoclonal antibodies (mAbs), hormones, growth factors, fusion proteins, cytokines, blood factors, enzymes, vaccines and anti-coagulants) mAbs are both the best-selling and most rapidly growing, with hundreds of candidates currently in clinical trials (Aggarwal, 2012). mAbs provide significantly improved therapies to many conditions, often covering large patient populations that require long-term treatment, and are usually engineered to improve functionality, for example by humanisation (Shukla *et al*., 2010). Indeed, the majority of modern biologics entering the market, including

fusion proteins and bispecific antibodies, are proteins specifically engineered to optimise their bioactivity, immunogenicity and pharmacokinetics, (Walsh, 2010; Czajkowsky *et al*., 2012).

1.1.2. Chinese hamster ovary cell factories

The majority of modern biopharmaceuticals are recombinant proteins produced in mammalian cell factories. Dominance of mammalian expression systems is primarily due to their capacity to perform human-like protein folding, assembly and post translational modifications (PTMs). This is particularly important for glycoproteins (approximately half of all approved biologics), where mammalian systems are required to produce proteins displaying therapeutically acceptable glycoprofiles with low immunogenicity and high bioactivity (Walsh 2010; Grainger *et al*., 2013). Given the relative high costs, low productivity and slow production pipelines there is continuing interest in developing non-mammalian production hosts. Accordingly, biopharmaceuticals that do not require human-like PTMs are predominantly produced in microbial cells, in particular *Escherichia coli* (Berlec *et al*., 2013). However, the majority of top-selling biologics do require PTMs (including mAbs, Tissue Plasminogen Activator, and Erythropoietin) and products produced in mammalian systems still account for over 70% of biopharmaceutical sales revenue (Zhu, 2012). Unsurprisingly, there have been significant efforts to synthetically engineer non-mammalian hosts to produce human-like proteins. For example, correctly folded and assembled (but not glycosylated) mAbs have recently been produced in *E. coli* (Huang *et al*., 2012). If the N-glycosylation pathway from *Campylobacter jejuni* can be successfully introduced into *E. coli*, these advances in protein processing could be exploited to enable faster, cheaper mAb production in microbial cell factories (Prandhal *et al*., 2012; Spadiut *et al*., 2014). There has also been significant progress in glycoengineering both yeast and plant systems to overcome their respective glycosylation problems of low serum half-life and high immunogenicity (Bosch *et al*., 2013; Spadiut *et al*., 2014).

Although developments in non-mammalian systems offer considerable promise in the long-term, mammalian cells are clearly established as the biopharmaceutical manufacturing factories of choice for the foreseeable future. Whilst multiple mammalian cell platforms have been utilised for production (baby

hamster kidney, NSO mouse myeloma, human embryo kidney), over 70% of therapeutic recombinant proteins are currently made in Chinese hamster ovary (CHO) cells. Moreover, the dominance of the CHO cell factory will likely increase in the near-future as most candidates currently in clinical development utilise a CHO cell platform (Walsh, 2010; Zhu, 2012). CHO cells have reigned as the preferred biopharmaceutical production system for > 25 years because they: i) produce therapeutically satisfactory glycoprofiles, ii) are well established safe hosts, being non-permissive for most human viruses, iii) are easily adapted to serum-free suspension culture in chemically defined media, providing compatibility with large-scale, regulatory-compliant bioreactor production processes, iv) benefit from decades of industrial optimisation, resulting in well-established CHO bioprocesses that have improved titers > 100 fold in 20 years, and v) have a good record of biologic approvals, facilitating an easier path through the strict regulatory procedures (Kim *et al*., 2012a; Lai *et al*., 2013). Although human cell lines such as Per.C6 are generating considerable interest due to potential advancements in viable cell densities and glycosylation profiles (Bandaranayake, 2014), CHO cells are clearly established as the most important biologics production factory of both the present and near-future.

1.2. Biopharmaceutical production processes

Production of a new therapeutic protein requires the development of a CHO clonal cell line delivering high titres of acceptable quality product under large-scale manufacturing conditions. The translation of a drug candidate gene sequence into vials of regulatory body-approved protein requires considerable investment in time, labour and costs and involves multiple discrete upstream (cell line isolation and bioreactor production) and downstream (product isolation and purification) bioprocesses (Shukla *et al*., 2010; Vijayasankaran *et al*., 2010). Each discrete production process presents both unique challenges and potential optimisation opportunities. Moreover, despite > twenty five years of optimisation there remains considerable pressure, and opportunity, to enact disruptive step-change innovation in all production process steps.

1.2.1. Process 1: CHO cell factory selection

Although CHO cells are the glycoprotein production factory of choice, not all CHO cells are created equal. Since the original CHO cell line was isolated in 1957 from a primary culture of Chinese hamster ovarian cells (Puck, 1985) multiple distinct CHO cell lineages have been developed with unique bioproduction characteristics via mutagenesis, directed evolution and genetic engineering. 'CHO cells' are in fact hundreds of discrete cell lines with varying genotypes and diverse phenotypes (Lewis *et al*., 2013; Wurm, 2013). Critical cell line characteristics required for biomanufacturing include i) growth in serum-free, chemically-defined suspension culture, ii) high specific protein productivity, iii) desirable bioreactor growth characteristics (*i.e.* maximised integral of viable cell density (IVCD), accelerated growth rate, *etc.*), iv) the capability to produce high-quality homogenous products, particularly uniform glycosylation events, and v) the ability to rapidly create stable, high yielding recombinant daughter cell lines (Birch *et al*., 2006; Kim *et al*., 2012a). Given that different CHO cell lines vary significantly in these key attributes, and further that they respond differentially to process platforms further downstream (media, feeding strategy, bioreactor conditions, purification, *etc.*) it is unsurprising that most biopharmaceutical companies have developed optimised production cell lines in-house. Utilising cell hosts specifically adapted for production processes improves product quality and yield, reduces development times, and enables implementation of standardised manufacturing bioprocesses. For example, the pre-adaption of parental lines to suspension growth in chemically defined, serum free media has enabled rapid switching between static and suspension growth, reducing timelines by approximately six months (De Jesus *et al*., 2013).

Development of next-generation factories is predominantly focused on genetically engineering cells to optimise key bioproduction functionalities, including apoptosis, proliferation, protein folding and secretion (Datta *et al*., 2013). Ultimately, most strategies to date have proved unsuccessful, and few genetically engineered hosts have been employed in manufacturing processes. However, Potelligent cells are an example of simple successful engineering, where gene-knockout of α-1, 6-fucosyltransferase resulted in production of fucose-free mAbs that exhibit enhanced antibody dependent cellular cytotoxicity activity (Shirita, 2009). This host has been licensed by multiple companies and is the parental cell line of several mAbs

currently in clinical trials. Development of the CHOK1SV (Lonza's proprietary growth-optimised host) Potelligent cell line is an example of multiple desirable bioproduction functionalities being combined in a single production-optimised factory.

1.2.2. Process 2: vector engineering

Plasmid vectors are utilised to introduce sequence-optimised product genes into the selected host cell line. Sequence engineering involves codon-optimisation and removal of elements that promote undesirable mRNA folding/ stability, such as cryptic polyA tails and splice sites (Birch *et al*., 2006; Hung *et al*., 2010). Appropriate gene expression control elements are employed to support efficient transcription and translation rates. These technologies are discussed in detail in section 1.4. Briefly, strong promoters (most commonly the human cytomegalovirus immediate early one (hCMV-IE1) promoter) are used to achieve high levels of transcription, and translational regulatory sequences (polyA tail, Kozak sequence, introns) are included to ensure favourable mRNA stability, export and translation (Kim *et al*., 2012a). Additionally, elements that minimise production instability, such as matrix attachment regions (MARs), are often incorporated flanking the expression cassette. Functioning to promote formation of open chromatin environments, these elements negate integration site-specific effects on transgene expression and prevent gene silencing (Rita Costa *et al*., 2010).

Successful plasmid genomic integration events are isolated utilising an appropriate selection system. Dihydrofolate reductase (DHFR) or glutamine synthetase (GS) selection marker genes are commonly used in conjunction with CHO hosts that require these enzymes for growth in either GHT (glycine, hypoxanthine, thymidine) or glutamine deficient media respectively. Selection of cell clones overproducing the selection marker is achieved by i) controlling selection-gene expression with 'weak' promoters and ii) exposing stable transfectants to increasing concentrations of specific enzyme inhibitors (methotrexate or methionine sulfoximine for DHFR and GS respectively) (Butler, 2005; Fan *et al*., 2012). Direct linkage of the selection and product genes in the expression vector therefore facilitates indirect selection of cells that also overproduce the product protein.

1.2.3. Process 3: transient expression

Before entering the time and cost-intensive process of developing a stable product producing cell line, small quantities of developmental material are produced via transient gene expression (TGE). Milligram to gram product quantities produced rapidly (< 5 weeks) and cheaply are utilised for toxicology studies, process development, and product quality evaluation (Baldi *et al*., 2007). Moreover, transient transfection provides an assessment of the products 'expressability' in the host cell factory. Rapid screening of candidate products utilising TGE prevents inefficient commitment of resources and improves speed-to-market (Cain *et al*., 2013).

1.2.4. Process 4: cloning

Following satisfactory results from developmental studies the expression vector is stably transfected into host cells via lipofection, polyfection or electroporation, prior to culture in selective media. Stably transfected cells represent unique integration events (*i.e.* vector inserted in different genomic loci at varying copy number) and vary significantly in critical bioproduction characteristics such as growth rate and productivity (Wurtele *et al*., 2003). Utilising this heterogeneous cell pool for protein manufacture would result in inconsistent bioprocesses, with unpredictable yields and variable product qualities. It is therefore a regulatory (and technical) requirement to perform single cell cloning (Agrawal *et al*., 2012). The production clonal cell line requires multiple key attributes that must be evaluated during the selection process, including high productivity, suitability for bioreactor scale manufacture, resistance to apoptosis, adaptability to suspension growth, production stability and high growth rate. Cells meeting these extensive criteria are rare events within the heterogenous population, necessitating the screening of several thousand clones (Lai *et al*., 2013). Traditionally this was achieved by serial limiting dilution. Requiring multiple rounds of dilutions to ensure monoclonality, and subsequent low throughput methods to assay clone productivity, this process was restrictively time, cost and labour intensive (De Jesus *et al*., 2011). Accordingly, under significant industry pressure, a number of optimised high-throughput screening technologies have been developed in recent years. For example, flourescence-activated cell sorting (FACS) can be utilised to specifically separate high producers from low producers, significantly reducing

the number of clones requiring screening. Multiple FACS-based methods have been developed to overcome the difficulties associated with monitoring production of secreted proteins. For example, linking easily detectable fluorescent markers such as green fluorescent protein (GFP) to the product gene facilitates indirect selection of high producers (Kim *et al*., 2012b). Alternatively, secreted protein can be 'captured' at the cell-surface using affinity matrices and detected with fluorescently labelled antibodies (Black *et al*., 2011).

Fully automated screening technologies have also been developed, including CellCelector and Clonepix systems (Mann 2007; Haupt *et al*., 2009). Growth in semi-solid media allows automatic isolation of distinctly separated clonal colonies, and further, results in localised close-proximity capture of secreted product. Following addition of an appropriate product-detection agent fluorescent halos form around clonal populations, whereby halo intensity is a measure of clone productivity. Clones are ranked according to growth rate, productivity and monoclonality and top performers are robotically transferred to 96-well plates for further characterisation (Dharshanan *et al*., 2011; Lai *et al*., 2013). Using automated systems several thousand clones are screened in a single week and poor performing clones are rapidly removed from the pipeline. Resultant panels of selected clones (~ 250) are robustly evaluated for key performance parameters such as productivity, growth rate and production stability. Subsequently, ~ 25 clones are re-adapted to suspension growth and their predicted performance in large-scale manufacturing processes is determined (Noh *et al*., 2013). Utilising shake flasks/ mini-bioreactors, production media and manufacturing feeding regimes critical bioproduction functionalities (cell growth characteristics, product titer, product quality, metabolism, *etc.*) are evaluated to select a final panel of ~ 5 clones. Production stability is monitored in these lines over several (> 50) generations to identify a single production clonal cell line and master/ working cell banks are prepared for use in large-scale manufacture (Li *et al*., 2010).

1.2.5. Upstream process 5: manufacturing process development

Producing sufficient cell mass to inoculate 20,000 L bioreactors requires stepwise expansion of cultivation volume. Starting with a single working cell-bank vial, biomass is accumulated through a series of shake flask and bioreactor stages,

typically ending with working volumes of 2,000 L (Davis, 2007). Several innoculum trains are run simultaneously to feed multiple production bioreactors. Stainless steel tanks remain the industry standard but disposable systems, with associated advantages in cost, flexibility and speed, are increasingly being utilised (Shukla and Gottschalk, 2013). Large-scale manufacture typically employs fed-batch bioreactor processes, where small feed volumes (< 10% volume) are added to culture to maintain key nutrient sufficiency. Perfusion culture systems, where 'old' media is continuously replaced, are infrequently utilised due to complexity and sterility issues (Pollock *et al*., 2013).

Media development requires definition of batch medium, feed composition, and feeding strategy. Companies frequently utilise a single chemically defined, protein free media comprising amino acids, lipids, inorganic salts, vitamins and trace elements. Whilst generic media is unlikely to be optimal for each product, owing to metabolic variance between cell lines, it significantly shortens development timelines and allows implementation of downstream platform processes (Li *et al*., 2010). In cases of unacceptably low productivity platform media can be optimised, for example by addition of animal-component-free hydrolysates. Additionally, antifoam and synthetic polymers (such as Pluronic F-68) are included in most media as protectants against sparging and agitation-related cell damage (Birch and Racher, 2006).

Feed compositions and regimen are designed to prevent nutrient depletion and toxic by-product accumulation. Controlled feeding of key substrates (e.g. glucose and glutamine) and consumed-nutrients (e.g. amino acids and iron) is critical to maximise both IVCD and specific productivity (Khattak *et al*., 2010). Although platform feeds are routinely employed, cell-specific variation usually necessitates feed optimisation via design of experiments approaches (Sellick *et al*., 2011).

Bioreactor operating conditions are optimised to maximise productivity, minimise protein aggregation/ degradation and prevent undesirable glycosylation modifications. Multiple chemical (pH, osmolality, dissolved oxygen/ carbon dioxide) and physical (temperature, mixing) parameters are controlled throughout the production process to prevent undesirable fluctuations significantly affecting product yield and quality. Operating conditions are optimised utilising scale-down bioreactor models, and parameters are either maintained (volume-independent) or

proportionally scaled up (volume-dependent) in full-scale validation studies prior to large-scale product manufacturing (Li *et al*., 2010).

Advances in media design, feeding regimen, and bioreactor control strategies have been the major source of yield increases during the past twenty years. Predominantly this has been achieved by facilitating improved cell growth profiles, whereby cell biomass is rapidly accumulated (proliferation phase) and subsequently maintained for extended periods of time (production phase) (Hacker *et al*., 2009). Cell growth dynamics are increasingly being optimised by employing biphasic bioreactor processes where proliferation and production phases are maintained at 37°C and 32°C respectively. Hypothermic conditions arrest cell growth, increase culture longevity, and enhance specific productivity (Masterton and Smales, 2014).

1.2.6. Downstream processes

Although this thesis focuses on novel upstream process technologies, a brief summary of downstream process steps provides a complete overview of biopharmaceutical manufacturing. Cells/ cell debris are first removed from bioreactor cell culture media by centrifugation and filtration using depth and membrane filters. Product protein is specifically captured by affinity chromatography (for example, Protein A chromatography for mAbs) to remove host cell proteins and DNA, prior to a low pH viral inactivation step (Hogwood *et al*., 2013). Polishing processes employing ion exchange and size exclusion chromatography remove host, product and process impurities to regulatory-standard concentrations. A final viral-removal filtration precedes an ultrafiltration step to concentrate the product in formulation buffer (Shukla *et al*., 2010). The resultant product can be utilised as a therapeutic drug, completing the pathway from potential drug candidate gene sequence to purified biopharmaceutical protein.

1.3. Biopharmaceutical manufacturing requires disruptive step-change innovation

Despite over two decades of bioprocess development and optimisation there is still considerable scope for innovation at each biopharmaceutical production process step.

Moreover, dramatic evolution of the biopharmaceutical industry is placing continually increasing demand, and necessity, to deliver these potential disruptive step-change advancements. Substantial pressure is currently being applied from key decision makers (governments, healthcare systems, insurance companies, and regulatory bodies), patients, and the biopharmaceutical companies themselves.

1.3.1. Industry Pressure

The blockbuster era has likely ended due to increased competition (*i.e.* multiple biologics available for the same conditions), reduced patient populations per drug (most new therapeutics coming to market treat conditions with relatively small patient populations) and the rise of biosimilars (Calo-fernandez, 2012). Accordingly, there is significant pressure to i) improve product qualities to provide superior therapies, ii) reduce manufacturing costs to improve cost/ profit per dose, and iii) shorten development timelines to maximise the period of patent-protected market exclusivity (Morton and Kyle, 2012). Further, bioindustrial portfolios of candidate biopharmaceuticals are diversifying to include many difficult-to-express (DTE) proteins, such as fusion proteins. Production yields of DTE proteins are typically low, with many candidates being inexpressible and impossible-to-manufacture (Pybus *et al*., 2014). Cell host capabilities are therefore critically restricting production, necessitating development of next-generation cell factories by either adopting new systems or engineering existing ones. Finally, the risk of product failure is increasing due to stricter regulatory approval processes and increasing target complexity. Of the small proportion of drug candidates that advance to clinical trials < 20% subsequently gain regulatory approval (DiMassi, 2014). Significant reductions in development costs are required to enable companies to absorb such product-failures whilst maintaining profits. Further, the risks of late-stage failure need to be minimised; potentially by building quality into products and implementing streamlined biomanufacturing processes with improved predictability and control (Elliott *et al*., 2013).

1.3.2. Patient pressure

Ultimately, the true value of biopharmaceuticals is their ability to extend, improve,

and save patient lives. As a result of extended life-expectancy and widespread adoption of cancer-causing behaviours (smoking, inactivity, *etc.),* increasing portions of the population are requiring treatment for cancer and central nervous system disorders (e.g. dementia, Alzheimer's disease) (Mariotto *et al*., 2010; Siegal *et al*., 2013). Biopharmaceuticals represent the most effective, and in many cases only, therapy for these conditions. Whilst this presents significant opportunities for industrial profits, at current cost-per-dose it is becoming economically impossible to provide biopharmaceutical therapy to such large patient populations. Indeed, increasingly patients are being denied access to potentially life-saving/ extending biologics. Further, millions of patients currently suffering life-changing morbidity could benefit from biopharmaceutical therapy if it were not considered economically intractable (Kelly and Smith, 2013). Moreover, the effects of globalisation are exponentially increasing the healthcare demands in developing countries (Jemal *et al*., 2011). Therefore, the worst-case scenario where cost, rather than capability, is the primary variable affecting patient outcomes is likely to be exacerbated in the near-future. Significant reductions in biopharmaceutical costs are therefore immediately required.

Development of modular, plug-and-play platform processes is a promising strategy to significantly reduce production costs, minimise development timelines and enable streamlined, flexible manufacturing. Such technology may facilitate patient optimised (stratified or personalised) medicine, which will likely require a pipeline capable of delivering high yields from quick, low-volume manufacturing processes with minimal optimisation.

1.3.3. Decision-maker pressure

Current models of healthcare spending are considered unfeasible and governments are attempting to reduce healthcare expenditure in the short term and provide a roadmap for sustainable spending in the future (Miller, 2012). Ultimately the solution to these challenges is to either buy fewer drugs or pay less for them. Accordingly, policy-makers, payers (in the U.S), and healthcare bodies (such as National Institute for Health and Clinical Excellence in the U.K) are increasingly choosing cost as their primary decision-making factor, reinforcing the requirement for disruptive innovation in biologics manufacturing to reduce cost-per-therapy.

Additionally, regulatory bodies have become progressively more cautious throughout the past decade, demanding both higher product standards and increased advancement over existing therapies (Morton and Kyle, 2012). Consequently, there is a critical need to both improve product qualities and enable production of DTE proteins.

1.3.4. Critically required biopharmaceutical manufacturing developments

Disruptive step-change innovations in biopharmaceutical manufacturing are clearly required to enable the ideal solution whereby decision-makers can select biologics therapies offering optimal patient outcomes whilst delivering sufficient industry profits. Currently, the ten most critically required biomanufacturing developments are as follows:

1. Increased titers via improvements in specific productivity and IVCD
2. Reduced costs of product development and manufacture
3. Faster product speed-to-market
4. Enhanced, and more predictable, product quality
5. Higher yields from TGE
6. Increased manufacturing flexibility
7. Development of standardised platform processes
8. Increased cell factory capability
9. Removal of production instability
10. Better prediction, and minimised risk, of product-failure

In simple terms we need to produce better drugs, cheaper and faster, and every biopharmaceutical production process step presents opportunities to achieve this.

1.4. Gene expression control technology represents a suboptimal biomanufacturing component

Innovative gene expression control technologies are needed to facilitate many of the critically required biomanufacturing developments. Expression cassettes contain

multiple sub-components including promoters, untranslated regions (UTRs), and expression stability elements. Cumulatively, these components control transcription and translation rates of both the product and selection marker genes. However, despite critical importance in determining productivity, CHO gene expression control technology is a severely unoptimised manufacturing component. Expression cassette elements would ideally enable sophisticated, optimisable control of all critical parameters (*i.e.* product and selection marker transcription and translation). However, currently, vector engineering involves the relatively primitive process of selecting components with rigid parameter outputs from a restricted toolbox. Precision controllability, predictability, and product-specific optimisation are therefore impossible. Given that i) suboptimal expression cassettes lock-in production limitations at the beginning of the manufacturing process that cannot be overcome with downstream optimisation, ii) currently utilised expression control technologies lack desirable functionality and exhibit well-known undesirable characteristics and iii) product gene transcription and translation rates are well-established critical rate-limiting factors in biopharmaceutical production, we simply cannot continue to engineer production vectors from a limited set of blunt-tool parts. The catalogue of currently available components will now be detailed, beginning with a brief overview of translation and expression stability control options, before turning to the focus of this thesis and major opportunity for delivering disruptive innovation, transcription control elements.

1.4.1. Defining the current toolbox: post-transcriptional regulatory elements

Recombinant mRNA translation rates have been shown to be a critical rate limiting factor in biopharmaceutical production (Lattenmayer *et al*., 2007; McLeod *et al*., 2011; O'Callaghan *et al*., 2011). Translation of a given mRNA is a function of mRNA export, localisation, and stability, and the specific rates of translation initiation, elongation, and termination. These six parameters are regulated by elements within 5' and 3' UTRs and the product coding sequence (Van Der Kalen *et al*., 2009; Jackson *et al*., 2010).

Under normal cellular conditions the first step in translation, ribosome-binding, is mediated by 5' mRNA caps. However, under cell stresses commonly found (e.g. hypoxia, stationary phase growth) and specifically employed

(hypothermia) during manufacturing processes, cap-dependent translation is repressed (Al-Fageeh *et al*., 2005). Internal ribosome entry sites (sequences able to specifically recruit ribosomes via cap-independent mechanisms) can be included within 5' UTRs to maintain translation rates throughout bioreactor production (Lee *et al*., 2005; Koh *et al*., 2013). 5' UTRs also typically contain optimised Kozak sequences to facilitate high-levels of translation initiation (Rita Costa *et al*., 2010). Further, introns are routinely employed to confer efficient mRNA export, cytoplasmic localisation and increased stability against degradation (Bicknell *et al*., 2012). As intragenic introns are undesirable due to potential formation of splicing variants (Zago *et al*., 2009) production vectors typically contain a single intron located within the 5' UTR (Ng *et al*., 2010; Skoko *et al*., 2011). The adenovirus tripartite leader sequence, which can both increase mRNA stability/ export and facilitate cap-independent translation initiation, is a further CHO-characterised 5' UTR element that can be used either in conjunction with, or as an alternative to, introns (Mariati *et al*., 2010). Multiple desirable 5' UTR components are often incorporated in a single cis-regulatory element by utilising 5' UTRs from viral genes (frequently hCMV-IE1) (Ng *et al*., 2010).

There are comparatively fewer component options for inclusion in 3' UTRs. The two most commonly employed elements are the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element which directs efficient mRNA export, and SV40 polyadenylation signals that function to increase translation initiation and inhibit mRNA degradation (Wulfhard *et al*., 2008; Rita Costa *et al*., 2010). Product coding sequences themselves significantly affect translation rates and are routinely optimised *in silico* using bioinformatic tools to remove cryptic splice sites, optimise codon usage, eliminate mRNA destabilising motifs, and reduce CpG content (Kalwy *et al*., 2006; Bauer *et al*., 2010; Fath *et al*., 2011). In the near-future next-generation translation control systems may enable sophisticated, precise regulation of translation rates that can be optimised specifically for each product. For example, Ferreira *et al.* describe the production of a synthetic RNA element library, comprising short upstream open reading frames (ORFs) and varying translation initiation sites, that facilitates user-defined translation levels of individual genes (Ferreira *et al*., 2013).

1.4.2. Defining the current toolbox: expression stability enhancing elements

Recombinant CHO clones typically exhibit low and/ or unstable product gene expression as a functional consequence of the local chromatin structure at the site of integration (SOI) (Zhu, 2012). Heterochromatic regions, characterised by highly condensed structures and unfavourable post-translational modifications of core histone tails (e.g. deacetylation, dephosphorylation, methylation and ubiquitylation), are inaccessible to the transcriptional machinery and therefore repressive to transgene expression (Huisinga, 2006). The vast majority of integration events are within these transcriptionally non-permissive regions, necessitating the screening of several thousand clones to isolate rare high-producers (Lai *et al*., 2013). Further, (and perhaps more importantly, given the development of high-throughput clone screening technologies) expression in high-producers is often silenced over time due to the spread of nearby heterochromatin regions (*i.e.* positional effect variegation) (Girod *et al*., 2005; Kim *et al*., 2012a). Moreover, even relatively minor unfavourable local chromatin structure remodifications can significantly impact protein titers by reducing the frequency of transcriptional bursts (Pilbrough *et al*., 2012).

Elements that maintain integrated gene copies in transcriptionally active genomic regions are increasingly being utilised in production vectors to i) increase the occurrence of high-producing clones, ii) enhance productivity of top performing clones, iii) facilitate long-term expression stability, and iv) enable copy-number-dependent expression levels, with increased productivity per copy number. Strategies are centred on either forming and maintaining euchromatic regions at the SOI, or specifically targeting the expression cassette to transcriptionally permissive sites in the CHO genome. The former most commonly employ scaffold/ matrix attachment regions (S/MARs), sequences that naturally occur at the boundaries of transcriptionally active genomic regions. S/MARs prevent heterochromatic spread by acting as boundary elements and promote localised euchromatin formation by recruiting histone acetyltransferases, chromatin remodelling complexes and components of the transcriptional machinery (Girod *et al*., 2007). Multiple endogenous and exogenous S/MARs have been shown to be extremely effective at promoting high-level, stable expression in CHO cells (Kwaks *et al*., 2003; Wang *et al*., 2008; Wang *et al*., 2012). Examples of the latter strategy most commonly

employ recombination mediated cassette exchange (RMCE), whereby both the identified genomic integration site and the expression cassette are tagged with the same specific recombinase recognition sequences. Whilst this approach is limited by the availability of genomic hotspots and the requirement to isolate new host cell lines, RMCE has been shown to produce stable, highly productive clones (Nehlsen *et al*., 2009; Zhou *et al*., 2010). Moreover, availability of genomic and transcriptome data should significantly improve RMCE in CHO cells by enabling identification of novel integration loci (Xu *et al*., 2011). Future vectors will likely benefit by combining both strategies (*i.e.* RMCE utilising S/MAR containing expression cassettes) to facilitate both insertion into, and maintenance of, transcriptionally-active chromatin regions.

1.4.3. Defining the current toolbox: transcriptional regulatory elements

Product gene transcription is the first step in protein expression and a critical parameter affecting productivity. Multiple studies evaluating the relative control that discrete cellular processes exert on production have identified transcription rate as a major rate limiting factor (Mead *et al*., 2008; Chusainow *et al*., 2009; McLeod *et al*., 2011; O'Callaghan *et al*., 2011). Transcription rates are controlled by gene-specific regulatory elements, comprising core promoters, proximal/ distal promoters and enhancers. Transcriptional output of a given gene is functionally determined by interactions between transcription factors (TFs) and their cognate binding sites (transcription factor regulatory elements (TFREs)) within these elements to overcome multiple rate-limiting process steps (Coulon *et al*., 2013). TFs function as activators or repressors to either relieve or impose regulation at each step by interacting with, and/ or recruiting accessory proteins to, the transcriptional machinery. The mechanistic function of TF-TFRE interactions has high-level complexity owing to the fact that i) TF expression and activity is highly regulated, ii) individual TFs can function as either activators or repressors depending upon cellular state and iii) spatial positioning of TFREs within promoters, and combinatorial interactions between locally bound TFs, significantly affect TF activity (Fuda *et al*., 2009). The rate of transcription of a given gene at a specific timepoint is therefore determined by the TFREs present in its promoter, their combinatorial and spatial

configurations, and the complement of TFs active in the cell. Promoter components for CHO cells fall into three discrete categories; viral, endogenous, and synthetic.

1.4.3.1. Viral promoters

Cell permissiveness to viral infection is often determined by virus promoter activity. Promiscuous viruses have accordingly evolved complex promoters containing multiple discrete TFREs to broaden their host cell ranges (Stinski and Isomura, 2008). Specifically designed to access TFs that are active in most cell types, viral promoters exhibit high activity in most mammalian cell lines. Unsurprisingly, they are therefore a popular choice for recombinant gene expression in divergent cell types, including CHO, and many production vectors utilise strong viral promoters to control product gene transcription (Birch and Racher, 2006; Rita Costa *et al*., 2010; Datta *et al*., 2013).

Historically, the hCMV-IE1 promoter has been utilised to drive constitutive strong expression in the vast majority of production vectors (Birch and Racher, 2006; Rita Costa *et al*., 2010; Datta et al, 2013). However, both mouse and rat CMV-IE1 promoters (mCMV-IE1, rCMV-IE1) have been shown to exhibit higher activity than hCMV-IE1 in transient and stable CHO cell transfectants (Xia et al 2006). Accordingly, mCMV-IE1 has been employed in some recent production vectors and has been adopted by at least one well-known contract manufacturer. Although infrequently utilised, multiple other viral promoters also have characterised high-activity in CHO cells, including simian virus 40 early promoter and enhancer (SV40E), adenovirus major late promoter, myeloproliferative sarcoma virus long terminal repeat (LTR), rous sarcoma virus LTR, and human immunodeficiency virus LTR (Makrides, 1999; Pontillier *et al*., 2008). All of these promoters must contain one or more constituent TFREs able to bind cognate TFs present in CHO cells. However, little is known about how they function mechanistically in the CHO cell and therefore strategies to precisely control or improve their transcriptional activity are not available. This is particularly problematic given that they are often associated with a number of undesirable bioproduction characteristics, such as induction of cellular stress and activation of apoptotic pathways (Pontillier *et al*., 2008). Further, many (including hCMV-IE1) have been shown to be cell cycle dependent, with greatest activity during S-phase; potentially limiting given that in a fed-batch

production over half of the integral cell area exists in the stationary phase (Dale, 2006; Prentice and Tonkin, 2007). Moreover, hCMV-IE1 is known to contribute to production instability via promoter methylation and gene deletion (Yang *et al*., 2010; Kim *et al*., 2011). Accordingly, viral promoters are increasingly being considered as unsuitable for manufacturing processes and in recent years many companies have replaced them with endogenous promoters.

1.4.3.2. Endogenous promoters

Mammalian cells require constitutive expression of multiple genes to maintain cellular homeostasis. Transcriptional regulation of these 'house-keeping' genes is accordingly controlled by constitutively active promoters (Schwanhäusser *et al*., 2011). Endogenous CHO 'house-keeping' gene promoters have therefore specifically evolved to deliver constitutive, stable expression in the host cell factory. Although they are generally not as active as viral alternatives their expression dynamics are therefore better suited to manufacturing processes, making endogenous promoters attractive candidates for controlling product gene transcription.

Strategies to isolate novel endogenous CHO promoters have involved i) utilising regulatory sequences from highly expressed genes, ii) employing promoters specifically suited to bioprocess conditions, and iii) screening the CHO genome for transcriptionally active elements. A successful application of strategy 1 showed that regulatory elements from the Chinese hamster elongation factor-1α (CHEF-1α) gene significantly increased (6 – 35 fold) expression of several proteins in stable CHO cell clones, relative to hCMV-IE1 (Running Deer and Allison, 2004). The CHEF1α promoter has since been utilised in many production vectors. More recently, Le *et al.* utilised transcriptomics data to identify CHO endogenous promoters with desired expression dynamics (Le *et al*., 2013). Whilst this is a promising avenue to identify promoters with discrete expression levels, it can be a significant challenge to define the genomic regulatory sequences controlling expression of specific genes. Strategy 2 has been utilised to isolate promoters exhibiting high activity during stationary phase bioreactor cell growth in order to synchronise expression with the production phase (for example, ferritin heavy chain and growth arrest and damage inducible 53 promoters) (de Boer *et al*., 2004; Prentice and Tonkin, 2007). Further, endogenous promoters highly active at 32°C have been described to optimise expression rates

during biphasic bioreactor processes (Thaisuchat *et al*., 2011). The third strategy has predominantly involved randomly inserting CHO genome fragments upstream of reporter genes and screening for protein expression to identify transcriptionally active sequences (Pontiller *et al*., 2008; Pontiller *et al*., 2010; Chen *et al*., 2013). These non-targeted approaches have isolated low-activity elements that may have application when high-level product expression is cytotoxic.

Whilst they offer some improved functionalities over viral alternatives, endogenous promoters have two significant limitations. Firstly, they exhibit inefficient transcriptional activity per unit DNA sequence (CHEF1α is > 15x larger than hCMV-IE1), often requiring extensive 5' and 3' sequence extensions to function effectively (likely due to MAR-like boundary elements within these regions). Secondly, and far more importantly, they are blunt-tools evolved for divergent functions that are extremely unlikely to enable optimal product gene transcription levels. Viral and endogenous promoters provide unpredictable, uncontrollable functionality and cannot facilitate the sophisticated optimisable transcriptional control that is required to achieve many critical biopharmaceutical manufacturing developments.

1.4.3.3. Synthetic promoters

Synthetic promoters can be defined as novel non-naturally occurring sequences capable of providing both i) significantly higher transcription levels than exist in nature, and ii) controllable activity, specifically tailored to desired outputs. Synthetic promoters therefore offer a potentially attractive solution in CHO cells, as they can replace functionally ill-defined and uncontrollable genetic elements in expression vectors with sophisticated, bespoke controllers with predictable function.

Synthetic promoters have been traditionally produced by screening either (i) randomised DNA sequences or synthetic oligonucleotide repeats or (ii) assemblies of known *cis*-regulatory elements (e.g. TFREs), both upstream of minimal core promoter motifs. Studies have demonstrated that the exquisite gene expression control offered by synthetic promoters is highly context-dependent, necessitating promoters to be specifically constructed for each host cell type (Schlabach *et al*., 2010). Whilst synthetic promoter libraries have been designed to function in a range of microbial bioproduction hosts, such as *Corynebacterium* (Yim *et al*., 2013),

Saccharomyces (Blazeck *et al*., 2012), *Pichia* (Stadlmayr *et al*., 2010) and *Streptomyces* (Seghezzi *et al*., 2011), and also in certain mammalian cells (Ferreira *et al*., 2011; Ogawa *et al*., 2007; Schlabach *et al*., 2010), few studies have previously explored the utilisation of synthetic promoters in CHO. Hitoshi *et al.* engineered a single synthetic hybrid by fusing the chicken AG and hCMV-IE1 promoters (Hitoshi *et al*., 1991). On a larger scale, Tornøe *et al.* created a small pool of promoters (< 20) with a tenfold range in activity by randomising the sequences separating TFREs within a chimeric promoter (Tornøe *et al*., 2002). Further, Grabherr *et al.* constructed five synthetic promoters with approximately equivalent activities by constructing contiguous sequences with nucleotide compositions mimicking those found in highly active promoters (Grabherr *et al*., 2011). Both of these studies targeted broad activity in mammalian cells (tested in multiple diverse cell lines) and neither sought to specifically design synthetic promoters to function in concert with the transactivational machinery of CHO cell factories. Accordingly, neither of these small libraries enable predictable, precise, robust control of CHO gene expression over broad dynamic ranges and synthetic promoter activities reported were significantly below that of hCMV-IE1.

Ligand-controllable synthetic gene switches are an alternative synthetic transcriptional control strategy that has seen considerable developmental effort. Typically these systems comprise a synthetic TF (comprising a mammalian transactivation domain fused to a prokaryotic response regulator), a synthetic promoter (comprising cognate TFREs for the synthetic TF), and an inducer molecule that modulates TF-TFRE binding kinetics to switch transcription on/ off (Tigges and Fussenegger, 2009; Weber and Fussenegger, 2010). Multiple such systems, utilising variable TFs, TFREs and inducers, have been described in CHO. Early systems were restricted by employing inducer molecules that were incompatible with bioreactor manufacturing due to adverse cellular effects, difficult downstream processing steps and regulatory concerns (e.g. antibiotics, hormones, metal ions) (Walter *et al*., 1991; James *et al*., 2000; Fux and Fussenegger, 2003). More recent systems have addressed these issues by utilising common regulatory-approved media components as inducer molecules, such as amino acids, gases and vitamins (Hartenbach *et al*., 2007; Weber *et al*., 2009; Gitzinger *et al*., 2012). Modern systems also benefit from optimised expression control dynamics with improved inducer responsiveness and more rapid on/ off switching. However, whilst these systems offer on/ off (and a

degree of ligand-titratable) transcriptional control, their implementation is inherently relatively complex, requiring coordinated expression of the synthetic TF and the engineered target gene construct and the inducer molecule. Moreover, precision control over a broad dynamic range is currently unachievable and all systems described to date are significantly weaker than hCMV-IE1. They are therefore an inherently less robust solution than the use of synthetic promoters.

1.5. Innovative gene expression control technology could enable multiple critically required biomanufacturing developments

This thesis has thus far presented the following facts: 1) biopharmaceutical production is of considerable importance from both a human and economic perspective, 2) CHO cells are the major biologics production cell host of the present and near-future, 3) all bioproduction process steps present opportunities for disruptive step change innovation, 4) this innovation is urgently required to meet demands from patients, industry and key decision makers, and 5) current transcriptional control technology represents a suboptimal, productivity-limiting manufacturing component. This section ties a thread between these points by detailing how sophisticated transcriptional control could facilitate development of next-generation biopharmaceutical manufacturing systems. This blueprint specifically outlines the opportunities available at each process step, the innovative promoter technologies required to exploit them, and the critical biomanufacturing developments that could be enabled.

1.5.1. Opportunity 1: engineering next-generation synthetic CHO cell factories

Process: upstream process 1 - cell factory selection.
Current limitation: current cell factories exhibit unoptimised bioproduction functionalities and are unfit-for-purpose.
Potential solution: engineer next-generation synthetic cell factories specifically optimised for biopharmaceutical production

Required technology: CHO-specific sophisticated gene expression control technology enabling predictable, controllable transcription regulation over a broad dynamic range.
Critical development enabled: increased cell factory capability

CHO cell factories are unoptimised for biopharmaceutical production and exhibit productivity-limiting performance in critical functionalities such as proliferation, apoptosis, protein folding and glycosylation (Kim *et al*., 2012a; Datta *et al*., 2013). To date, improvements in factory performance have been achieved via directed evolution and mutagenesis - for example adaptation to serum-free suspension culture (Hacker *et al*., 2009). Whilst these approaches enable incremental advancements, they are uncontrollable and unpredictable, and it is therefore widely acknowledged that next-generation cell factory engineering will require implementation of synthetic biology strategies (Pasotti *et al*., 2012; Esvelt and Wang 2013).

Previous CHO cell synthetic engineering attempts have predominantly involved tailoring the expression of single genes. Whilst this approach has seen some success, for example the Glycart-Roche and Potelligent glyco-engineering technologies, these simple, small-scale attempts have unsurprisingly proved largely ineffective (Umaña *et al*., 1999; Yamane-Ohnuki *et al*., 2004; Kildegaard *et al*., 2013). Tinkering with a few components from amongst thousands within the cell factory's incredibly complex design space is unlikely to provide optimised functions. Biopharmaceutical production involves thousands of endogenous proteins in multiple cellular pathways and next-generation CHO cell factories will accordingly require sophisticated multigene engineering strategies where the expression of multiple genes is stoichiometrically balanced (Dinnis and James, 2005; Guye *et al*., 2013).

Large-scale synthetic engineering requires seven key enabling technologies: genomics (Lewis *et al*., 2013), transcriptomics/ proteomics/ systems modelling (Kildegaard *et al*., 2013), DNA synthesis (Ma *et al*., 2012), multigene engineering systems (Torella *et al*., 2014), high-throughput screening tools (Lai *et al*., 2013) and sophisticated gene expression control technology. Of these, only the latter is currently unavailable for CHO cells. Stoichiometrically balancing the expression of multiple genes will require the ability to select (or rapidly design) promoters with appropriate desired activities. Predictable control of transcriptional activity over a

broad dynamic range would facilitate construction of completely novel design spaces with bioproduction-optimised cellular functionalities (e.g. factories that are resistant to apoptosis, have increased viable cell densities, and provide optimised product glycosylation). Currently unfit-for-purpose CHO cell factories could accordingly be demolished and rebuilt according to specific design instructions. Synthetic cell factories, optimised for biomanufacturing and operating at maximal productivity, could potentially enable all ten of the critical bioproduction developments detailed in section 1.3.4 to facilitate manufacture of better drugs, cheaper and faster.

1.5.2. Opportunity 2: enabling product-specific optimised transcription rates

Process: upstream process 2 - vector engineering.
Current limitation: product gene transcription rates routinely limit productivity.
Potential solution: specifically optimise the transcription rate of each recombinant protein.
Required technology: system-specific sophisticated gene expression control technology enabling precise, predictable transcription control over a broad dynamic range.
Critical development enabled: increased titers via improvements in specific productivity.

Recombinant gene transcription has repeatedly been identified as a rate limiting factor in biopharmaceutical production. Productivity is a function of multiple discrete cell factory processes (*i.e*. transcription, translation, protein folding, assembly, glycosylation, and secretion) and therefore optimal transcription rates are protein-specific, dependent upon the factory's capacity to perform downstream processes. For ETE proteins, where transcription rates have been shown to exert a high level of control over production, transcription rates are typically too low, failing to take full advantage of downstream factory capability (McLeod *et al*., 2011; O'Callaghan *et al*., 2010). For DTE proteins, where maximising transcription is unlikely to be beneficial as ER folding and assembly processes assume proportionately greater control over synthetic flux as proteins become more DTE, transcription rates are typically too high, inducing downstream bottlenecks leading to unfolded protein responses, cellular stress and apoptosis (Girod *et al*., 2013; Le

Fourn *et al*., 2014). For mAbs, where light chain : heavy chain transcription rate ratios are a critical parameter effecting productivity, transcription ratios are typically suboptimal resulting in inefficient folding and assembly (Ho *et al*., 2013; Pybus *et al*., 2014). Clearly, maximising productivity of most biopharmaceuticals will require the ability to provide optimised recombinant gene transcription rates that are kinetically coordinated with polypeptide-specific downstream process rates. However, this is currently intractable given that available promoters are restricted to enabling a very limited set of discrete activities. Libraries of CHO-specific promoters with hundreds (or indeed thousands) of discrete activities covering a broad dynamic range would facilitate protein-promoter matchmaking to specifically optimise transcription rates for each product. Such technology could potentially be utilised to increase productivity, and titers, for the vast majority of biopharmaceuticals.

1.5.3. Opportunity 3: maximising transient expression yields

Process: upstream process 3 – transient expression
Current limitation: low recombinant gene expression levels limit productivity
Potential solution: increase recombinant gene transcription rate
Required technology: promoters with significantly increased activity compared to existing alternatives.
Critical development enabled: higher yields from transient production systems

TGE is utilised to rapidly produce grams of developmental material for use in toxicology studies, process development, and product quality evaluation (Baldi *et al*., 2007). Moreover, significant time and cost advantages over stable expression systems have seen growing interest in utilising TGE for large-scale manufacturing. If regulatory concerns surrounding protein quality, process consistency and product safety can be addressed, TGE could potentially enable significantly increased manufacturing flexibility (Cain *et al*., 2013). However, despite significant progress, TGE systems currently provide relatively poor yields, where low recombinant gene expression levels are a critical productivity-limiting factor (Zhu *et al*., 2012). Currently available promoters (the vast majority of these systems utilise hCMV-IE1) are therefore incapable of maximising transcription rates in these systems.

Availability of new promoters with significantly increased activity compared to the current TGE promoter of choice (hCMV-IE1) would enable increased transgene expression rates and facilitate significant improvements in transient production yields. This would further reduce the cost and time associated with production of developmental material and, more importantly, would be a significant step towards enabling TGE as a method for large-scale therapeutic protein manufacturing.

1.5.4. Opportunity 4: shortening the cloning process

Process: upstream process 4 – cloning.
Current limitation: isolation of stable high-producing clones is time-intensive and significantly slows speed-to-market.
Potential solution: increase the frequency of high-producers and minimise expression instability.
Required technology: promoters that enable optimised selection marker transcription rates and improved product gene expression stability.
Critical development enabled: faster product speed-to-market

Selection marker gene expression levels are a key determinant of cloning efficiency (Lai *et al*., 2013). Ideally, selection gene expression would be optimised to enable specific selection of high producing clones containing either multiple gene copy numbers or integration events in highly transcriptionally active genomic loci. This would require 'Goldilocks' transcription rates that are high enough to prevent survival of low-producers and low enough to allow survival of sufficient clone numbers. However, this level of control is currently unachievable and the vast majority of production vectors utilise the relatively weak SV40E promoter to regulate selection marker gene transcription. Predictably, this element provides unoptimised functionality, routinely resulting in low average clone productivity and necessitating both multiple rounds of gene amplification and/ or screening of several thousand clones (De Jesus *et al*., 2011; Agrawal *et al*., 2012).

The ability to predictably and precisely control transcription in CHO cells would enable optimised selection marker gene expression to increase both average clone productivity and the frequency at which high-producers occur. Moreover, if bespoke promoters could be engineered to remove expression instability issues

associated with current product gene promoters (e.g. hCMV-IE1 methylation) the failure-rate of late-stage clones may be reduced, negating the requirement for time-consuming stability studies (Li *et al*., 2010; Kim *et al*., 2011). Sophisticated transcriptional control technology could therefore potentially be utilised to significantly improve cloning efficiency and enable isolation of higher producers in shorter timescales.

1.5.5. Opportunity 5: co-ordinate product gene expression with bioreactor conditions and processes.

Process: upstream process 5 – manufacturing process development.
Current limitation: product gene expression dynamics are not ideally suited to bioreactor conditions.
Potential solution: synchronise expression levels with proliferation/ production phases.
Required technology: promoters specifically designed to function optimally in bioprocess conditions.
Critical development enabled: increased titers via improvements in specific productivity and IVCD.

Protein yields are optimised when cell biomass is rapidly accumulated (proliferation phase) and subsequently maintained for extended periods of time (production phase) (Hacker *et al*., 2009; Li *et al*., 2010). Accordingly, product gene expression would ideally be minimised during the proliferation phase (to maximise cell growth rate) and maximised during the production phase. However, currently available promoters do not facilitate biphasic expression kinetics, for example hCMV-IE1 is preferentially active during proliferation phases (Dale, 2006). Ligand-controllable synthetic gene switches may potentially offer attractive solutions but inherent complexity and relatively low expression levels limit their utilisation (Weber and Fussenegger, 2010). Similarly, endogenous promoters preferentially active during production phases have been described but typically provide insufficient levels of activity (Prentice and Tonkin, 2007). Moreover, their functionality is ill-defined, preventing exploitation or control of desirable qualities. Synthetic promoters could be specifically designed to provide optimal transcription dynamics by constructing

them from *cis*-regulatory elements that are preferentially active during production phases. Similarly, promoters could be further refined for optimal functionality both during desirable bioprocess conditions, such as hypothermia, and common stresses, such as hypoxia (Masterton and Smales, 2014). Utilising bioreactor-optimised, specifically engineered promoters could both increase product titers and enable implementation of desirable bioreactor conditions that may be currently intractable.

1.6. Thesis overview

As outlined in section 1.5, innovative transcription control technologies have the potential to deliver multiple critically required biomanufacturing developments. The work in this thesis now describes the development of three such technologies that cumulatively enable sophisticated, next-generation transcriptional control in CHO cell factories.

Chapter two details the materials and methods that were used in this study. Chapter three presents the first-ever synthetic promoters designed specifically to regulate recombinant gene transcription in CHO cells. TFRE function in CHO cells was functionally screened and active elements were utilised to construct libraries of synthetic promoters that exhibited variable activity over two orders of magnitude, significantly exceeding that of hCMV-IE1. Promoter functionalities were validated in both different CHO host lines and through a fed-batch transient production process. The precision control of recombinant gene transcription enabled by these synthetic promoters will facilitate the design of novel, predictable synthetic constructs for diverse CHO cell engineering applications.

Chapter four presents a novel form of TF decoys that are specifically designed for use in CHO cells (TF decoys are short synthetic oligodeoxynucleotides (ODNs) that sequester cognate transcription factors and prevent their binding at target promoters). A method was developed whereby blocks containing discrete transcription factor binding sites (TFRE-blocks) are combined into circular molecules, enabling rapid construction of chimeric decoys containing stoichiometrically optimised ratios of input TFRE-blocks. It was demonstrated that block-decoys were able to inhibit expression from multiple target elements simultaneously in CHO cells using a bespoke chimeric decoy. Enabling investigation

of any multi-TF mediated cell function or phenotype, block-decoys are a valuable new tool for characterising and controlling CHO cell transcription.

In chapter five the mechanistic functionality of the hCMV-IE1 promoter is systematically deconstructed to identify the discrete TFREs that control its activity in CHO cells. hCMV-IE1 has been employed in the vast majority of regulatory approved production vectors, is currently used to drive expression of many biopharmaceutical products, and is the current promoter of choice for TGE systems, yet surprsisingly little was previously known about how it functioned in CHO cells. *In silico* bioinformatics analysis (of both hCMV-IE1s TFRE composition and CHO cell's TF complement) and evaluation of the activity of discrete promoter regions identified a design space that was interrogated via TF sequestration (utilising block-decoys) and TFRE knockout (utilising synthetic CMV constructs with scrambled TFRE sequences). It was determined that i) the vast majority of CMVs activity within CHO cells is dependent upon just two TFREs, NFkB and CRE, and ii) YY1 is a negative regulator of the CMV promoter in CHO. This mechanistic understanding of hCMV-IE1's functional regulation enables strategies to predictably control and improve the activity of this commonly-utilised genetic component by engineering the promoter's TFRE composition or the cell factory's TF abundances.

Chapter 2: Materials and Methods

This chapter details the materials and methods that were used in this study.

2.1. Vector construction

2.1.1. Promoterless reporter vectors

A promoterless vector was subcloned from pSEAP2control (Clontech, Oxford, UK) by PCR amplification of appropriate vector regions with Phusion high fidelity polymerase (New England Biolabs, Hitchin, UK) using the following primer pairs (Sigma, Poole, UK): PVForward1: 5'-ATGCC TCGAG CTTCG AATCG CGAAT TCG and PVReverse1: 5'-AGTCA AGCTT TTACC ACATT TGTAG AGG; PVForward2: 5'-ATGCA AGCTT ACCGG TGGAT CCGTC GACCG ATGC and PVReverse2: 5'-ATGCA ATGAT CTCGA GCCCG GGCTA GC. Amplified fragments were digested with XhoI and HindIII (Promega, Southampton, UK), gel extracted (Qiaquick gel extraction kit; Qiagen, Crawley, UK) and ligated with T4 DNA ligase (Life Technologies, Paisley, UK). Recombinant plasmid was amplified in *E. coli* DH5α cells and purified using a Qiagen plasmid mini kit (Qiagen). Finally, correct plasmid construction was confirmed by restriction analysis and DNA sequencing. A second reporter plasmid was created by replacing the secreted alkaline phosphatase (SEAP) open reading frame (ORF) with the turbo green fluorescent protein (GFP) ORF. Unless otherwise stated, all vectors in this study were constructed and analysed using the same methodology.

2.1.2. Core promoter-reporter vectors

The core promoter sequences shown in Table 2.1 were synthesised (Sigma), PCR amplified, and cloned into XhoI and EcoRI sites directly upstream of the SEAP ORF in the promoterless SEAP-reporter vector.

Table 2.1: DNA sequences of core promoter constructs used in this study

Core promoter construct	Sequence (5' - 3')
CMV	AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGC CTAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGAC
CHEF1α	GAACGGTATAAAAGTGCGGTAGTCGCGTTGGACGTTCTTTTTCG CAACGGGTTTGCCGTCAGAACGCAGGTGAGTGGCGGGTTTG
Supercore	AGGCTTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGC CTAGATAGACATCCACGAGCGGACGTGCGTTTTAGAAGAC
Synthcore2	AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATTCTT CCGGATAGCTGTCTACGCTGTCAGCTGCTCCATAGAAGAC
TA box	AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGGTGACACGC CTAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGAC
INR	AGGTCACGTCCGTCAGAGCTCGTTTAGTGAACCGTCAGATCGCC TAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGAC
MTE.DPE	AGGCTACGTCCGTCAGAGCTCGTTTAGTGAACCGGTGACACGC CTAGATAGACATCCACGAGCGGACGTGCGTTTTAGAAGAC
DCE	AGGTCACGTCCGTCAGAGCTCGTTTAGTGAACCGGTGACATCTT CCGGATAGCTGTCTACGCTGTCAGCTGCTCCATAGAAGAC

2.1.3. TFRE-Reporter Vectors

Synthetic oligonucleotides containing 7x repeat copies of the TFRE consensus sequences in Table 2.2 were synthesised (Sigma), PCR amplified, and inserted into KpnI and XhoI sites upstream of the CMV core promoter in the promoterless SEAP and GFP-reporter plasmids.

2.1.4. Synthetic Promoter Libraries

Synthetic promoter building blocks were constructed from complementary single stranded 5' phosphorylated oligonucleotides (Sigma), annealed in STE buffer (100 mM NaCl, 50 mM Tris-HCl, 1 mM EDTA, pH 7.8, Sigma) by heating at 95°C for 5 min, prior to ramp cooling to 25°C over 2 h. Oligonucleotides were designed such

that the resulting double stranded blocks contained the specific TFRE (Table 2.2) and a 4 bp TCGA single stranded overhang at each 5' termini.

Table 2.2: DNA sequences of transcription factor regulatory elements (TFREs) used to construct TFRE-reporter vectors.

Transcription Factor Regulatory Element (RE)	**Sequence**
Activator protein 1 (AP1)	TGACTCA
$CC(A/T)_6GG$ element (CArG)	CCAAATTTGG
CCAAT displacement protein (CDP)	GGCCAATCT
CCAAT-enhancer binding protein alpha (C/EBPα)	TTGCGCAA
Cellular myeloblastosis (cMyb)	TAACGG
cAMP RE (CRE)	TGACGTCA
Elongation factor 2 (E2F)	TTTCGCGC
E4F1	GTGACGTAAC
Early growth response protein 1 (EGR1)	CGCCCCCGC
Estrogen-related receptor alpha RE (ERRE)	AGGTCATTTTGACCT
Enhancer box (E-box)	CACGTG
GATA-1 (GATA)	AGATAG
GC-box	GGGGCGGGG
Glucocorticoid RE (GRE)	AGAACATTTTGTTCT
Growth factor independence 1 (Gfi1)	AAAATCAAC
Helios RE (HRE)	AATAGGGACTT
Hepatocyte nuclear factor 1 (HNF)	GGGCCAAAGGTCT
Insulin promoter factor 1 (IPF1)	CCCATTAGGGAC
Interferon-stimulated RE (ISRE)	GAAAAGTGAAACC
Myocyte enhancer factor 2 (MEF2)	CTAAAAATAG
Msx homeobox (MSX)	CGGTAAATG
Nerve growth factor-induced gene-B RE (NBRE)	AAAGGTCA
Nuclear factor 1 (NF1)	TTGGCTATATGCCAA
Nuclear factor of activated T cells (NFAT)	AGGAAATC
Nuclear factor kappa B (NFκB)	GGGACTTTCC
Octamer motif (OCT)	ATTAGCAT
Retinoic acid RE (RARE)	AGGTCATCAAGAGGTCA
Yin yang 1 (YY1)	CGCCATTTT
Random 8mer (8mer)	TTTCTTTC

For example the sequences used for the NFκB-RE block were as follows (RE site underlined): 5'-TCGATGGGACTTTCCA-3' and 5'-TCGATGGAAAGTCCCA-3'. Synthetic promoter libraries were constructed by ligating block molecules at appropriate stoichiometric molar ratios with high concentration T4 DNA ligase (Life Technologies). A 'cloning-block' containing KpnI and XhoI sites was included in ligation mixes at a 1:20 molar ratio of the RE blocks. The ligated molecules were digested with KpnI and XhoI (Promega), gel extracted (Qiaquick gel extraction kit, Qiagen), and inserted upstream of the CMV core promoter in the promoterless SEAP reporter vector.

2.1.5. Plasmid-decoys

A minimal TFRE-acceptor vector containing only the sequences required for propagation in bacteria (pUC Ori, AMPr) and a multiple cloning site (MCS) was subcloned from pSEAP2control by PCR amplification of appropriate vector regions using the following primer pairs (Sigma): PDForward1: 5'-AAGTG AATTC AGAAT CAGGG GATAA CGCAGG and PDReverse1: 5'-TACGA AGTTT CATGA GACAA TAACC CTG; PDForward2: 5'-ATGCA AGCTT AGGTA CGGGA GGTAC TTGG and PDReverse2: 5'-ATCTG AATTC ATCTC GAGCC CGGGC TAGC. Synthetic oligodeoxynucleotides containing 7x repeat copies of target TFRE consensus sequences (TFRE-ODNs) were synthesised (Sigma), PCR amplified, and inserted into the MCS to construct TFRE-specific plasmid-decoys. The TFRE consensus sequences used to construct TFRE-ODNs were as follows: NFκB-RE, GGGACTTTCC and NFκB-RE-scrambled, AATCGCAAGT.

2.1.6. hCMV-IE1 promoter-variant reporter vectors

Discrete hCMV-IE1 promoter regions were PCR amplified using the primer pairs shown in Table 2.3 and inserted upstream of the CMV core promoter in the promoterless GFP-reporter vector.

Wild-type hCMV-IE1 (nucleotides 595 – 1194 in accession number M60321.1) and synthetic promoter-variants with varying TFRE sites 'knocked out' (*i.e.* scrambled) were synthesised (GeneArt, Life Technologies), gel extracted and cloned upstream of the SEAP ORF in the promoterless SEAP reporter-vector. TFRE

sites were scrambled as follows: NFkB, AATCGCAAGT; CRE, CTACTGTG; and YY1, TGTC.

Table 2.3: Primers used to amplify discrete hCMV-IE1 promoter regions. F = forward, R = reverse, TSS = transcriptional start site, CRM = Cis-regulatory module.

hCMV-IE1 promoter region	**Primer sequences (5' – 3')**
Distal + proximal (-560 to -50 relative to TSS)	F: TTATCTCGAGAGTTCATAGCCCATATATGGAGTTCC R: ATCGGAATTCGTCTTCTATGGAGGTC
Distal (-560 to -300)	F: TGGCGGTACCAGTTCATAGCCCATATATGG R: ATAACTCGAGCAGGCGGGCCATTTACCG
Proximal (-300 to -50)	F: TAAACTCGAGGCATTATGCCCAGTACATG R: ATCGGAATTCGTCTTCTATGGAGGTC
CRM 1 (-560 to –448)	F: ATATGGTACCGGGTCATTAGTTCATAGCC R: ATATCTCGAGTACGTCATTATTGACGTC
CRM 2 (-467 to –354)	F: ATATGGTACCTTGACGTCAATAATGACG R: ATATCTCGAGTACACTTGATGTACTGC
CRM 3 (-375 to –259)	F: ATATGGTACCCTTGGCAGTACATCAAG R: ATATCTCGAGAGTAGGAAAGTCCCGTAAGG
CRM 4 (-302 to –192)	F: ATATGGTACCCTGGCATTATGCCCAGTAC R: ATATCTCGAGCATTGGTGTACTGCCAAAAC
CRM 5 (-210 to –92)	F: ATATGGTACCTTTGGCAGTACACCAATG R: ATATCTCGAGTTTGGAAAGTCCCGTTG
CRM 6 (-143 to –31)	F: ATATGGTACCTGACGTCAATGGGAGTTTG R: ATATCTCGAGGACCTCCCACCGTACACG

2.2. Cell culture and transfection

CHO-S and CHO-K1 cells were cultured in CD-CHO medium (Life Technologies) supplemented with 8 mM and 6mM L-glutamine (Sigma) respectively. CHO-DG44 cells were cultured in CD-DG44 medium (Life Technologies) supplemented with 8 mM L-glutamine and 18 mL/L pluronic F68 (Life Technologies). All cells were routinely cultured at 37°C in 5% (v/v) CO_2 in vented Erlenmeyer flasks (Corning, Surrey, UK), shaking at 140 rpm and subcultured every 3-4 days at a seeding density

of 2 x 10^5 cells/ml. Cell concentration and viability were determined by an automated Trypan Blue exclusion assay using a Vi-Cell cell viability analyser (Beckman-Coulter, High Wycombe, UK). Two hours prior to transfection, 2 x 10^5 cells from a mid-exponential phase culture were seeded into individual wells of a 24 well plate (Nunc). Cells were transfected with DNA-lipid complexes comprising DNA and Lipofectamine (Life Technologies) at a 1: 3 ratio (µg/µl), prepared according to the manufacturer's instructions. Transfected cells were incubated for 24 – 72 h prior to protein expression analysis. To eliminate potential promoter-promoter interference, individual transfections did not include a co-transfected reporter vector to compare transfection efficiencies. However, each 24 well plate included a set of three external reporters (CMV-SEAP, SV40-SEAP and 7x CRE-SEAP) to confirm reproducible transfection performance.

2.2.1. Fed-batch transient transfection

Two hours prior to transfection 6 x 10^6 cells from a mid-exponential phase CHO-S culture were seeded into 50 mL CultiFlask bioreactors (Sartorius, Surrey, UK) at a working volume of 6 mL. Cells were transfected with DNA: lipofectamine complexes, prepared according to the manufacturer's instructions. Fed-batch cultures were maintained for seven days by nutrient supplementation with 10% v/v CHO CD Efficient Feed A (Life Technologies) on day 2, 4 and 6. SEAP expression and cell growth were measured at 24 h intervals.

2.3. Quantification of reporter expression

SEAP protein expression was quantified using the Sensolyte pNPP SEAP colorimetric reporter gene assay kit (Cambridge Biosciences, Cambridge, UK) according to the manufacturer's instructions. GFP protein expression was quantified using a Flouroskan Ascent FL Flourometer (Excitation filter: 485 nm, Emission filter: 520 nm). Background fluorescence/ absorbance was determined in cells transfected with a promoterless vector.

2.4. Bioinformatics analyses

2.4.1. *In silico* analysis of transcription factor regulatory elements

The following promoter sequences were retrieved from GenBank: hCMV-IE1 (accession number M60321.1), mouse CMV-IE1 (M11788), rat CMV-IE1 (U62396), guinea pig CMV-IE1 (CS419275), mouse CMV-IE2 (L06816.1), simian virus 40 early promoter and enhancer (NC_001669.1), adenovirus major late promoter (KF268310), myeloproliferative sarcoma virus long terminal repeat (LTR) (K01683.1), rous sarcoma virus LTR (J02025.1), and human immunodeficiency virus LTR (K03455.1). Promoters were analysed using the Transcription Element Search System (TESS: http://www.cbil.upenn.edu/cgi-bin/tess/tess) and the Transcription Affinity Prediction tool (TRAP: http://trap.molgen.mpg.de/cgi-bin/trap_form.cgi) according to the methods previously described by Schug (Schug 2008) and Manke et al. (Manke *et al*., 2008).

2.4.2. Analysis of hCMV-IE1's 4mer composition

To determine the relative abundance of every possible 4mer sequence, the hCMV-IE1 promoter was analysed using the Regulatory Sequence Analysis (RSA, http://rsat.ulb.ac.be) oligo-analysis tool (van Helden *et al*., 1998).

2.4.3. *In silico* analysis of synthetic promoter libraries

The relationship between synthetic promoters' discrete TFRE compositions and relative activities was analysed using the statistical analysis software R (R Development Core Team, 2013; http://r-project.org). The R script developed to execute this analysis (synpro.anal.R) utilised the following R packages (available at http://cran.r-project.org/web/packages): ggplot2, leaps, relaimpo, DAAG, RDCOMClient, R2wd, plyr, and stringr. This script can be downloaded at https://sourceforge.net/projects/syntheticpromoteranalysis/files/

2.5. Block-decoy construction and analysis

2.5.1. Construction of block-decoys

Regulatory element (RE)-block molecules were developed by annealing two complementary, single stranded 5' phosphorylated DNA ODNs (Sigma) in STE buffer (100 mM NaCl, 50 mM Tris-HCl, 1 mM EDTA, pH 7.8, Sigma). ODNs were heated at 95°C for 5 min and then ramp cooled to 25°C over 2 h to create RE-blocks that contain a transcription factor binding site and a 4 bp TCGA single stranded overhang at each 5' termini. RE-blocks (12 µg) were then ligated with 5 units of high concentration T4 DNA ligase (Life Technologies) at room temperature for 3 h to form block-decoys. Chimeric decoys were constructed by ligating varying molar concentrations of different RE-blocks. The sequences of all RE-blocks employed in this study are shown in Table 2.4.

2.5.2. Analysis of Block Decoy Structure

Block-decoy population size distribution was analysed by ethidium bromide agarose gel electrophoresis utilising molecular weight markers (Hyperladder II, Bioline, London, UK). To confirm block-decoys circularisation, 1.5 µg of purified block-decoy was added to 5 units of high concentration T4 DNA ligase before gel analysis. To test the stability of block-decoys against exonuclease, 4 µg of block-decoy was incubated with 300 units of Exonuclease III (Promega) and the mixture was incubated at 37 °C. A mixture of linear ODNs spanning the molecular weight range of the block-decoys was used as a positive control.

Table 2.4: DNA sequences of oligonucleotides used to construct TFRE-specific block-decoys. Complementary oligonucleotides were annealed to create RE-blocks that were subsequently ligated to form TFRE-specific block-decoys.

TFRE Target	**Oligonucleotide sequences (5' – 3')**
NFκB	TCGATGGGACTTTCCA TCGATGGAAAGTCCCA
CRE	TCGATTTGACGTCATT TCGAAATGACGTCAAA
E-box	TCGAAACACGTGAGA TCGATCTCACGTGTT
YY1	TCGATCGCCATTTTAA TCGATTAAAATGGCGA
8mer	TCGAAGTTTCTTTCGA TCGATCGAAAGAAACT
E4F1	TCGATTGTGACGTAACTT TCGAAAGTTACGTCACAA
GC-box	TCGATGGGGCGGGGA TCGATCCCCGCCCCA
RARE	TCGATAGGTCATCAAGAGGTCATT TCGAAATGACCTCTTGATGACCTA
C/EBPα	TCGATTTTGCGCAATT TCGAAATTGCGCAAAA
NF1	TCGATTTGGCTATATGCCAATT TCGAAATTGGCATATAGCCAAA
NFκB-scrambled	TCGATAATCGCAAGTA TCGATACTTGCGATTA
CRE-scrambled	TCGATTGACTAGAGTT TCGAAACTCTAGTCAA
E-box-scrambled	TCGAAAGCTCAGAGA TCGATCTCTGAGCTT

Chapter 3: Synthetic Promoters for CHO Cell Engineering

This chapter presents the first-ever synthetic promoters designed specifically to regulate recombinant gene transcription in CHO cells. TFRE function in CHO cells was functionally screened and active elements were utilised to construct libraries of synthetic promoters that exhibited variable activity over two orders of magnitude, significantly exceeding that of hCMV-IE1. Promoter functionalities were validated in both different CHO host lines and through a fed-batch transient production process. The precision control of recombinant gene transcription enabled by these synthetic promoters will facilitate the design of novel, predictable synthetic constructs for diverse CHO cell engineering applications.

Based on the work within this chapter the following article has been published (shown in Appendix C):

Brown AJ, Sweeney B, Mainwaring DO, James DC. 2014. Synthetic promoters for CHO cell engineering. Biotechnol. Bioeng. doi: 10.1002/bit.25227

Based on the work within this chapter the following patent application has been filed:

Brown AJ, James DC. UK patent application number GB 1321109.9 (29 November 2013) Synthetic promoters for CHO cells

3.1. Introduction

As discussed in section 1.5, novel transcription control technologies could enable disruptive step-change innovation in biopharmaceutical manufacturing. The precise transcriptional control required to implement multiple critically required biomanufacturing developments (outlined in section 1.3) is currently unachievable with viral and endogenous promoters that exhibit uncontrollable and unpredictable functionality (Pontillier *et al*., 2008; Kim *et al*., 2011; Datta *et al*., 2013). These naturally evolved blunt-instruments are restricted to offering a very limited set of

discrete activities. In contrast, synthetic promoters can be engineered to exhibit specifically tailored activities over broad dynamic ranges. They therefore represent an attractive solution for achieving sophisticated, predictable transcriptional control in CHO cells. Given that multiple studies have demonstrated synthetic promoter activity to be highly context (*i.e.* cell type) dependent, novel synthetic promoters need to be specifically designed to harness the transcriptional activation machinery of CHO cell factories (Ogawa *et al*., 2007; Schlabach *et al*., 2010).

Synthetic biology strategies aim to achieve desirable systems (or behaviours) by constructing novel devices with predictable functionality from characterised, modular parts (bioblocks) (Passoti *et al*., 2012). When applied to transcriptional control this typically involves constructing libraries of promoters (devices) that exhibit activities over broad dynamic ranges (behaviours) from parts that are either i) random synthetic oligonucleotides (e.g. 6 - 10 bp oligomers) (Ferreira *et al*., 2011), ii) known TFREs (Koschmannet *et al*., 2012), or iii) existing promoters (e.g. by producing mutagenised or hybrid variants) (Smith *et al*., 2013). This strategy has been employed to construct synthetic promoter libraries for a range of microbial protein production hosts, including *Saccharomyces* (Blazeck *et al*., 2012), *Streptomyces* (Seghezzi *et al*., 2011) and *E. coli* (De Mey *et al*., 2007), and has been particularly successful in *Pichia pastoris* where considerable synthetic promoter engineering efforts have facilitated improved bioprocesses, optimised expression levels and increased yields in the manufacture of industrial and therapeutic enzymes (Hartner *et al*., 2008; Ruth *et al*., 2010; Mellitzer *et al*., 2012; Vogl *et al*., 2013). Mammalian cell system synthetic promoters have predominantly been constructed by randomly assembling small numbers (2 – 6) of discrete system-active TFREs (although a few studies have utilised random synthetic oligonucleotides (Edelman *et al*., 2000; Schlabach *et al*., 2010)). Such promoters have typically been developed for use in cell type-specific gene therapy applications, for example in macrophages (He *et al*., 2006), neurons (Hwang *et al*., 2004), hepatocytes (Han *et al*., 2011), myocytes (Jianwei *et al*., 2012) and tumour cells (Xiong *et al*., 2011; Chen *et al*., 2012). Construction of most mammalian synthetic promoters has therefore required *a priori* knowledge of host system TFRE functionality - something that is currently unavailable for CHO cells.

Few studies have previously explored the utilisation of synthetic promoters in CHO (discussed in detail in section 1.4.3.3). Whilst two studies have characterised

the activity of small synthetic promoter libraries in CHO cells, both targeted broad mammalian cell activity (tested in multiple diverse cell lines) and neither sought to specifically design promoters to function in concert with CHO cell factory transactivational machinery (Tornøe *et al*., 2002; Grabherr *et al*., 2011). Cumulatively, these libraries contain a small number of promoters (< 25) that are only capable of controlling gene expression over a narrow range (< 10-fold) and have relatively low top activities (significantly less than hCMV-IE1). Accordingly, no promoters to date have enabled predictable, precise, robust control of CHO gene expression over a broad dynamic range.

The work in this chapter describes for the first time the creation of a library of 140 synthetic promoters specifically designed to regulate the expression of recombinant genes in CHO cells. TFRE function in CHO cells was functionally screened and active elements were utilised to construct libraries of synthetic promoters that exhibited variable activity over two orders of magnitude, significantly exceeding that of hCMV-IE1 in transient production processes. Moreover, relative promoter activities across a broad dynamic range were maintained in both different CHO cell hosts and across a fed-batch transient production process. The precision control of recombinant gene transcription enabled by this synthetic promoter technology will facilitate the design of novel, predictable synthetic constructs for diverse CHO cell engineering applications.

3.2. Results

3.2.1. Selecting a core promoter for use in synthetic promoter library construction

3.2.1.1 Identifying a core promoter design space

In order to construct CHO-cell specific synthetic promoters the activity of core promoter regulatory elements (CPREs) was first evaluated in CHO-S cells. CPREs within core promoter regions (typically -50 - +50 relative to the transcriptional start site (TSS)) bind cognate general TFs to form transcription preinitiation complexes (PICs). As transcriptional activators (both discrete TFREs and larger elements such

as proximal promoters and enhancers) have been shown to function specifically with different CPREs, it is optimal to comparatively evaluate different TFREs and synthetic promoters using a single fixed core promoter (Theisen *et al*., 2010). However, the design space for constructing synthetic core promoters is limited to the following seven previously described CPREs: TA box; initiator element (Inr); motif 10 element (MTE); downstream promoter element (DPE); TFIIB recognition elements (BREu/BREd); X core promoter element (XCPE1/2); and the downstream core element (DCE) (Juven-Gershon and Kadonaga, 2010). Moreover, the design space is further reduced as i) BRE and XCPE have been shown to exhibit repressive and unpredictable functionality and ii) each CPRE's functionality is dependent upon strict spatial positioning relative to the TSS (Juven-Gershon *et al*., 2008)). The positional requirements of the five CPREs (TA box, INR, DPE, MTE, and DCE) that can be utilised to construct synthetic core promoters are shown in Figure 3.1.

Multiple studies have demonstrated that CPREs function synergistically to increase transcription initiation rates (Juven-Gershon and Kadonaga, 2010). Accordingly, it was hypothesised that optimised core promoter functionality in CHO cells would be achieved by maximising CPRE numbers. Given that DCE inclusion prevents utilisation of DPE and MTE blocks (see Figure 3.1), the design space was narrowed to two synthetic core construct compositions; Ta box: INR: MTE: DPE and Ta box: INR: DCE. The former has previously been shown to be highly active in HeLa cells and is patented as the 'supercore' (Juven-Gershon *et al*., 2006), whilst the latter has never previously been constructed and will hereafter be referred to as synthcore2. It was predicted that these two synthetic cores would exhibit higher activities than naturally occurring alternatives. However, core promoters from the hCMV-IE1 (-34 to +48 relative to the TSS, containing a TA box and an INR, hereafter referred to as CMV core) and CHEF1α (-34 to +48 relative to the transcriptional start site, containing a TA box, hereafter referred to as CHEF1α core) promoters were evaluated in parallel as they have (indirectly) proven functionality in driving high levels of recombinant gene expression in CHO cells. It was hypothesised that this design space of four core promoters with varying CPRE compositions (Figure 3.1) would yield a 'part' with desirable, characterised functionality for inclusion in subsequent synthetic promoter libraries.

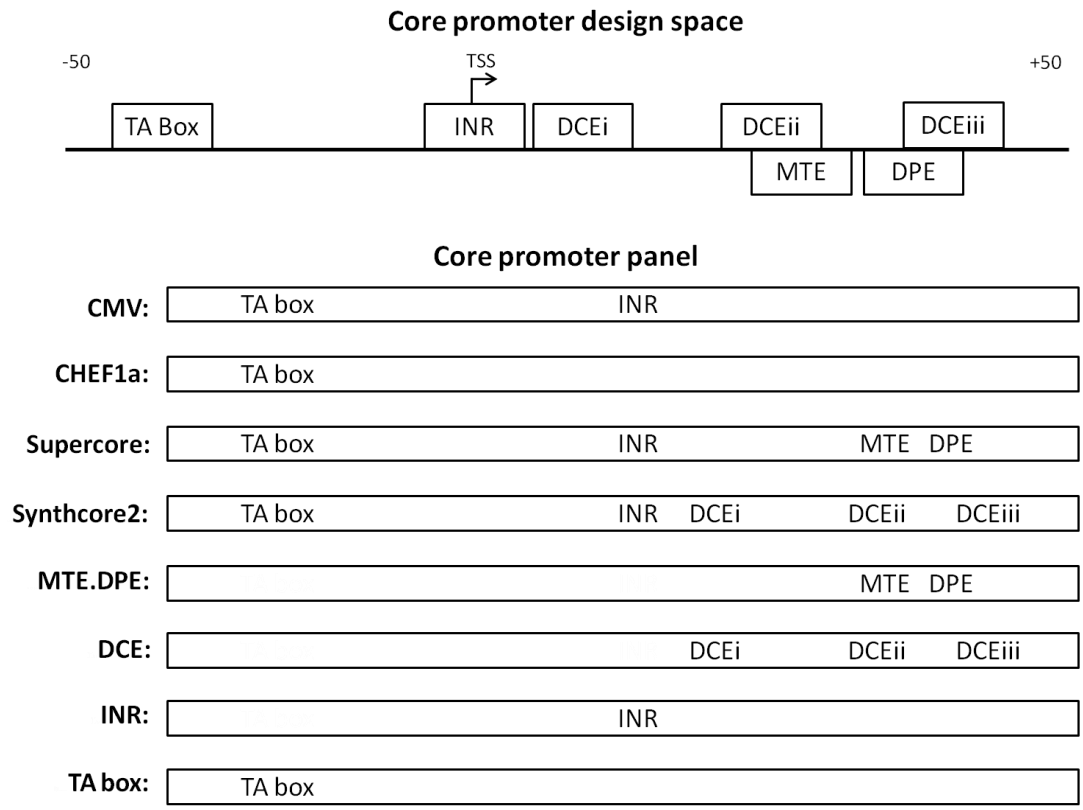


Figure 3.1: Core promoter engineering design space. The design space for core promoter construction is limited to five discrete core promoter regulatory elements (CPREs), with associated spatial positioning requirements. The varying CPRE compositions of the seven core promoters evaluated in this study are shown.

3.2.1.2 Different core promoters exhibit variable activities in CHO cells

In order to determine the relative transcriptional activity of core promoters in CHO-S cells reporter constructs containing each core upstream of a secreted alkaline phosphatase (SEAP) reporter gene were constructed. These vectors were otherwise 'promoterless' (i.e. devoid of proximal/ distal promoter sequences), and therefore SEAP expression levels were directly dependent on relative core activities. Measurement of SEAP reporter production after transient transfection of CHO-S cells with each core-reporter plasmid is shown in Figure 3.2. These data show that core constructs exhibit varying 'basal level' activities in CHO, where relative SEAP production from each core occurred at the ratio supercore 4.73: synthcore2 2.03: CMV 1.00: CHEF1α 0.33.

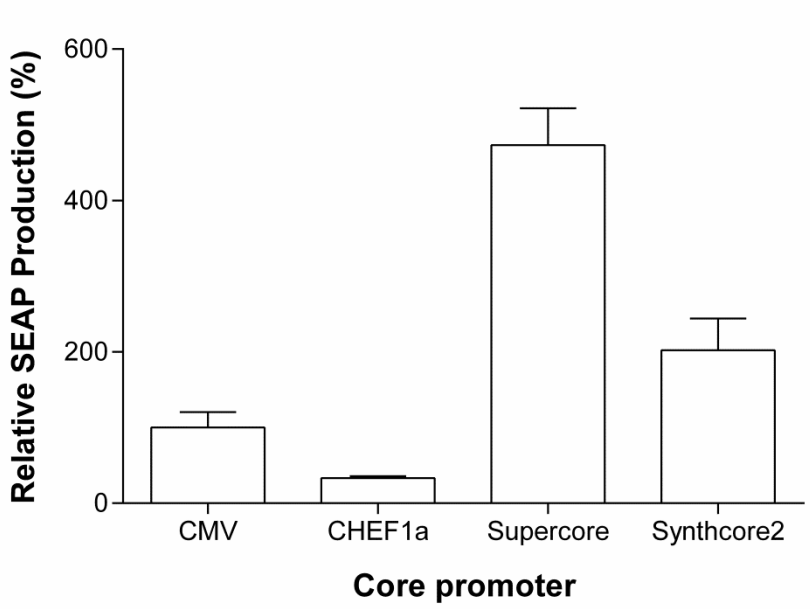


Figure 3.2: Core promoter constructs exhibit variable 'basal level' activities in CHO cells. Each core promoter construct (CPRE compositions shown in Figure 3.1) was cloned upstream of SEAP reporters in otherwise 'promoterless' vectors. CHO-S cells (2 x 10^5) in 24-well plates were transfected with 1 μg of core reporter-vector and SEAP activity in cell culture supernatant was measured 72 h post-transfection. Data are expressed as a percentage of the production exhibited by the CMV core promoter. Bars represent the mean + SD of three independent experiments each performed in triplicate.

Previous studies have demonstrated that MTE, DPE and DCE elements are incapable of initiating transcription in isolation, requiring either TA boxes or INRs for functionality (Lim *et al*., 2004). Evaluation of the activity of these CPREs in the absence of TA boxes and INRs (*i.e.* core promoters containing only MTE.DPE or DCE) confirmed that neither of these elements were active in CHO cells in isolation as both were incapable of driving detectable levels of SEAP production (Figure 3.3). It was therefore inferred that the higher activities exhibited by supercore and synthcore2, relative to CMV core, were a result of synergistic interactions between MTE/ DPE/ DCE elements and TA boxes/ INRs.

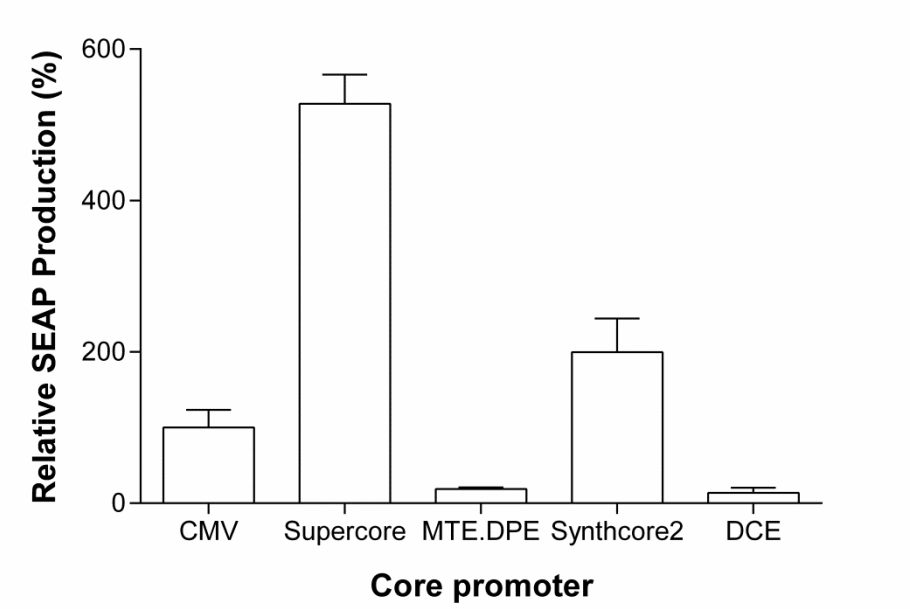


Figure 3.3: MTE, DPE and DCE elements are incapable of initiating transcription in isolation. Core promoter constructs containing only MTE.DPE or DCE elements were cloned upstream of SEAP reporters in otherwise 'promoterless' vectors. CHO-S cells (2 x 10^5) in 24-well plates were transfected with 1 μg of core reporter-vector and SEAP activity in cell culture supernatant was measured 72 h post-transfection. Bars represent the mean + SD of three independent experiments each performed in triplicate.

To determine if relative activities were maintained at higher levels of expression (*i.e.* > basal transcription levels) core promoter function was evaluated in the context of relatively weak (SV40E, - 155 to - 35 relative to TSS) and strong (hCMVIE1 distal + proximal promoters, -559 to -35 relative to the TSS) expression cassettes. This analysis showed that relative core promoter activities were not maintained in either SV40E-core-SEAP or hCMVIE1-core-SEAP reporters, where relative SEAP production from each core occurred at the ratios supercore 1.94: synthcore2 0.89: CMV 1.00: CHEF1α 0.48 and supercore 0.75: synthcore2 0.67: CMV 1.00: CHEF1α 0.38 respectively (Figure 3.4). These data reveal that the increased activities of supercore and synthcore2 relative to the CMV core at basal transcription levels are significantly reduced in weak SV40E expression cassettes. Moreover, in the context of the strong hCMV-IE1 cassette both cores exhibited reduced SEAP production relative to the CMV core. Therefore it was inferred that

relative core promoter functionalities are variable in CHO cells, dependent on surrounding expression cassette elements.

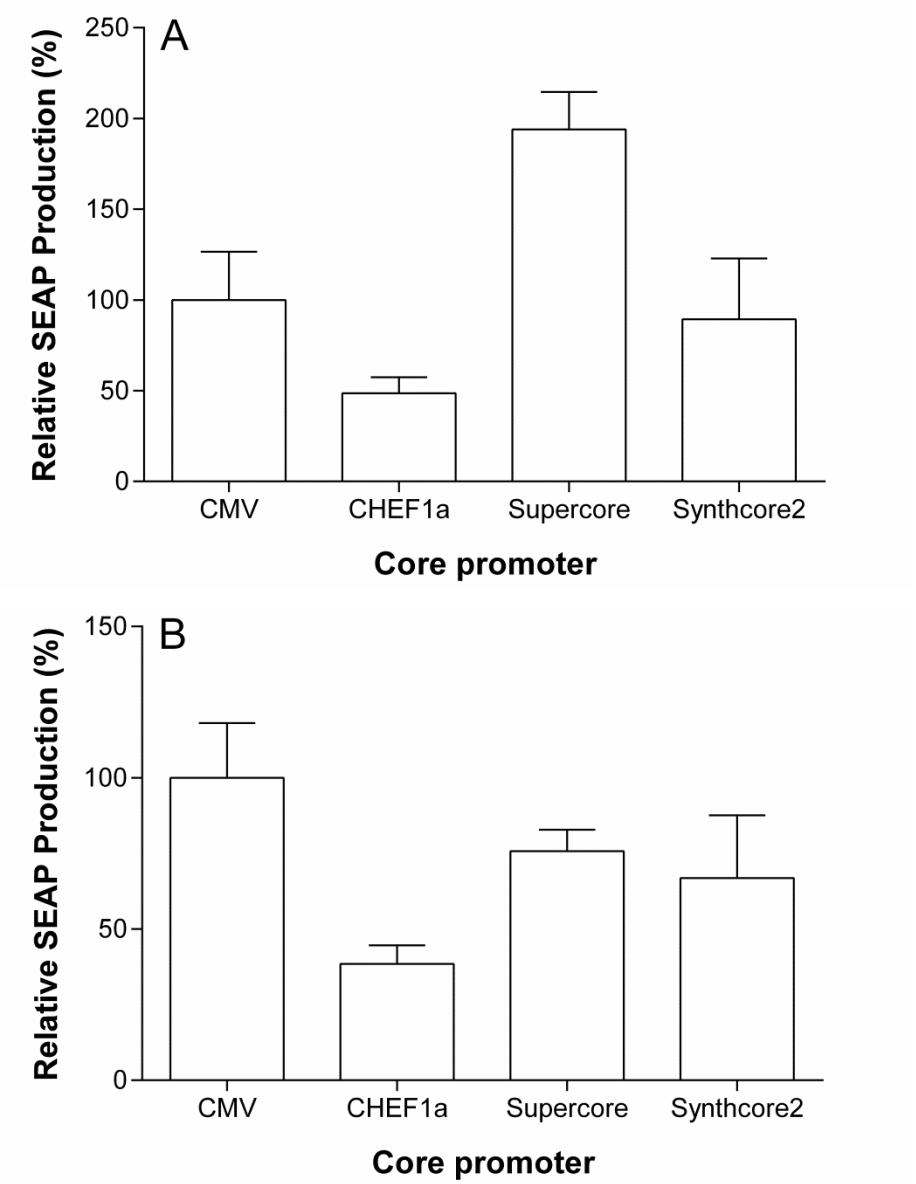


Figure 3.4: Core promoters exhibit variable relative activities in the context of different expression cassettes. Each core promoter construct was cloned downstream of either SV40E (A) or hCMV-IE1 (B) enhancer elements in SEAP reporters. CHO-S cells (2 x 10^5) in 24-well plates were transfected with 1 μg of SEAP reporter-vector and SEAP activity in cell culture supernatant was measured 72 h post-transfection. Data are expressed as a percentage of the production exhibited by CMV core promoter in each expression cassette. Bars represent the mean + SD of three independent experiments each performed in triplicate.

Lastly, the activity of TA box and INR elements in isolation was determined by constructing hCMV-IE1-core-SEAP reporters containing either the wildtype CMV core or TA box/ INR knockout variants (*i.e.* CMV cores with either the TA box or INR element sequence scrambled). Measurement of SEAP reporter production after transient transfection of CHO-S cells with each core-reporter plasmid is shown in Figure 3.5. This analysis showed that i) removal of either CPRE from the CMV core promoter resulted in > 55% decrease in SEAP expression and ii) in contrast to DPE, MTE and DCE, TA boxes and INRs are capable of initiating transcription in isolation in CHO cells.

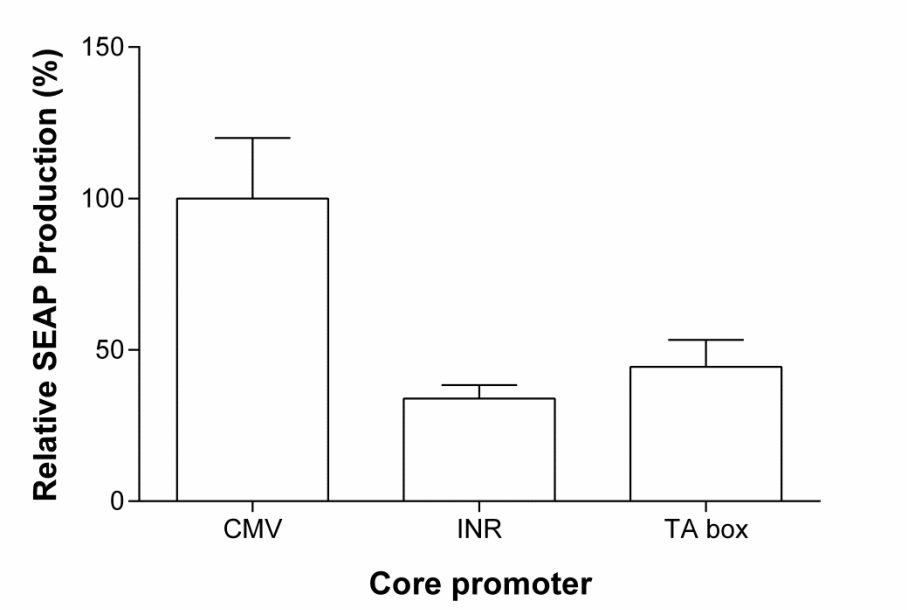


Figure 3.5: TA box and INR elements are capable of initiating transcription in isolation in CHO cells. Core promoter constructs containing only TA box or INR elements were cloned downstream of the hCMV-IE1 enhancer in SEAP reporters. CHO-S cells (2 x 10^5) in 24-well plates were transfected with 1 μg of core reporter-vector and SEAP activity in cell culture supernatant was measured 72 h post-transfection. Data are expressed as a percentage of the production exhibited by the wild-type CMV core promoter (containing both TA box and INR elements) in the same hCMV-IE1 expression cassette. Bars represent the mean + SD of three independent experiments each performed in triplicate.

3.2.1.3 CMV core is the best available part for inclusion in synthetic promoter libraries

Whilst discrete CPRE compositions exhibited variable activity in different expression cassettes, at relatively low levels of expression a rank order of strength was identified; TA box + INR + MTE + DPE > TA box + INR > TA box or INR (Figure 3.6). Addition or subtraction of CPRE blocks therefore represents a simple method for tailoring the transcription rate of lowly expressed genes in CHO. However, at higher levels of expression, supercore, synthcore2 and CMV core drove similar SEAP production levels. This is likely a function of TFRE-TF interactions within enhancers reducing the rate-limitation of core-mediated steps such as PIC formation and transcription initiation (Juven-Gershon and Kadonaga, 2010).

The CMV core has proven functionality in driving high levels of recombinant gene expression in CHO cells (as part of the hCMV-IE1 expression cassette) during production processes. Accordingly, alternative core promoters would need to exhibit significantly improved functionality to replace CMV core in synthetic promoter libraries - particularly the patented supercore which has associated licensing costs. Evaluation of the core promoter engineering design space in CHO-S cells both i) confirmed the CMV core was a robust CHO-active element and ii) revealed that rationally designed synthetic cores did not exhibit enhanced functionality over the CMV core. Therefore the CMV core will be utilised for construction of synthetic promoter libraries.

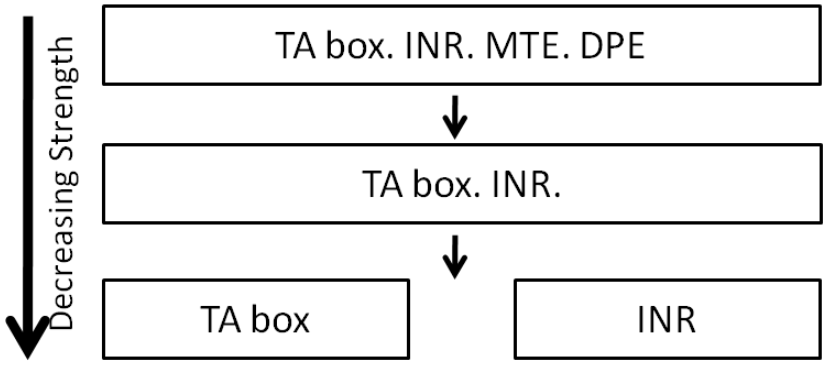


Figure 3.6: Variable CPRE compositions enable three discrete core promoter activities at low levels of expression in CHO cells.

3.2.3. Identification of active transcription factor regulatory elements in CHO-S cells.

Construction of synthetic promoter libraries required a pool of CHO-active TFREs. As it was considered intractable to functionally screen all previously described (> 250) elements (Mathelier *et al*., 2013), a bioinformatics analysis was employed to narrow the design space to a set of TFREs significantly likely to exhibit CHO cell activity. Accordingly, in order to identify discrete TFREs capable of recombinant gene transactivation in CHO-S cells putative TFREs in ten viral promoters (listed in Table 3.1) generally known to be active in CHO cells were surveyed *in silico*. It was rationalised that these promoters must contain one or more constituent TFREs able to bind cognate transcription factors present in CHO cells. Using online search tools that scan DNA sequences for transcription factor (TF) binding sites, specifically Transcription Element Search System (TESS) and Transcription Affinity Prediction tool (TRAP), stringent search parameters (Manke *et al.*, 2008; Schug, 2008) were employed to minimise false positives. Across all viral promoter sequences, 67 discrete TFREs were identified as being present in one or more. To further minimise this pool (design space) TFREs that did not occur in at least two promoters were filtered out. Viral promoter-specific TFRE compositions are listed in Table 3.1. Table 3.2 lists the final set of 28 TFREs incorporated into the functional screen and their associated consensus sequences.

To determine the relative transcriptional activity of each TFRE in CHO-S cells sets of both GFP and SEAP reporter constructs that each contained seven repeat copies of a specific TFRE in series, upstream of the CMV core promoter, were created. Measurement of GFP and SEAP reporter production after transient transfection of CHO-S cells with each TFRE-reporter plasmid is shown in Figure 3.7. This analysis identified seven TFREs that significantly increased expression of both SEAP and GFP over basal expression from the core promoter up to 100-fold (NFκB, E-box, AP1, CRE, GC-box, E4F1, C/EBPα). A further two elements (OCT and RARE) increased SEAP expression but showed no observable increase in GFP, likely due to differences in reporter measurement (e.g. intracellular turnover of GFP, extracellular accumulation of SEAP). No other TFREs mediated a significant increase in reporter expression above core control levels. These data identify a

group of TFREs that can independently mediate activation of recombinant gene transcription in CHO-S cells using available transcription factor activity. The relative level of reporter expression is a function of TF relative abundance, affinity of the TF for its cognate TFRE and the discrete mechanism of transcriptional activation. Lastly, the data implies that many TFREs that exist in common viral promoters used to drive recombinant gene expression in CHO cells may be functionally redundant.

Table 3.1: Transcription factor regulatory elements identified by bioinformatic survey of viral promoters. Ten viral promoters known to exhibit activity in CHO cells were surveyed for the presence of discrete transcription factor regulatory elements (transcription factor binding sites) using Transcription Element Search System (TESS) and Transcription Affinity Prediction (TRAP) algorithms using stringent search parameters to minimise false positives. 28 TFREs that occur in more than one viral promoter are listed.

Promoter	**Transcription Factor Regulatory Elements**
Human Cytomegalovirus immediate early 1 (hCMV-IE1)	AP1, CArG, C/EBPα, CRE, E4F1, EGR1, GC-box, Gfi1, IPF1, NF1, NFκB, RARE, YY1
Mouse Cytomegalovirus immediate early 1 (mCMV-IE1)	AP1, CRE, E-box, E4F1, ERRE, Gfi1, HRE, IPF1, NF1, NFκB, NFAT, NBRE, RARE
Rat Cytomegalovirus immediate early 1 (rCMV-IE1)	AP1, E2F, ERRE, ISRE, NFκB, NFAT, NBRE, RARE
Guinea pig Cytomegalovirus immediate early 1 (gpCMV-IE1)	AP1, GATA, GC-box, GRE, HNF, MSX, NF1, NFκB, OCT, RARE, YY1
Mouse Cytomegalovirus immediate early 2 (mCMV-IE2)	CArG, GC-box, cMyb, E2F, EGR1, GATA, HRE, MSX, RARE
Simian virus 40 early promoter and enhancer (SV40E)	AP1, C/EBPα, cMyb, E-box, GATA, GC-box, MSX, NFκB, OCT
Adenovirus major late promoter (AdMLP)	CDP, E-Box, EGR1, GATA, GC-box, HNF, NF1, YY1
Myeloproliferative sarcoma virus long terminal repeat (LTR) (MPSV LTR)	CDP, cMyb, ERRE, GATA, GC-box, Gfi1, GRE, NF1, RARE, YY1
Rous sarcoma virus LTR (RSV LTR)	CArG, CDP, C/EBPα, ISRE, OCT
Human immunodeficiency virus LTR (HIV LTR)	E-box, GC-box, GATA, HNF, NFκB, NF1

Table 3.2: DNA sequences of transcription factor regulatory elements identified by bioinformatic survey of viral promoters. Ten viral promoters known to exhibit activity in CHO cells were surveyed for the presence of discrete transcription factor regulatory elements (transcription factor binding sites) using Transcription Element Search System (TESS) and Transcription Affinity Prediction (TRAP) algorithms using stringent search parameters to minimise false positives. DNA sequences of single TFREs that occur in more than one viral promoter are listed. Measurement of their relative ability to activate transcription of recombinant reporter genes in CHO-S cells is shown in Figure 3.7.

Transcription Factor Regulatory Element (RE)	**Sequence**
Activator protein 1 (AP1)	TGACTCA
CC(A/T)$_6$GG element (CArG)	CCAAATTTGG
CCAAT displacement protein (CDP)	GGCCAATCT
CCAAT-enhancer binding protein alpha (C/EBPα)	TTGCGCAA
Cellular myeloblastosis (cMyb)	TAACGG
cAMP RE (CRE)	TGACGTCA
Elongation factor 2 (E2F)	TTTCGCGC
E4F1	GTGACGTAAC
Early growth response protein 1 (EGR1)	CGCCCCCGC
Estrogen-related receptor alpha RE (ERRE)	AGGTCATTTTGACCT
Enhancer box (E-box)	CACGTG
GATA-1 (GATA)	AGATAG
GC-box	GGGGCGGGG
Glucocorticoid RE (GRE)	AGAACATTTTGTTCT
Growth factor independence 1 (Gfi1)	AAAATCAAC
Helios RE (HRE)	AATAGGGACTT
Hepatocyte nuclear factor 1 (HNF)	GGGCCAAAGGTCT
Insulin promoter factor 1 (IPF1)	CCCATTAGGGAC
Interferon-stimulated RE (ISRE)	GAAAAGTGAAACC
Myocyte enhancer factor 2 (MEF2)	CTAAAAATAG
Msx homeobox (MSX)	CGGTAAATG
Nerve growth factor-induced gene-B RE (NBRE)	AAAGGTCA
Nuclear factor 1 (NF1)	TTGGCTATATGCCAA
Nuclear factor of activated T cells (NFAT)	AGGAAATC
Nuclear factor kappa B (NFκB)	GGGACTTTCC
Octamer motif (OCT)	ATTAGCAT
Retinoic acid RE (RARE)	AGGTCATCAAGAGGTCA
Yin yang 1 (YY1)	CGCCATTTT
Random 8mer (8mer)	TTTCTTTC

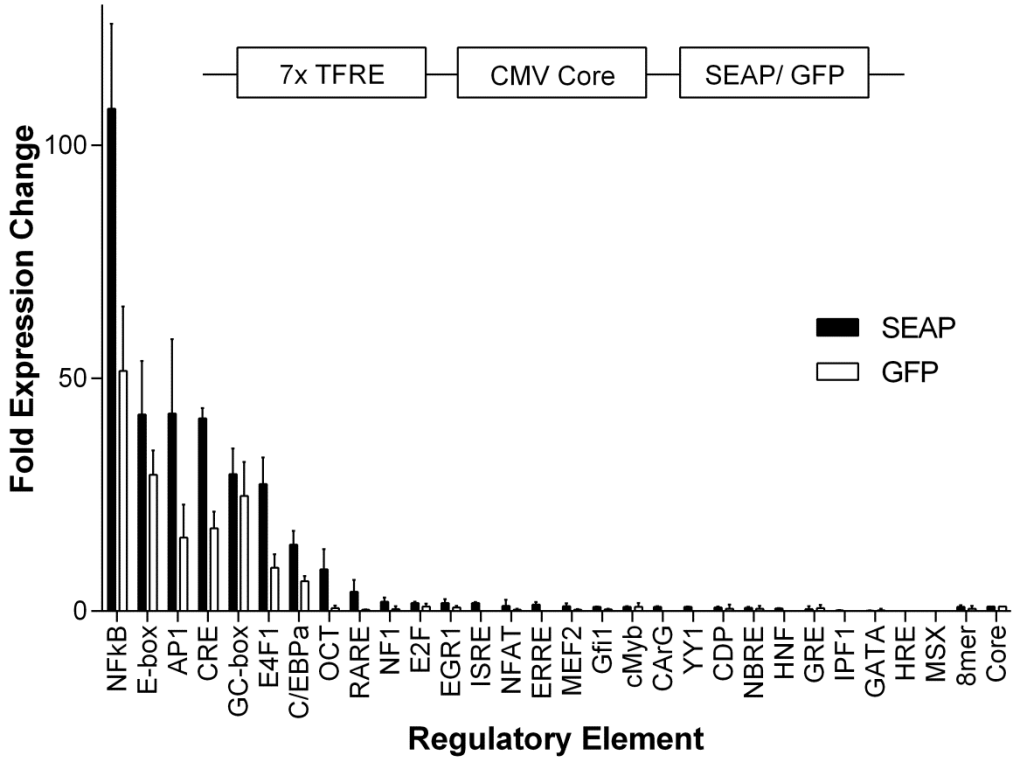


Figure 3.7: Identification of active transcription factor regulatory elements in CHO-S cells. Seven copies of each TFRE (as described in Table 3.2) were cloned in series upstream of a minimal CMV core promoter in reporter vectors encoding either GFP or SEAP reporters. CHO-S cells (2 x 10^5) in 24-well plates were transfected with 1 μg of SEAP (black bars) or GFP (white bars) TFRE reporter-vector. SEAP activity in cell culture supernatant and intracellular GFP were measured 24 h post-transfection. Data are expressed as a fold-change with respect to the activity of a vector containing only a minimal CMV core promoter (Core). A random 8bp sequence with no known homology to TFRE sequences (8mer) was also used as a control. Bars represent the mean + SD of three independent experiments each performed in triplicate, using three clonally derived plasmids for each TFRE-reporter construct.

3.2.4. First generation synthetic promoters exhibit a broad activity range up to that of hCMV-IE1

In order to construct a first generation synthetic promoter library all TFREs identified as transcriptionally active in CHO-S cells were utilised. Oligonucleotide building blocks containing a single copy of each TFRE sequence (TFRE blocks) were chemically synthesised (NFκB, CRE, E-box, GC-box, E4F1, and C/EBPα; AP1

was omitted from the library due to previously observed functional redundancy between CRE and AP1 sites (Hai and Curran, 1991)), and ligated at an equal ratio to assemble random TFRE-combinations which were inserted upstream of the CMV core promoter in SEAP reporter plasmids. A control CMV promoter reporter plasmid was constructed using the hCMV-IE1 promoter (-559 to +48 relative to the TSS, *i.e.* the complete hCMV-IE1 enhancer containing the distal, proximal and core promoter regions, hereafter referred to as CMV) upstream of the SEAP ORF. Purified plasmid DNA from 110 transformed *E.coli* colonies picked at random was utilised for measurement of SEAP reporter production.

Transient production was employed to determine the relative activity of synthetic promoters as it both maximises throughput and provides a direct readout of synthetic promoter transactivation without potential interference from integration-specific effects or silencing. Whilst SEAP production is not a direct measurement of transcriptional activity, previous experiments in this laboratory have confirmed that SEAP activity in cell culture supernatant is linearly correlated with SEAP mRNA levels post-transfection. Moreover assay conditions were optimised such that control CMV-SEAP reporter activity was in the centre of the linear assay range with respect to plasmid copy number (DNA load) and measured SEAP output. SEAP production at 24 h post transfection was measured for each synthetic promoter, and each promoter was sequenced to reveal its TFRE-block composition. A small proportion (14) of reporter plasmids were found to be lacking a promoter insert and these were excluded from further analysis. The relative transcriptional activity of the remaining 96 promoters is shown in Figure 3.8, and their TFRE-block compositions are listed in Supplementary Table 1 in Appendix A. These data show that generation 1 synthetic promoter activities spanned two orders of magnitude, where the most active synthetic promoter exhibited a 1.2-fold increase in SEAP production over that deriving from the CMV control vector.

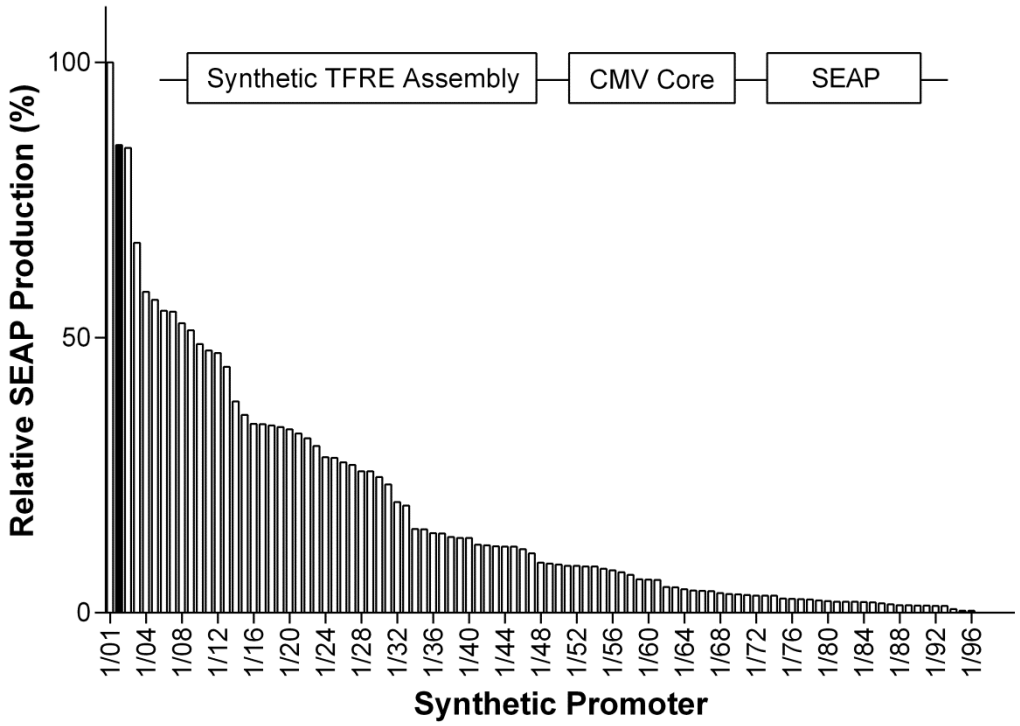


Figure 3.8: First generation synthetic promoters vary in activity up to that of CMV. First generation synthetic promoters were constructed by random ligation of NFκB, CRE, E-box, GC-box, E4F1 and C/EBPα TFREs in equal proportion. Synthetic promoters were inserted upstream of a minimal CMV core promoter in SEAP reporter plasmids and transfected into CHO-S cells. SEAP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by promoter 1/01. SEAP production from the control CMV-SEAP reporter is shown as the black bar. Each bar represents the mean of two transfections, for each promoter less than 10% variation in SEAP production was observed.

3.2.5 Development of a fully automated synthetic promoter library analysis platform to identify optimised design spaces

It was hypothesised that the broad variance in first generation synthetic promoter activities was a function of varying discrete TFRE block compositions (as shown in Appendix A). Determination of the design rules governing promoter activity would accordingly enable rational design of second generation promoters with optimised functionality. It was envisaged that an automated promoter analysis platform facilitating simple selection of optimised design spaces would enable industrial companies to develop desirable synthetic promoters in-house. Statistical analysis software R was utilised to develop the analytics platform as it i) is open-source and

free of charge, ii) is compatible with all operating systems, iii) has a vast array of available statistical analysis tools that are regularly extended by a dedicated community, and v) employs user-friendly computing language (Kreutzer *et al*., 2012; R Development Core Team, 2013). A promoter library analysis R script (synpro.anal.R) was developed that automatically generates a word document analysis report from a CSV file detailing promoter activities and sequence compositions. The data provided within this report enables rapid identification of an optimal design space for second generation promoter library construction. Key components of the four stage analysis will now be outlined using synthetic promoter library 1 data as an example (see Appendix B for the full analysis report). **Note that this section (3.2.5.1 – 3.2.5.5) is intended to detail the generic benefits of the developed analytics platform for rapidly identifying optimal next-generation synthetic promoter library construction strategies. Specific analysis of the design rules that determined first generation synthetic promoter activities is presented in section 3.3.6.**

3.2.5.1. Synthetic promoter library analysis platform: section 1 – key library statistics

The first analysis section provides the data required to evaluate the following assumptions:

1. The relative abundance of TFRE blocks across the promoter library is as designed
2. Relative promoter activity is not simply a function of promoter length
3. Promoter activities cover the desired/ expected expression range.

The analysis first converts the TFRE blocks into letter identifiers, in this example A - F. Data is generated showing i) the distribution of each TFRE block across the promoter library (Table 3.3), ii) the relationship between relative transcriptional activity and promoter length (Figure 3.9), and iii) the range of promoter activities (Figure 3.10). The data in this example showed that i) the relative abundance of each TFRE per promoter was approximately equivalent (as expected given that TFRE-blocks were ligated at an equal ratio), ii) there was no significant correlation between promoter length and activity, and iii) promoter activities spanned over two orders of magnitude. Therefore all three assumptions were confirmed, justifying progression to further analysis.

Table 3.3: Relative abundance at which discrete TFRE blocks (A – F) occurred per promoter in synthetic promoter library 1.

A	B	C
Min. :0.000	Min. :0.000	Min. :0.000
1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000
Median :2.000	Median :2.000	Median :2.000
Mean :1.897	Mean :1.814	Mean :2.237
3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:3.000
Max. :7.000	Max. :7.000	Max. :7.000

D	E	F
Min. :0.000	Min. :0.000	Min. :0.000
1st Qu.:0.000	1st Qu.:1.000	1st Qu.:1.000
Median :1.000	Median :2.000	Median :2.000
Mean :1.629	Mean :1.969	Mean :2.072
3rd Qu.:2.000	3rd Qu.:3.000	3rd Qu.:3.000
Max. :6.000	Max. :7.000	Max. :7.000

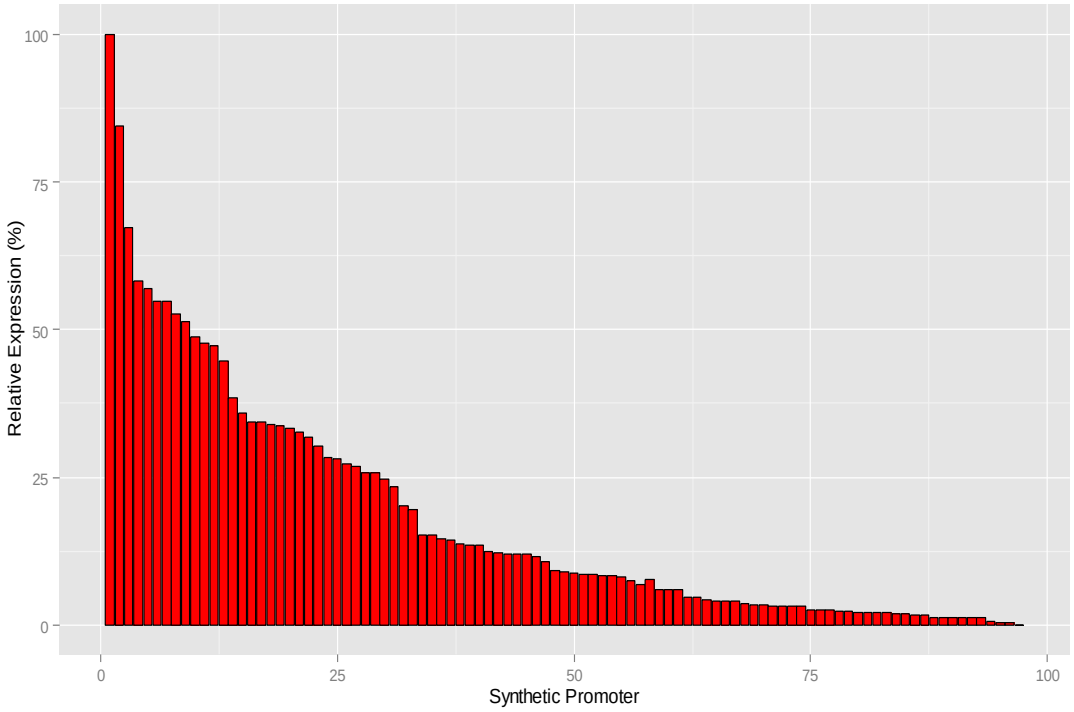


Figure 3.9: Range of promoter activities across the library. Data are expressed as a percentage of the expression exhibited by the most active promoter.

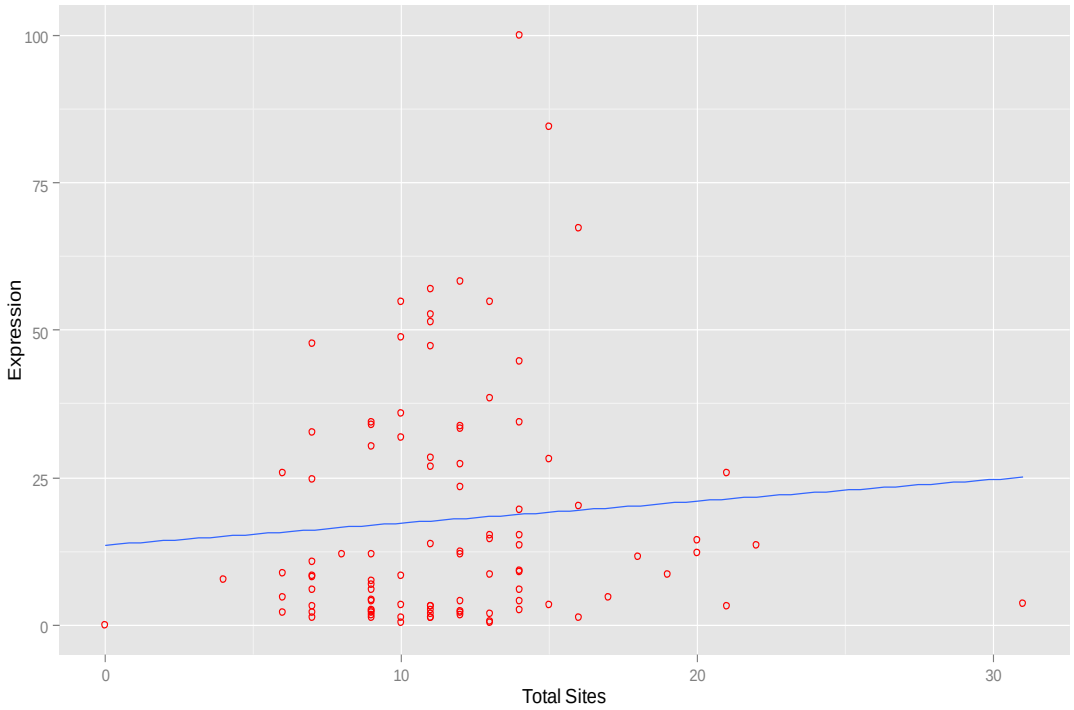


Figure 3.10: Correlation between promoter length and relative transcriptional activity. The slope of the linear regression line (blue) indicates the extent to which total TFRE block number determines relative promoter strengths.

3.2.5.2 Synthetic promoter library analysis platform: section 2 - identification of TFREs that are positive, neutral and negative regulators of promoter activity.

The second analysis section provides the data required to evaluate the contribution discrete TFREs make to promoter activities. Three figures are generated for each TFRE that enables blocks to be designated as positive, negative or neutral effectors of promoter strength. Figure 3.11 shows an example figure from the synthetic promoter library 1 report, generated for block D. These data clearly identify block D as a negative effecter of promoter strength that is abundant in low activity promoters and lacking in high-activity promoters. Determination of each TFRE block's functionality facilitates binary (yes/ no) decisions regarding formation of next-generation design spaces; for example in this case block D would not be included in second generation library construction.

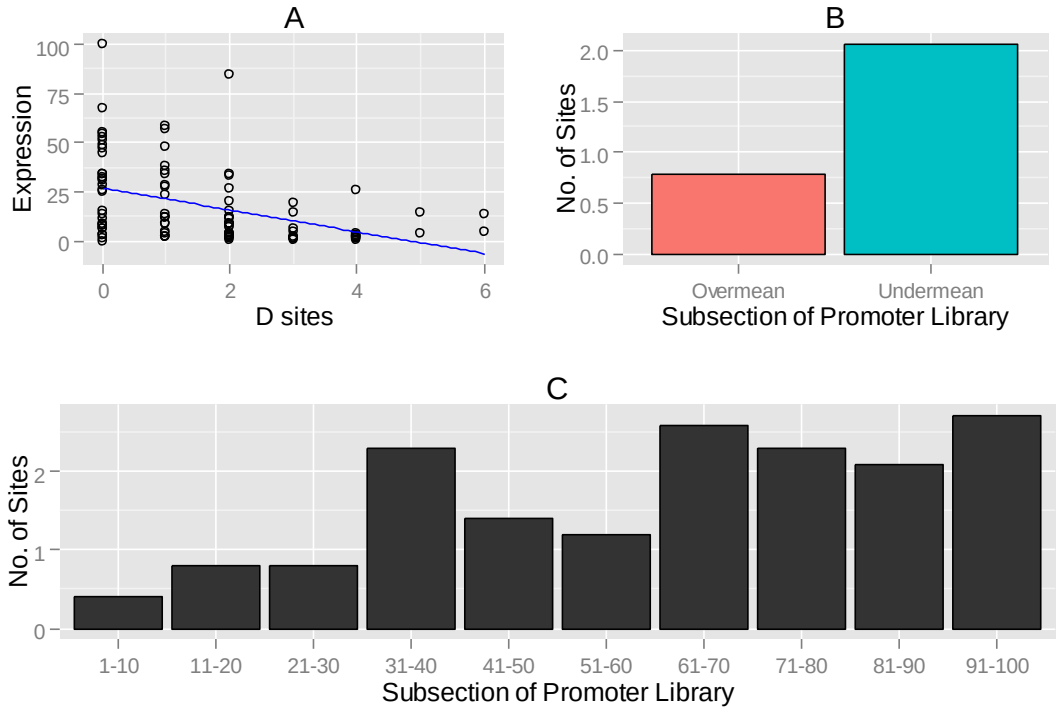


Figure 3.11: Correlation between discrete TFRE block abundances and relative promoter activities. Three figures are generated for each discrete TFRE-block - this example shows the analysis of block D in synthetic promoter library 1. **A)** The number of the TFRE block in each synthetic promoter is plotted against relative activity of that promoter. The linear regression line is shown, where the slope of the line indicates the extent to which the TFRE occurs in promoters of varying activity. **B)** The mean number of the TFRE block in higher or lower activity promoters (over or under mean promoter activity). **C)** The mean number of the TFRE block in promoters within discrete library subsections. Subsection 1-10 contains the top 10% of promoters (ranked by activity).

3.2.5.3. Synthetic promoter analysis platform: section 3 - multiple linear regression analysis

The previous section identified TFREs to be excluded from/ included in second generation library construction. The third analysis section provides the data required to manipulate this design space by identifying optimal TFRE block stoichiometries that can be employed in next-generation library construction. Every possible multiple linear regression model (*i.e.* each combination of TFRE block variables; e.g. A, A + B, A + B + C, *etc.*) explaining promoter activity is ranked (see Appendix B for all models) and the 'best' model (ranked by r^2) for each number of parameters is reported (Table 3.4). The data generated enables identification of parameters

(TFRE blocks) that significantly contribute to explaining promoter strengths. In this example model fit was maximised in a four parameter model, indicating that promoters' relative abundance of blocks A and F had minimal effect on their activities. The most parsimonious model is then reported with associated key statistics, such as the model summary (Table 3.5), analysis of variance table and the contribution of each variable to r^2 (Figure 3.11). The analysis in this example indicated that discrete TFRE blocks were positive (E and C) and negative (B + D) effectors of promoter activity. Moreover, these data identified potential design solutions to specifically tailor next-generation library activities. For example, in this case an optimal stoichiometric ratio of TFRE blocks to increase promoter strengths was predicted to be E 1.58 : C 1. Accordingly, at the conclusion of this analysis stage second generation library construction strategies can be implemented by employing design-led TFRE block ratios in ligation reactions.

Table 3.4: Multiple linear regression models explaining relative promoter activities. The best fitting model for each possible number of parameters (discrete TFRE blocks A- F) is reported. Block inclusion/ exclusion within a model is indicated by a 1 or a 0 respectively. The r^2 statistic is reported for each model.

	TFRE blocks						
Number of parameters	**A**	**B**	**C**	**D**	**E**	**F**	**r^2**
1	0	0	0	0	1	0	0.255
2	0	0	0	1	1	0	0.425
3	0	0	1	1	1	0	0.512
4	0	1	1	1	1	0	0.561
5	1	1	1	1	1	0	0.565
6	1	1	1	1	1	1	0.566

Table 3.5: Key statistics of the 'best' multiple linear regression model explaining synthetic promoter activities.

Residuals:

Min	1Q	Median	3Q	Max
-28.708	-7.958	-1.922	5.880	45.711

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	13.3184	3.5032	3.802	0.000258 ***
B	-3.2190	0.9963	-3.231	0.001713 **
C	3.7386	0.8859	4.220	5.72e-05 ***
D	-5.8172	0.9259	-6.282	1.09e-08 ***
E	5.8472	0.8301	7.044	3.33e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.62 on 92 degrees of freedom

Multiple R-squared: 0.5618, Adjusted R-squared: 0.5427

F-statistic: 29.49 on 4 and 92 DF, p-value: 8.842e-16

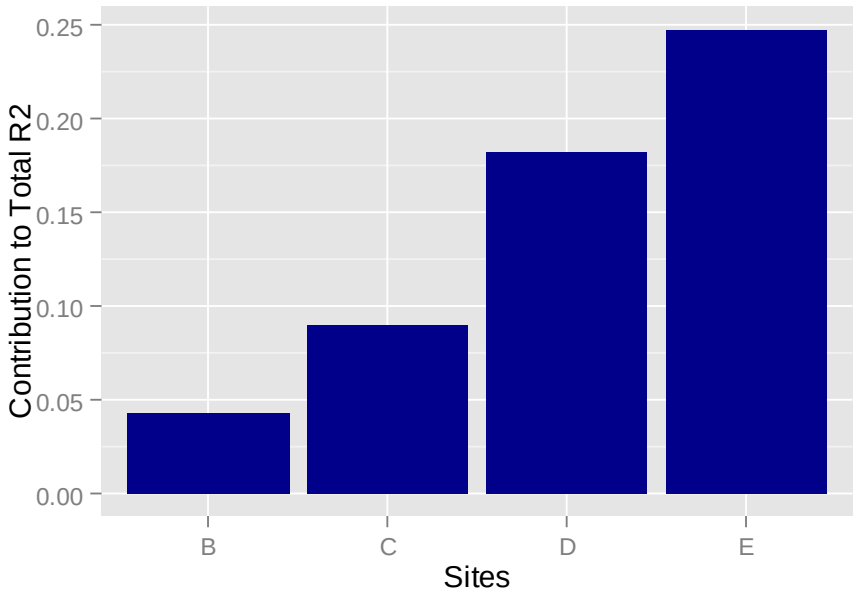


Figure 3.12: Relative contribution of each TFRE-block parameter to the 'best' model's r^2 statistic.

3.2.5.4 Synthetic promoter analysis platform: section 4 – string analysis

The fourth analysis section provides the data required to evaluate the effects of combinatorial interactions between neighbouring blocks within promoters. Design rules governing promoter activity are likely to be more complex than simply being a function of relative TFRE block abundances. Combinatorial interactions between specific combinations of neighbouring TFRE blocks (*i.e.* strings) and spatial effects are likely key determinants of promoter strength. Statistically significant conclusions from string analyses are computationally intractable with the sample size of the first generation promoter library (96). However, the data generated for this analysis section will be detailed in order to provide an example of the output that could be utilised for larger promoter libraries.

The functionality of every two and three block TFRE-string is analysed to both i) determine how neighbouring sites impact each blocks function and ii) allow characterisation of larger parts that could be utilised in future design spaces (*i.e.* blocks containing two or three discrete TFREs could be included in ligation reactions). Strings are evaluated both in the context of each discrete TFRE (Figures 3.12 and 3.13 show the figures generated for block A in synthetic promoter library 1), and as a collection of comparative block parts (Figure 3.14, Table 3.6). The data generated identifies whether discrete TFRE-strings need to be avoided (for example if neutral regulators negatively affect neighbouring positive regulators) or employed in next-generation library construction. Accordingly, second generation library design spaces are finalised at the conclusion of this final analysis stage.

3.2.5.5. The developed analytics platform enables simplified, modular synthetic promoter library construction pipelines.

The R script synpro.anal.R can be utilised to automatically generate the comprehensive data analysis required to optimise next-generation design spaces. This analytics platform can be utilised as a black box technology within biopharmaceutical companies or academic groups and requires minimal computational programming or bioinformatics skills. The platform can be employed to rapidly identify optimal library construction strategies to achieve desired promoter activities – for example to tailor/ optimise the expression levels of a specific

recombinant protein. By facilitating analytics-led design space manipulation the analysis package therefore enables a complete synthetic promoter library engineering pipeline in CHO, which could also be utilised in divergent mammalian cell types.

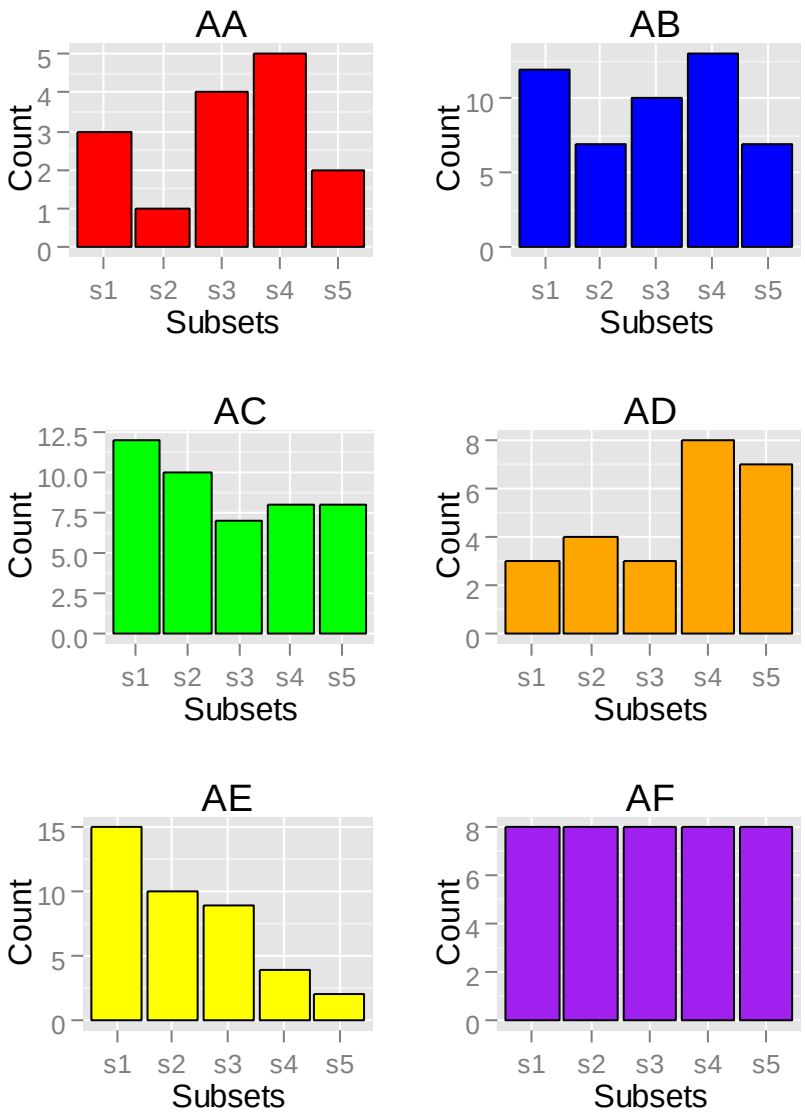


Figure 3.13: Relative abundance of 2-block TFRE strings in synthetic promoters of varying activities. The number of promoters within discrete library subsections (subsection 1 contains the top 20% of promoters, ranked according to strength) that contain each possible 2-block string is reported. Figures are grouped to show the relative differences between every possible 2-block string containing each discrete TFRE (this example shows every 2-block string containing TFRE block A). Counts represent either orientation, e.g. AB could occur as either 'A B' or 'B A'.

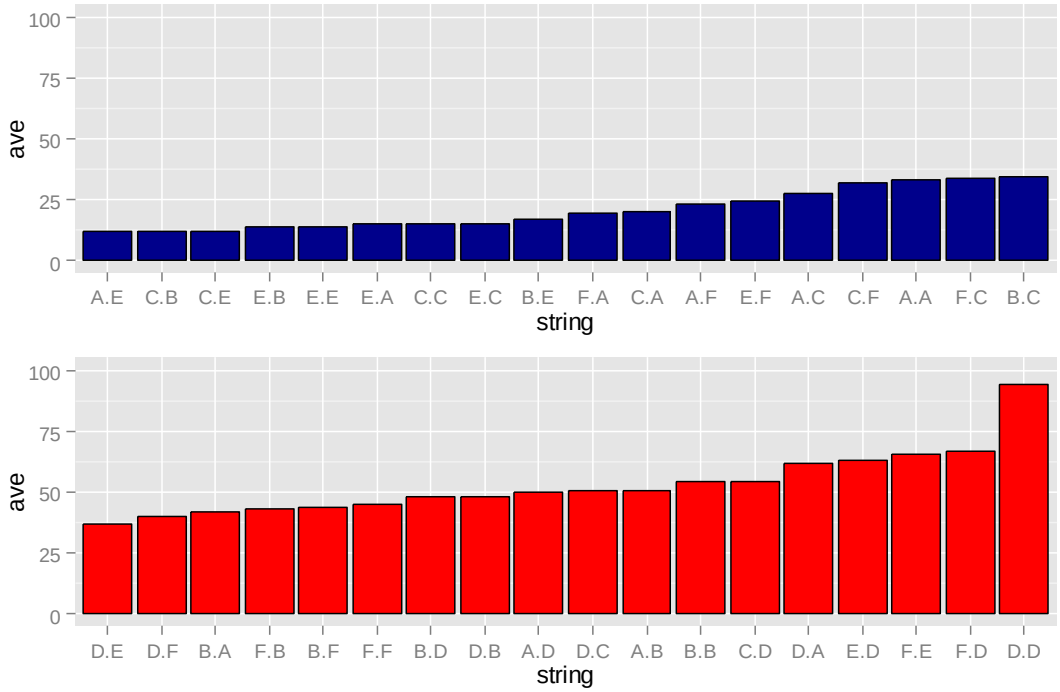


Figure 3.14: Average library position of promoters containing 3-block TFRE strings. Promoters were ranked according to relative activity (1 = most active) and analysed to identify constituent 3-block TFRE strings. For each discrete TFRE a figure is generated detailing the average library position of promoters containing each possible 3-string where that TFRE is the central block (this example shows the output for block E; e.g. A.E and C.B refer to AEE and CEB respectively). Each TFRE figure is split into two sections to highlight higher (blue) and lower (red) performing 3-strings respectively

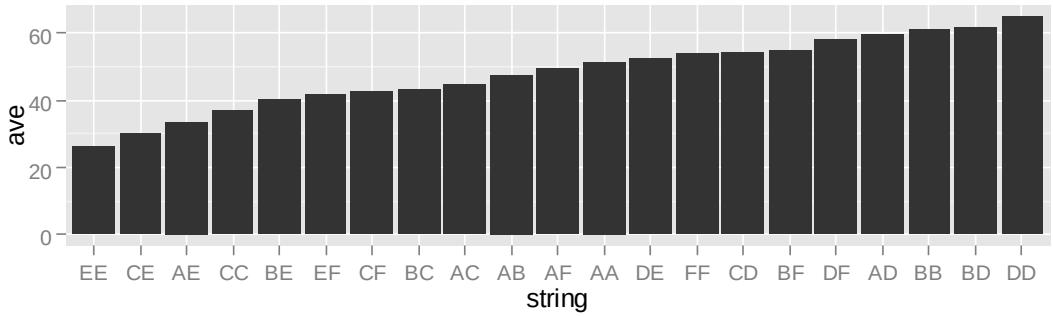


Figure 3.15: Relative activities of each 2-block TFRE string. Promoters were ranked according to relative activity (1 = most active) and analysed to identify constituent 2-block TFRE strings. The average library position of promoters containing each possible 2-string is shown. Strings can occur in either orientation, e.g. CE could occur as either ‘C E’ or ‘E C’

Table 3.6: Relative activities of each 3-block TFRE string. Promoters were ranked according to relative activity (1 = most active) and analysed to identify constituent 3-block TFRE strings. The average library position of promoters containing each possible 3-string is shown.

String	Ave. pos.	String	Ave. pos.	String	Ave. pos.	String	Ave. pos.	String	Ave. pos.	String	Ave. pos.
AAA	73.50	**ABA**	49.50	**ACA**	55.53	**ADA**	NA	**AEA**	33.28	**AFA**	58.80
AAB	56.00	**ABB**	76.50	**ACB**	36.60	**ADB**	63.75	**AEB**	50.66	**AFB**	43.50
AAC	36.50	**ABC**	36.30	**ACC**	54.80	**ADC**	60.25	**AEC**	27.16	**AFC**	50.00
AAD	57.00	**ABD**	67.00	**ACD**	39.80	**ADD**	NA	**AED**	49.66	**AFD**	77.50
AAE	14.00	**ABE**	51.60	**ACE**	20.87	**ADE**	NA	**AEE**	11.83	**AFE**	40.83
AAF	74.75	**ABF**	51.83	**ACF**	37.44	**ADF**	84.20	**AEF**	23.00	**AFF**	35.50
BAA	44.00	**BBA**	74.67	**BCA**	32.60	**BDA**	74.00	**BEA**	41.71	**BFA**	64.66
BAB	64.14	**BBB**	54.50	**BCB**	29.00	**BDB**	55.00	**BEB**	54.00	**BFB**	77.20
BAC	41.60	**BBC**	31.50	**BCC**	43.50	**BDC**	55.25	**BEC**	34.00	**BFC**	34.75
BAD	84.00	**BBD**	NA	**BCD**	66.37	**BDD**	94.00	**BED**	47.75	**BFD**	78.50
BAE	33.12	**BBE**	59.75	**BCE**	7.75	**BDE**	55.67	**BEE**	17.00	**BFE**	50.20
BAF	62.25	**BBF**	59.50	**BCF**	21.00	**BDF**	62.50	**BEF**	43.71	**BFF**	52.71
CAA	64.66	**CBA**	50.33	**CCA**	61.28	**CDA**	49.75	**CEA**	20.25	**CFA**	44.80
CAB	49.88	**CBB**	32.00	**CCB**	33.75	**CDB**	70.33	**CEB**	12.00	**CFB**	36.14
CAC	40.20	**CBC**	34.75	**CCC**	44.50	**CDC**	71.67	**CEC**	14.75	**CFC**	35.66
CAD	77.00	**CBD**	47.00	**CCD**	38.28	**CDD**	51.00	**CED**	54.00	**CFD**	33.33
CAE	18.66	**CBE**	26.83	**CCE**	15.25	**CDE**	42.22	**CEE**	12.00	**CFE**	31.00
CAF	35.60	**CBF**	55.80	**CCF**	25.85	**CDF**	52.75	**CEF**	31.50	**CFF**	52.40
DAA	80.00	**DBA**	77.00	**DCA**	64.40	**DDA**	61.50	**DEA**	61.60	**DFA**	31.50
DAB	77.50	**DBB**	54.75	**DCB**	77.00	**DDB**	52.50	**DEB**	48.00	**DFB**	74.25
DAC	30.66	**DBC**	63.25	**DCC**	50.33	**DDC**	82.00	**DEC**	50.33	**DFC**	50.20
DAD	NA	**DBD**	86.33	**DCD**	75.83	**DDD**	79.25	**DED**	94.00	**DFD**	62.62
DAE	28.66	**DBE**	34.75	**DCE**	44.00	**DDE**	NA	**DEE**	37.00	**DFE**	48.50
DAF	55.66	**DBF**	71.20	**DCF**	65.50	**DDF**	38.00	**DEF**	39.67	**DFF**	73.42
EAA	28.40	**EBA**	31.71	**ECA**	22.00	**EDA**	40.00	**EEA**	14.67	**EFA**	33.00
EAB	23.33	**EBB**	60.00	**ECB**	21.50	**EDB**	59.67	**EEB**	14.00	**EFB**	59.00
EAC	38.90	**EBC**	NA	**ECC**	18.11	**EDC**	55.83	**EEC**	14.75	**EFC**	38.42
EAD	15.00	**EBD**	40.00	**ECD**	28.00	**EDD**	62.33	**EED**	62.67	**EFD**	57.00
EAE	30.85	**EBE**	32.50	**ECE**	40.25	**EDE**	52.50	**EEE**	14.00	**EFE**	15.50
EAF	83.00	**EBF**	52.00	**ECF**	39.20	**EDF**	57.67	**EEF**	24.33	**EFF**	23.66
FAA	72.00	**FBA**	57.25	**FCA**	49.33	**FDA**	40.00	**FEA**	19.60	**FFA**	43.71
FAB	59.67	**FBB**	56.50	**FCB**	56.75	**FDB**	63.00	**FEB**	43.25	**FFB**	53.50
FAC	23.16	**FBC**	33.75	**FCC**	40.67	**FDC**	NA	**FEC**	33.42	**FFC**	76.66
FAD	71.00	**FBD**	79.00	**FCD**	56.00	**FDD**	67.80	**FED**	66.62	**FFD**	61.75
FAE	39.80	**FBE**	45.60	**FCE**	35.00	**FDE**	59.50	**FEE**	65.33	**FFE**	52.80
FAF	28.50	**FBF**	73.00	**FCF**	65.25	**FDF**	62.77	**FEF**	44.67	**FFF**	37.80

3.2.6. Variation in first generation synthetic promoter activity was a consequence of the differing relative abundance of specific TFRE blocks

The developed analytics platform was utilised to identify the design rules that determined first generation synthetic promoter activities. Analysis of synthetic promoter composition revealed that (i) synthetic promoter length varied between 7 and 31 TFRE blocks (mean = 11.9 ± 4.2 blocks; 189 ± 66 bp), although relative transcriptional activity was unrelated to promoter length (ii) across the generation 1 library the relative abundance of the six TFRE building blocks was approximately equivalent and (iii) individual TFRE blocks could occur in either forward or reverse orientation (*i.e.* the consensus TF recognition sequence could occur on either DNA strand) but this was not apparently related to synthetic promoter activity, either with respect to the general frequency of occurrence or with respect to the relative orientation of specific TFRE blocks. Therefore, it was inferred that variation in synthetic promoter activity was a consequence of the differing relative abundance of specific TFRE blocks within promoters and/or positional effects (*i.e.* that specific neighbouring or distal combinations of TFRE blocks may affect promoter strength). Whilst the latter was computationally intractable given the size of the library, the former was addressed by determination of the relative frequency with which individual TFRE blocks occurred within synthetic promoters of varying activity. These data are shown in Figure 3.15. Whilst no single TFRE block exhibited an obviously dominant influence over synthetic promoter strength, individual TFRE blocks were either relatively abundant in active promoters (NFκB, E-box), equally distributed across promoters (C/EBPα, GC-box) or relatively abundant in low activity promoters (E4F1, CRE). This bias was confirmed by multiple linear regression analysis, where either an all factor model (inclusion of all six TFREs, r^2 = 0.57, p = 1.7 x 10^{-14}) or a parsimonious model excluding C/EBPα and GC-box TFREs (as these do not improve model fit; r^2 = 0.56, p = 8.84 x 10^{-16}) predicted the optimal stoichiometry of TFRE blocks to be NFκB 1.58 : E-box 1. The other TFRE blocks were either neutral (C/EBPα, GC-box) or negative effectors (E4F1, CRE). Analysis of specific promoter sequences throughout the library confirmed site designations as positive, neutral or negative. For example, the strongest promoter (1/01) contains the highest ratio of positive (NFκB, E-box) : negative (E4F1, CRE)

sites (9 : 1) in the library. Moreover, the three most active promoters (1/01 – 1/03) are the only promoters in the library containing more than 7 positive sites and fewer than 3 negative sites. There are also multiple examples where high numbers of positive sites are apparently counteracted by high numbers of negative sites to produce relatively weak promoters, for example promoters 1/37, 1/52 and 1/68 which have positive : negative ratios of 8 : 8, 8 : 8, and 9 : 11 respectively.

3.2.7 Second generation synthetic promoters achieve twice the activity of CMV.

In order to further improve synthetic promoter activity a second generation library using random ligation of a mixture of TFRE blocks at an optimal ratio derived from analysis of the composition of first generation promoters was created. Specifically, negative TFREs were omitted (E4F1, CRE, Figure 3.15), positive TFREs were included at the ratio NFκB 5 : E-box 3, and neutral TFREs were included at the ratio C/EBPα 1 : GC-box 1. The latter were included based on the hypothesis that increased complexity could be advantageous. For example, the three most active synthetic promoters in the first generation library all contained at least two copies of both neutral TFREs (see Appendix A) and thus they could contribute to unknown positional effects. It was expected that second generation promoters would contain the same average number of TFRE blocks (12) as first generation promoters.

A second generation library was created as described previously; 50 transformed *E.coli* colonies were picked at random, synthetic promoters in purified plasmid DNA were sequenced and 44 reporter plasmids containing promoter sequences were utilised for measurement of SEAP reporter production. The relative transcriptional activity of second generation promoters is shown in Figure 3.16, and their TFRE-block compositions are listed in Supplementary Table 2 in Appendix A. Second generation promoters exhibited significantly increased activity. The mean expression level (relative to CMV) shifted from 21.2% for first generation promoters to 116% for the second generation library. Twenty five synthetic promoters (57% of the library) achieved a higher SEAP production than the CMV control, with the strongest promoter (2/01) exhibiting a 2.2-fold increase.

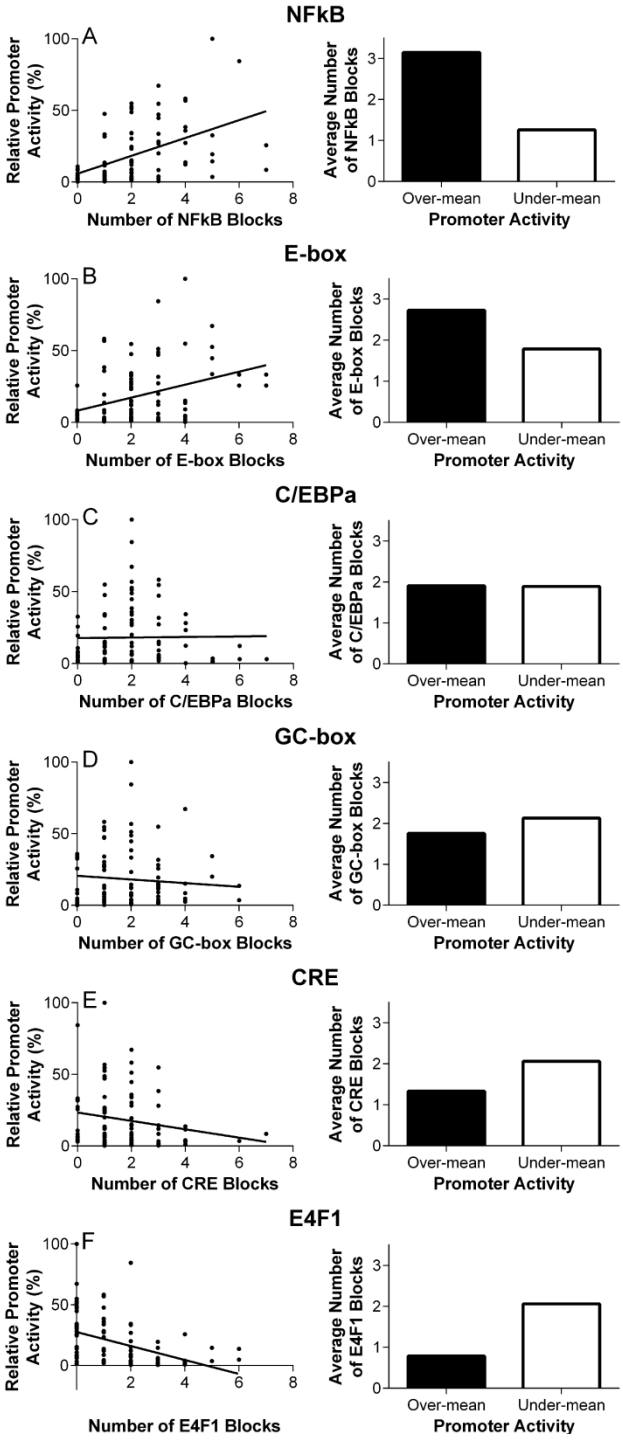


Figure 3.16: Relative abundance of transcription factor regulatory elements in first generation synthetic promoters. First generation synthetic promoters were sequenced to enable assignment of TFRE block composition (listed in Appendix A). The number of each TFRE block in each synthetic promoter is plotted against the relative activity of that promoter (A-F). In each case the linear regression line is shown, where the slope of the line indicates the extent to which each TFRE occurs in promoters of varying activity. The mean number of each TFRE block in higher or lower activity promoters (over or under mean first generation promoter activity) is indicated in each case.

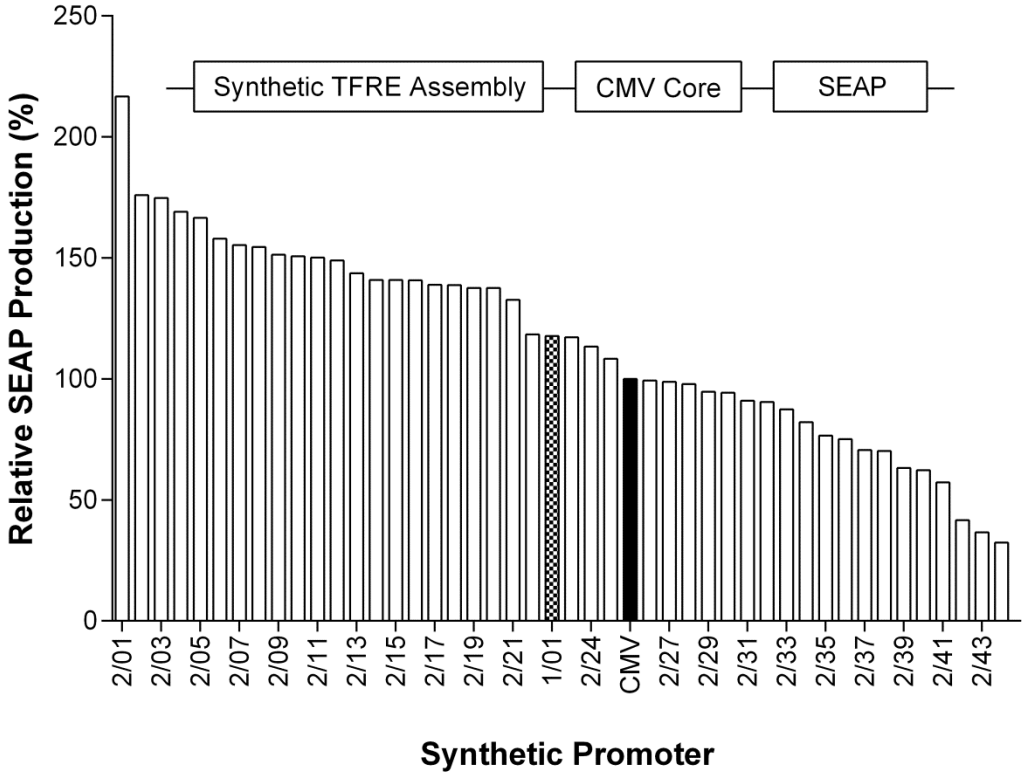


Figure 3.17: Generation two synthetic promoters achieve twice the activity of the CMV promoter. Second generation synthetic promoters were constructed by random ligation of NFκB, E-box, GC-box, C/EBPα TFREs in the ratio 5 : 3 : 1 : 1. Synthetic promoters were inserted upstream of a minimal CMV core promoter in SEAP reporter plasmids and transfected into CHO-S cells. SEAP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by CMV control promoter (black bar). SEAP production from the most active promoter from the first generation library (1/01; Fig. 2) reporter is shown as a checked bar. Otherwise, each bar represents the mean of two transfections, for each promoter less than 10% variation in SEAP production was observed.

3.2.8. Variation in second generation synthetic promoter activity was a consequence of the differing relative abundance of specific TFRE blocks

The analytics platform (synpro.anal.R) was utilised to identify the design rules that determined second generation synthetic promoter activities. Analysis of the TFRE block composition of second generation promoters revealed that the relative stoichiometry of TFRE blocks across the library was approximately as designed

(NFκB 5 : E-box 2.81 : GC-box 1.32 : C/EBPα 1.15). *In silico* analysis utilising TFRE-prediction tools confirmed that unexpected, additional TFRE sites had not formed at TFRE-block junctions during promoter assembly. In contrast to first generation promoters the slope of the fitted linear regression line relating total TFRE block number (synthetic promoter length) and promoter activity was slightly positive ($r^2 = 0.31$, $p = 8.9 \times 10^{-5}$), although this was not regarded as a significant factor. As shown in Figure 3.17, for second generation promoters the influence of GC-box and C/EBPα is generally negative, whereas NFκB and E-box remain positive effectors. However, considering the composition of second generation promoters listed in Appendix A, the data do not support the conclusion that either NFκB or E-box TFRE blocks could support high transcriptional activity alone – clearly a combination of both is necessary. The most powerful promoters (2/01- 2/03) contain relatively high numbers of both TFREs in approximately equal proportion, with a correspondingly low number of negative GC-box and C/EBPα blocks. Some lower activity promoters do contain relatively large numbers of NFκB or E-box blocks (e.g. 2/11, 2/13, 2/17) but (i) contain a sub-optimal ratio of NFκB : E-box (2/11, 2/17) or (ii) also contain relatively large numbers of GC-box and C/EBPα blocks (2/13).

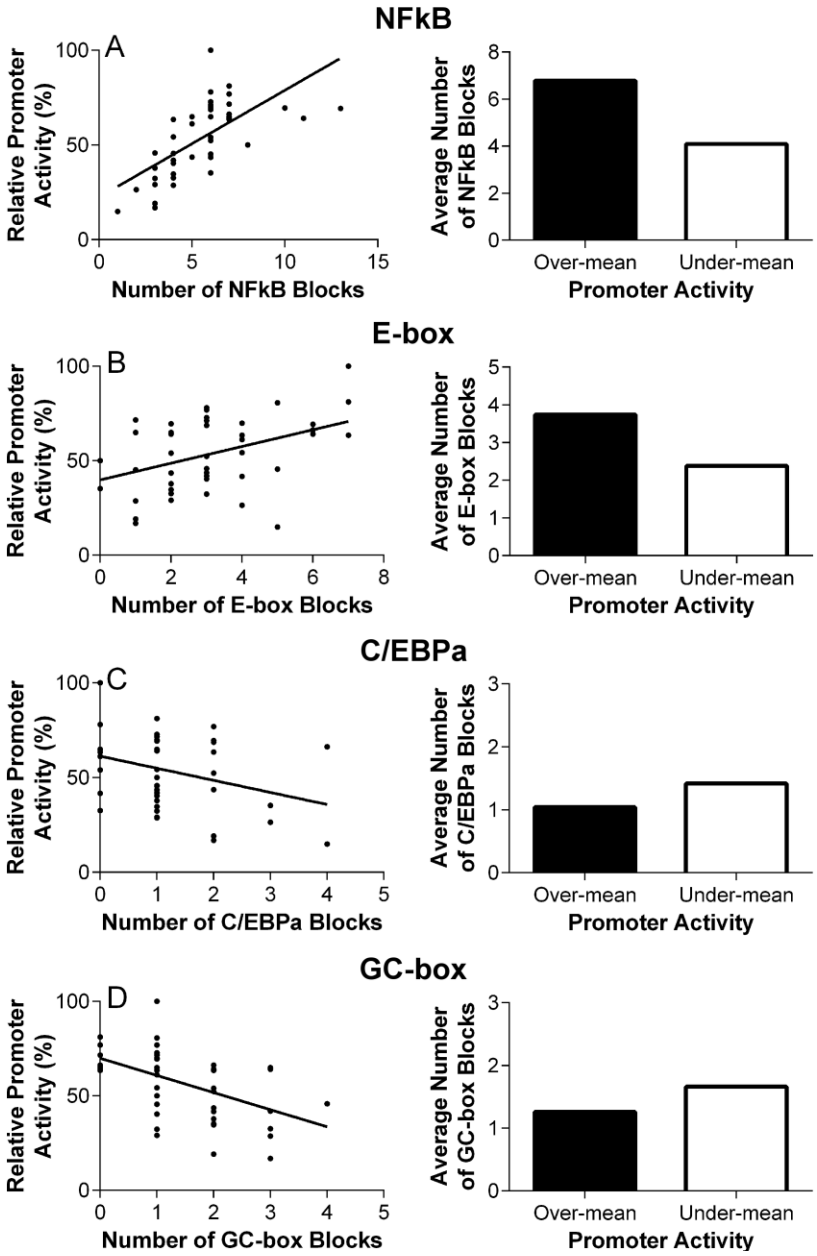


Figure 3.18: Relative abundance of transcription factor regulatory elements in second generation synthetic promoters. Second generation synthetic promoters were sequenced to enable assignment of TFRE block composition (listed in Appendix A). The number of each TFRE block in each synthetic promoter are plotted against the relative activity of that promoter (A-D). In each case the linear regression line is shown, where the slope of the line indicates the extent to which each TFRE occurs in promoters of varying activity. The mean number of each TFRE block in higher or lower activity promoters (over or under mean second generation promoter activity) is indicated in each case.

3.2.9. Synthetic promoters exhibit conserved relative activity in different CHO host cell lines and through a fed-batch transient production process.

In order to determine if synthetic promoters performed robustly and predictably their relative functional capability in different CHO host lines and through a fed-batch transient production process was evaluated. With respect to the former, it was hypothesised that different CHO hosts may contain varying proportions of transcription factors that could markedly influence synthetic promoter function, especially as the complexity of synthetic promoters is significantly reduced compared to CMV. With respect to the latter, it was hypothesised that transition of CHO cells through a production process, with the associated dynamic variation in cell physiology and function (e.g. growth rate, suspension culture) may change the relative proportion of endogenous transcription factors affecting synthetic promoter activity.

A panel of seven promoters from both first and second generation libraries were selected that cover a broad range of promoter activity (1/51 < 1/17 < 1/04 < 1/02 < 2/19 < 2/03 < 2/01). These were compared to the activity of CMV. Figure 3.18 shows transient SEAP production from all promoters in three commonly utilised host lines; CHO-S, CHO-DG44 and CHO-K1. The relative rank order of promoter activity is maintained in all three cell lines, with the exception that 2/03 outperforms 2/01 in CHO-K1. In contrast to the original screen, promoters 2/03 and 2/01 have approximately equivalent expression in CHO-DG44 and CHO-S. In each cell line the top performing synthetic promoter drives significantly higher SEAP production than CMV; 3.1-fold 1.9-fold, and 1.7-fold in CHO-DG44, CHO-S, and CHO-K1 cells respectively. It is noted that in general CHO-DG44 cells exhibited significantly less reporter production than either CHO-S or CHO-K1 cells, presumably due to their reduced "transfectability" by lipofection.

Lastly, it was hypothesised that transfection of cells with synthetic promoters specific for a small number of transcription factors may function as decoys, effectively competing with the host cell genome for transactivation, thus reducing host cell proliferation. Therefore, the same panel of promoters were evaluated in a longer-term fed-batch SEAP production process over 7 days, utilising CHO-S host cells in suspension (Figure 3.19). A transient system was employed (rather than

stable) to ensure production variability was directly linked to differences in promoter activity rather than cell line specific, site-specific integration or promoter silencing (e.g. methylation, deletion) artifacts. It was observed that the differences in synthetic promoter activity observed in static microplates were maintained in the fed-batch transient system. The highest SEAP titer, driven by promoter 2/03, was over 1.65-fold that obtained by CMV-mediated expression. It was observed that no synthetic promoter had a significant effect on the integral of viable cell density at the end of the transient production process, disproving the hypothesis.

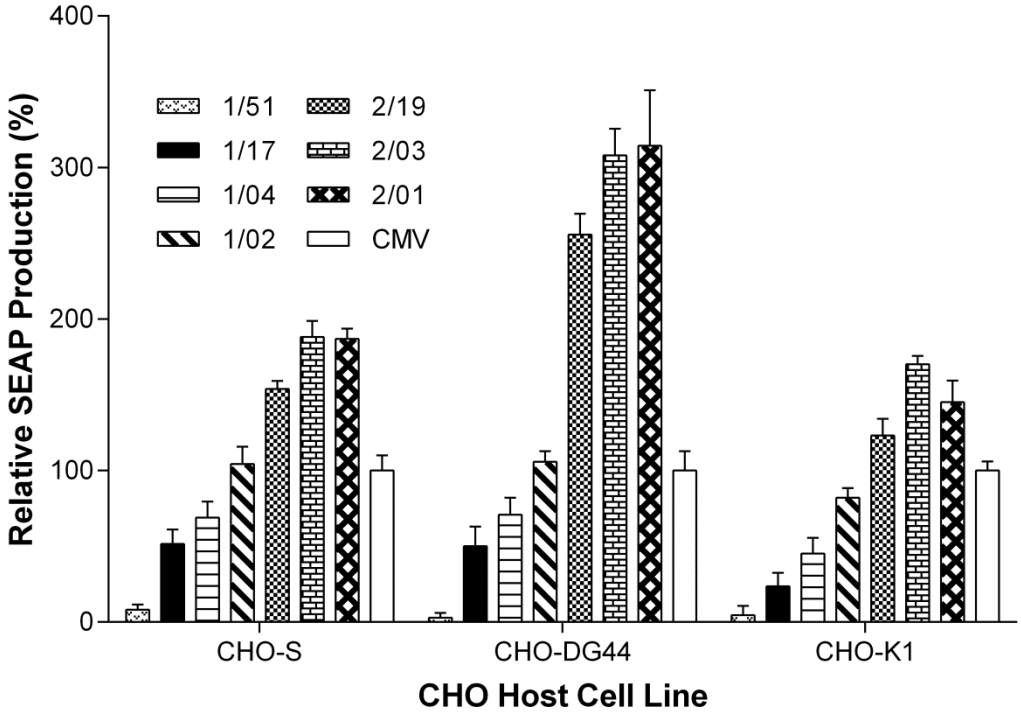


Figure 3.19: Synthetic promoters exhibit conserved relative activity in different CHO host cell lines. The relative activity of seven synthetic promoters with differing relative activity was determined in CHO-S cells, CHO-K1 cells and CHO-DG44 cells. Cells (2 x 10^5) were transfected with 250 ng SEAP-reporter vector, and SEAP production was quantified 24 h post-transfection. Data are expressed as a percentage of the activity of the CMV promoter in each cell line. Note that in general CHO-DG44 cells exhibited significantly less reporter production than either CHO-S or CHO-K1 cells (SEAP production from the CMV promoter in each cell line (100%) occurred at the ratio CHO-S 1: CHO-K1 0.81: CHO-DG44 0.32). Values represent the mean + S.D of three independent experiments performed in triplicate.

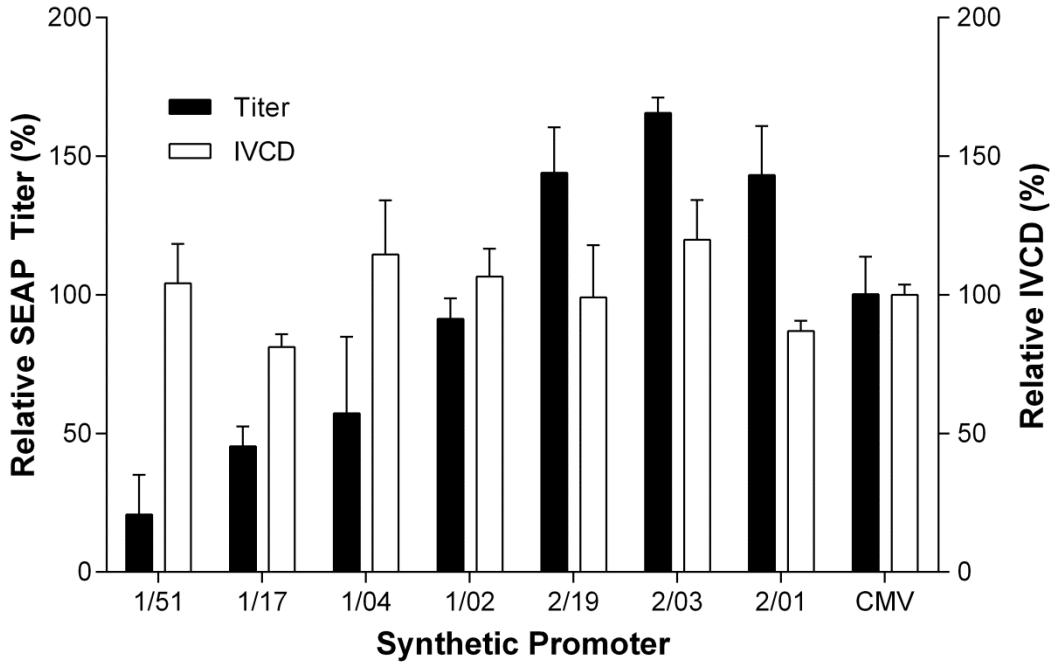


Figure 3.20: Synthetic promoters exhibit conserved relative activity in a fed-batch transient production process. CHO-S cells (6 x 10^6) were transfected with 7.5 μg of SEAP reporter-vectors, where SEAP expression was under the control of synthetic promoters with varying activity or the control CMV promoter. SEAP production and viable cell concentration were measured over the course of a 7-day fed-batch process in tube-spin bioreactors. The mean IVCD at Day 7 (white bars) and SEAP titer (black bars) are shown. SEAP data are expressed as a percentage of the control CMV promoter activity. Two independent transfections were performed in duplicate.

3.3. Discussion

The work in this chapter describes construction of the first bespoke synthetic promoters designed specifically to function with the transcriptional machinery of CHO cell factories. The characterised transcriptional 'parts' (CPREs, TFREs, core promoters and proximal/ distal promoters) and suite of synthetic promoter construction tools (TFRE-reporter plasmids, analytics platform, modular construction pipeline methodology) can be utilised to replace functionally ill-defined and uncontrollable genetic elements in expression vectors with sophisticated, bespoke controllers that can engineer host cell function predictably.

Synthetic promoter activity was primarily a function of a promoter's relative composition of NFκB and E-box TFREs. TFs that bind these sites are amongst the

most frequently occurring oncogenes, and their impaired regulation is a principal mechanism by which cancerous cells achieve uncontrolled proliferation and suppressed apoptosis (Dolcet *et al*., 2005; Lin *et al*., 2012). It is therefore likely that these synthetic promoters will be functional in a range of transformed mammalian cell types and may have use in cancer-targeted gene therapeutic applications. It is anticipated that there are more complex design rules governing NFκB and E-box function (spatial effects, combinatorial interactions) and that many additional combinations of TFREs could be utilised to construct CHO cell synthetic promoters. Determination of the former would require construction of larger, specifically designed synthetic promoter libraries with the associated computational modelling, whilst the latter would require a more comprehensive screen of CHO cell TFRE activities. Moreover, it is envisaged that emerging transcriptomics data will enable identification of endogenous TFREs that are active in different CHO cell hosts, or under specific bioprocess conditions (Datta, 2013).

TFREs that do not exhibit activity in CHO-S cells were indirectly identified. Whilst inactivity may be explained by assay-design features (*i.e.* suboptimal TFRE sequences, TFs unable to drive transactivation independently) it could be evidence that their cognate TFs are not expressed in CHO cells. It would not be surprising if CHO cells expressed a particularly narrow range of transactivators given that CHO cells have lost specific functionalities over the course of their long-term (> 50 years) culture *in vitro* (Lewis *et al*., 2013; Xu *et al*., 2011). Unrecognised TFREs could be utilised with their cognate TFs as inducible, controllable elements to construct regulated gene expression systems (switches, oscillators, synthetic regulatory cascades) for synthetic biology applications in CHO cells (Tigges and Fussenegger 2009; Weber and Fussenegger 2010). Many of these inactive TFREs are present in the CMV promoter, suggesting that a large proportion of the CMV sequence (and that of other viral promoters) may be functionally redundant in CHO cells. The CMV promoter has evolved to access a broad range of TFs to enable the virus to infect a wide range of host cells with varying physiology and TF activity (Stinski and Isomura, 2008). Specifically designed, rather than evolved, the synthetic promoters have far more efficient transcriptional activity per unit DNA sequence (synthetic promoter 2/01 exhibited a 2.2-fold increase in activity over CMV but is less than half the size).

Synthetic promoters that achieved higher titers than CMV in fed-batch transient SEAP production could be utilised to optimise the transient production of early stage products (*i.e.* development material for toxicology and clinical trials testing (Daramola *et al*., 2013)). Although the promoters have not been tested in a stable expression format, and therefore it cannot be definitively claimed that they will maintain either their relative activities or expression stability in stable transfectants, they may offer additional advantages beyond high level transcriptional activity. For example, the CMV promoter can contribute to production instability via promoter methylation and gene deletion (the latter likely via homologous recombination, *i.e.* 'looping-out' caused by two identical CMV sequences surrounding a light chain gene copy (Kim *et al*., 2011)). This homologous recombination also predisposes synthetic genetic circuits to failure (Sleight *et al*., 2010). The most potent synthetic promoters do not contain CpG islands and two different synthetic promoter sequences could be utilised within constructs to minimise gene copy loss. Further, as CpG islands in CMV have been shown to interfere with the functionality of matrix attachment regions (Girod *et al*., 2005), the synthetic promoters could be more compatible with existing transcription enhancing technologies such as ubiquitously-acting chromatin opening elements, bacterial artificial chromosomes and site-specific integration systems (Zhou *et al*., 2010; Mader *et al*., 2012). Moreover, they could be utilised with other expression control technologies in research applications requiring highly precise regulation. For example, they could be employed in conjunction with recently described synthetic elements that control translation initiation rates (Ferreira *et al*., 2013). Lastly, it is suspected that by utilising specific TFRE combinations the synthetic promoters could be refined to exhibit desirable bioproduction functionalities such as increased activity at sub-physiological temperatures (Al-Fageeh *et al*., 2006) or during stationary phase cell growth (Prentice *et al*., 2007).

The synthetic promoters presented in this chapter enable precise control of recombinant gene expression in CHO host cells over a broad dynamic range. For easy to express proteins, where transcription rates have been shown to exert a high level of control over production, (O'Callaghan *et al*., 2010; McLeod *et al*., 2011), they could be utilised to maximise recombinant gene transcription levels (for example, by using synthetic promoter 2/01). For difficult to express proteins (e.g. bispecific antibodies, fusion proteins), where maximising transcription is unlikely to

be beneficial, they could be utilised to provide optimised protein-specific transcription activity kinetically coordinated with polypeptide-specific folding and assembly rates. An obvious potential application is in monoclonal antibody (mAb) expression where synthetic promoters of varying activity could be used to achieve mAb-specific light chain : heavy chain (LC : HC) expression ratios to optimise mAb production (Ho *et al*., 2013; Pybus *et al.,* 2013). For example, the recent study by Pybus *et al*., demonstrated that three different difficult to express mAbs had three discrete optimal LC : HC expression ratios of 9 : 1, 4 : 1 and 2.3 : 1. By utilising synthetic promoter 2/01 (to maximise LC gene expression) in conjunction with synthetic promoters 2/25, 1/03 or 1/24, each of these ratios could be functionally achieved. Lastly, the provision of 140 discrete promoter activities, covering over three orders of magnitude, will enable CHO cell engineers to precisely engineer the cell factory, where systems level control of cell function may require the constitutive expression of several genes to be stoichiometrically balanced. By enabling sophisticated multigene engineering systems these synthetic promoters will therefore facilitate exploitation of the current revolution in CHOmics (Datta *et al*., 2013).

Chapter 4: Transcription Factor Block-Decoys

This chapter presents a novel form of TF decoys that are specifically designed for use in CHO cells. A method was developed whereby blocks containing discrete transcription factor binding sites (TFRE-blocks) are combined into circular molecules, enabling rapid construction of chimeric decoys containing stoichiometrically optimised ratios of input TFRE-blocks. It was demonstrated that block-decoys were able to inhibit expression from multiple target elements simultaneously in CHO cells using a bespoke chimeric decoy. Enabling investigation of any multi-TF mediated cell function or phenotype block-decoys are a valuable new tool for characterising and controlling CHO cell transcription.

Based on the work within this chapter the following article has been published (shown in Appendix D):

Brown AJ, Mainwaring DO, Sweeney B, James DC. 2013. Block decoys: transcription factor decoys designed for in vitro gene regulation studies. Anal. Biochem. 443(2): 205-210

Based on the work within this chapter the following patent application has been filed:

Brown AJ, James DC.UK patent application number GB1310853.5 (18 June 2013) Transcription factor block-decoys

4.1. Introduction

Transcriptional output of a given gene at a specific time point is determined by the composition of transcription factor regulatory elements (TFREs) within its promoter and the availability of cognate transcription factors (TFs) within the cell (Coulon *et al*., 2013). Cellular transcriptomes are therefore a functional consequence of multiple TF-TFRE interactions occurring at thousands of discrete genomic loci. A mechanistic understanding of the TF-TFRE interactions regulating individual

promoters' transcription would enable strategies to predictably control, manipulate and improve their activities. A mechanistic understanding of the TF-TFRE interactions regulating multiple promoters' activities within discrete pathways would enable strategies to engineer entire cellular processes. Characterisation of TF-TFRE interaction functionalities within the CHO cell factory could accordingly enable i) optimisation of product gene transcription rates throughout biomanufacturing processes and ii) cell line engineering strategies to achieve desirable bioproduction phenotypes, such as resistance to apoptosis and increased proliferation.

Physical disruption of TF binding to target sites is the most effective and well-established method of investigating TF-TFRE interactions. An effective method to achieve this is the use of transcription factor decoys (Tomita *et al*., 1999; Bezzerri *et al*., 2011; Renard *et al*., 2012); short synthetic oligodeoxynucleotides (ODN) that contain a specific TFRE motif. When introduced into a cell the decoys compete for available TFs, preventing their association at target promoters (Bielinska *et al*., 1990). This site-specific sequestration of TFs makes decoys an attractive method to determine the functional contribution of individual TFREs to a promoter's activity.
The key determinants of decoy effectiveness are stability, specificity, and uptake (Osako *et al*., 2012). Multiple methods of decoy formation have been developed to improve these factors, primarily focusing on their stability against intracellular nucleases. These include chemical modifications such as the use of phosphorothioate groups (Bielinska *et al*., 1990), and circular dumbbell ODNs that have enzymatically ligated termini (Osako *et al*., 2007), conferring resistance to exonucleases (the primary cause of intracellular degradation (Gamper *et al*., 1993)). Although such advancements have greatly improved decoy functionality, particularly in potential therapeutic applications (Gambari *et al*., 2011), currently available methods are not ideally suited to *in vitro* gene regulation studies.

As most promoters contain binding sites for multiple TFs, gene regulation studies utilising decoys are likely to require multiple decoys, targeting varying combinations of different TFREs. Ideally, where multiple TFREs are targeted at once they would be included on a single decoy molecule to avoid the uneven distribution of different decoys across the transfected cell population. Phosphorothioate and dumbbell decoys targeting two (Miyake *et al*., 2006; Lee *et al*., 2012) and three (Gao, 2006) TFREs have been described (and shown to be far superior to using individual decoys against each site) but these formation methods do

not allow for the rapid creation of bespoke chimeric decoys. Further, they do not provide the capability to fine-tune the molar ratio of different sites within one molecule. Currently available tools are therefore poorly suited for *in vitro* investigations into multi-transcription factor mediated processes that may require multiple regulatory elements to be inhibited in varying combinations. Determination of the TF-TFRE interactions regulating promoters/ cellular pathways in CHO cells is therefore intractable with current decoy methods.

This chapter describes a novel method for decoy formation, specifically designed to enable characterisation of TF-TFRE interaction functionalities in CHO cells. This method utilises blocks containing discrete TFREs (TFRE-blocks) to construct circular, exonuclease-resistant molecules (block-decoys). Unlike currently available methods block-decoys allow rapid construction of chimeric decoys targeting multiple regulatory elements. Further, they enable fine tuning of binding site copy ratios within chimeras, allowing sophisticated control of the CHO cellular transcriptional landscape. This novel method offers significant advantages for multi-target decoy studies investigating multi-TF mediated phenotypes in CHO, and is particularly suited to gene regulation studies. Block-decoys therefore represent a valuable new tool for investigating, characterising and controlling CHO cell transcription.

4.2 Results

4.2.1 'Plasmid-decoys' exhibit undesirable decoy function

In order to construct decoys optimised for use in CHO cells, plasmid vectors were evaluated as a TFRE delivery vehicle. It was hypothesised that plasmids would exhibit desirable decoy functionality as they i) are exonuclease resistant, ii) have well-established transfection protocols and iii) can be produced via low-cost, high-yield systems. A minimal TFRE-acceptor vector (< 2500bp) containing only the sequences required for propagation in bacteria (pUC Ori, AmpR) and an extended multiple cloning site (MCS) was constructed in order to maximise TFRE copies per unit DNA sequence (*i.e.* maximise the number of TFRE copies that can be transfected per cell) (Figure 4.1). TFRE-specific decoys were constructed by

inserting appropriate synthetic ODNs (TFRE-ODNs) containing multiple TFRE copies into the MCS. As low-cost DNA synthesis is restricted to ODN sizes of < 120 nucleotides, approximately 7x TFRE copies can be inserted per TFRE-ODN. It was predicted that bespoke chimeric decoys could be constructed by inserting different TFRE-ODNs into discrete MCS acceptor regions (nine TFRE-ODNs can be inserted per decoy-plasmid).

The use of decoy-plasmids to inhibit the activity of specific TFREs was evaluated using a GFP reporter plasmid containing 7 copies of the NFkB-RE motif upstream of a core CMV promoter (*i.e.* the NFkB-reporter plasmid described and utilised in section 3.3.3). An NFkB-RE decoy was constructed by separately inserting two NFKB-RE-ODNs (*i.e.* 14 NFkB-RE copies) into the decoy plasmid MCS. Measurement of GFP production after transient co-transfection of CHO-S cells with reporter vector and varying concentrations of decoy plasmid is shown in Figure 4.1. This analysis identified that NFkB-RE-reporter expression was not inhibited by NFkB-RE decoy plasmids. It was hypothesised that decoy-plasmids did not enable delivery of sufficient TFRE copies per cell to sequester significant quantities of cognate TFs. Therefore, it was concluded that the requirement for 'accessory' DNA sequences (*i.e.* pUC Ori, AmpR) was critically limiting to decoy-plasmid functionality, rendering it an ineffective method for decoy formation.

4.2.2 Block-decoy formation and stability

In order to construct circular decoy molecules with significantly increased TFRE copies per unit DNA sequence an improved formation method was developed. This method utilises blocks containing discrete TFREs (TFRE-blocks) to construct circular molecules (block-decoys). The method of block-decoy construction is shown schematically in Figure 4.2. TFRE-blocks containing a single transcription factor binding site and a 4 bp TCGA single stranded overhang at each 5' terminus are created by annealing two complementary, single stranded 5' phosphorylated DNA ODNs. The cohesive ends enable RE-blocks to be ligated together into extending concatamers. At sizes greater than 100 bp DNA molecules are likely to bend (Ulanovsky *et al*., 1986; Zhang *et al*., 2003), allowing ligation of cohesive termini (Shore *et al*., 1991; Révet *et al*., 1998). Therefore, ligation of input TFRE blocks theoretically results in covalently closed circular block-decoys containing multiple

copies of the target binding sites. It was predicted that block-decoys would exhibit desirable functionalities of decoy-plasmids (easy-to-transfect, exonuclease-resistant, low cost production) whilst increasing TFRE copies per unit DNA sequence by > 100 fold. Moreover, it was hypothesised that bespoke chimeras could be rapidly constructed by adjusting the stoichiometric ratio of different TFRE blocks in ligation reactions.

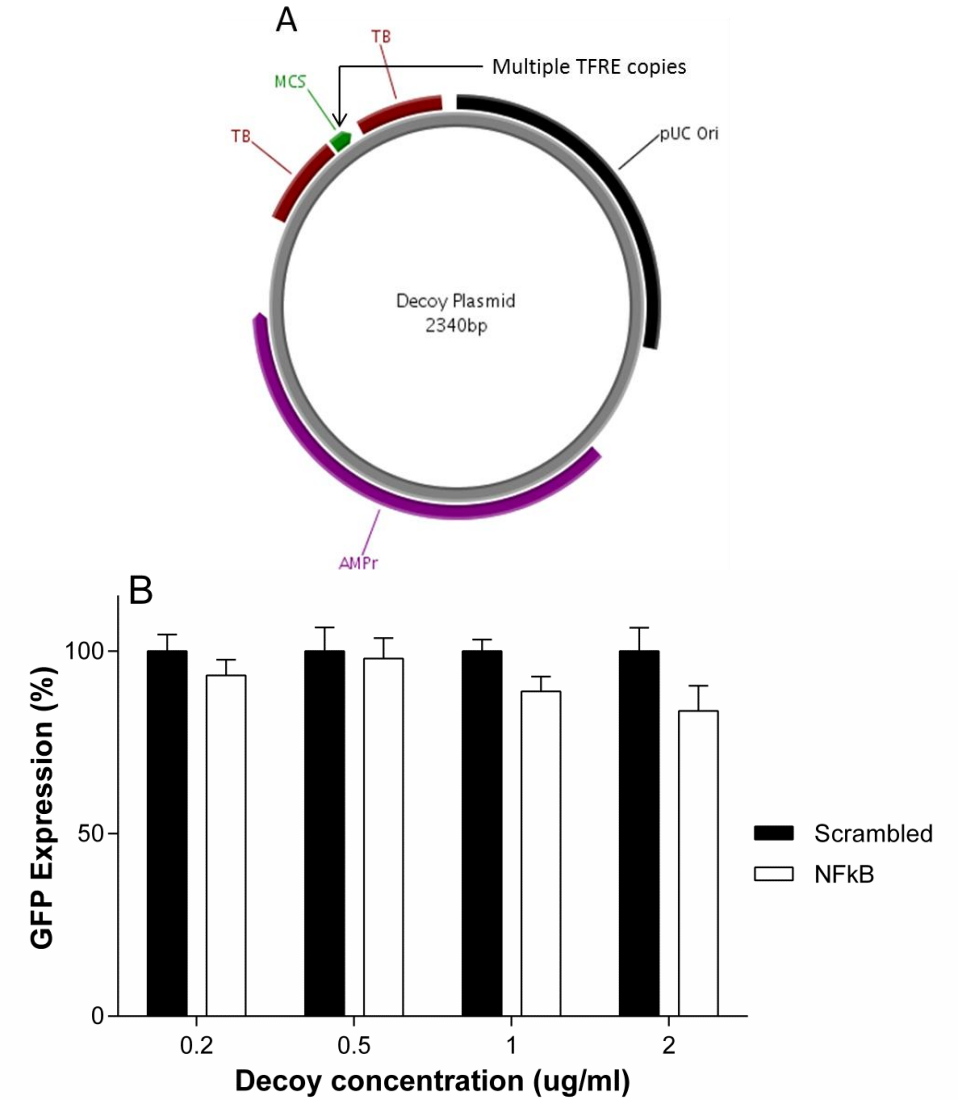


Figure 4.1: Decoy-plasmids do not inhibit expression mediated by discrete regulatory elements. A) Decoy-plasmid map. TFRE specific decoy-plasmids are constructed by cloning synthetic ODNs containing multiple TFRE copies into the MCS. **B)** CHO-S cells (2 x 10^5) were co-transfected with an NFkB-RE-dependent GFP reporter plasmid with either an NFkB-RE decoy-plasmid (white bars) or a scrambled NFkB-RE decoy-plasmid (black bars) at concentration of 0.2 – 2 µg/ml DNA per transfection. Data were normalised with respect to GFP expression in the presence of the scrambled decoy in each case. Bars represent the mean + SEM of three independent experiments each performed in triplicate.

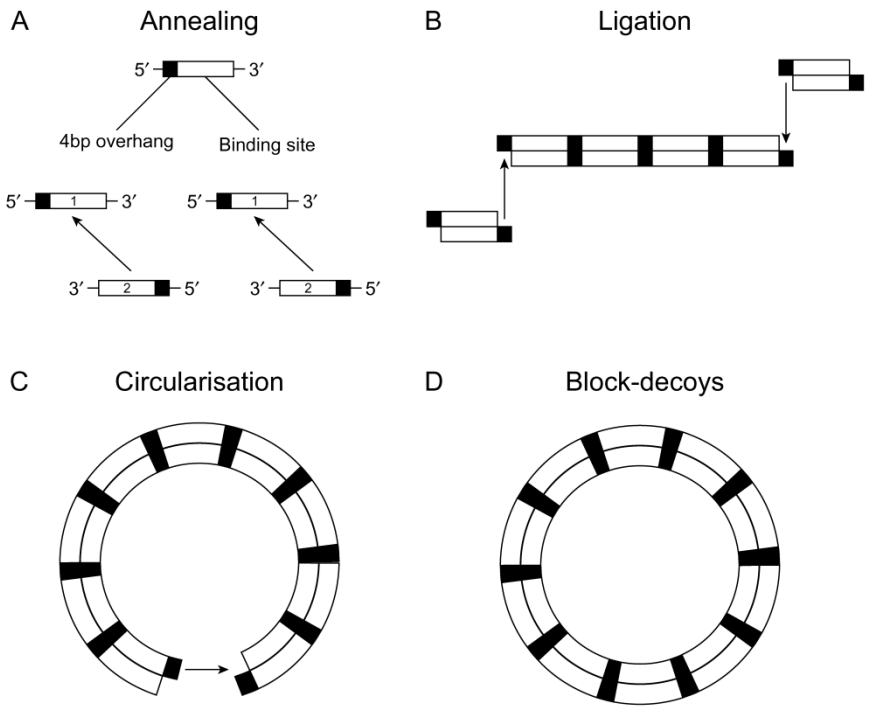


Figure 4.2: Schematic of block-decoy formation. (A) Single stranded oligonucleotides are annealed to form regulatory element-blocks containing a transcription factor binding site and a 4 bp single stranded overhang at 5' termini. (B) Regulatory element-blocks are ligated together into extending concatamers which circularise (C), allowing intramolecular ligation of cohesive termini, yielding covalently closed circular block-decoys containing multiple copies of the target binding site (D).

Block-decoy formation was confirmed by gel electrophoresis. This analysis showed that different TFRE-specific decoys constructed using the appropriate TFRE-blocks exhibited near-identical size distributions, with the vast majority of molecules between 100 – 300 bp in size (Figure 4.3). To test the hypothesis that circularisation of decoys prevented further ligation (thus limiting their size) purified block-decoys were utilised as the substrate in further ligation reactions (Figure 4.3). No variation in block-decoy size distribution was observed, indicating an absence of ligatable single stranded overhangs. Further, block-decoy stability against digestion by Exonuclease III, active against linear DNA, was evaluated (Figure 4.3). This analysis showed decoys were resistant to exonculease digestion and it was therefore concluded that this method of block-decoy construction yields circular ODNs containing approximately 7 – 20 copies of a target TFRE.

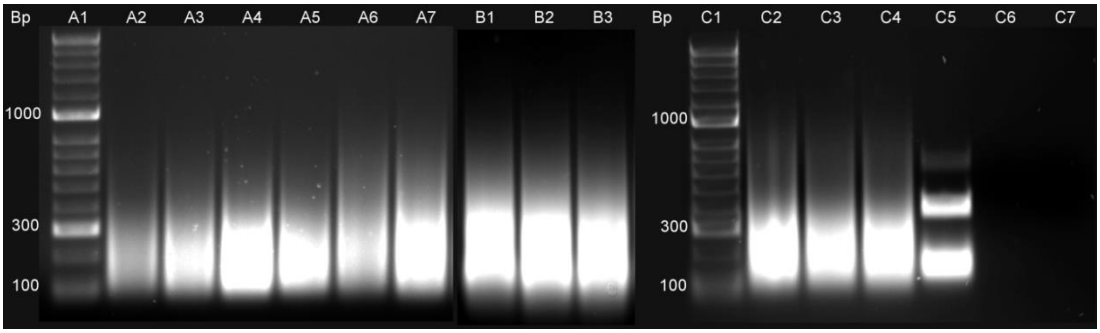


Figure 4.3: Circular block decoys contain numerous regulatory element binding sites. Agarose gel analysis of block decoys constructed from NFkB-RE (lanes A2,3), E-box (lanes A4,5) and CRE (lanes A6,7) regulatory element binding site blocks. Circularisation of a purified block decoy (B1) was demonstrated by (i) two further sequential ligation reactions (B2,3) which showed no additional increase in decoy size distribution and (ii) stability on digestion with Exonuclease III for 0, 1 and 6 h at 37°C (lanes C2-C4) respectively. Lanes C5, 6 and 7 show digestion of linear DNA sampled at the same time points.

4.2.3 Block-Decoy Function and Specificity

The work in chapter 3 (section 3.3.3) identified seven discrete TFREs that are transcriptionally active in CHO-S cells. Moreover, a panel of GFP and SEAP reporter plasmids containing 7 copies of each of these discrete TFREs upstream of a core CMV promoter were constructed and validated, where minimal reporter expression was observed with the core promoter alone (1 - 3% of reporter activity of TFRE-containing plasmids). These CHO-active TFRE reporters provide an ideal system for evaluating the use of block-decoys to inhibit the activity of specific TFREs.

Initially, reporter plasmids utilising NFkB-RE (the strongest element identified in the CHO-S TFRE functional screen) to drive reporter expression were used to validate the specific inhibitory effects of block-decoys *in vitro*. Measurement of GFP production after transient co-transfection of CHO-S cells with NFkB-reporter vector and varying concentrations of NFkB block-decoy is shown in Figure 4.4. This analysis identified that the NFkB-RE block-decoy inhibited expression from NFkB-RE-GFP reporter plasmid in a dose-dependent manner. In order to validate that the NFkB-RE block decoy specifically inhibited NFkB-RE reporter expression, CHO-S cells were co-transfected with i) NFkB-RE block-decoy and different GFP plasmids utilising varying TFREs (E-box, CRE) to drive reporter

expression and ii) NFkB-RE-GFP reporter and different TFRE-specific (E-box, CRE) block decoys. As shown in Figure 4.5 the concentration of NFkB-RE block-decoy exhibiting maximal inhibition of NFkB-RE-GFP reporter expression (2 μg/ml) had no significant effect on GFP expression from either CRE or E-box reporter plasmids. Further, block-decoys constructed from E-box and CRE TFRE-blocks did not significantly affect expression from NFkB-RE-GFP reporter plasmid. It was therefore concluded that block-decoys function to specifically sequester cognate TFRE-binding transcription factors, inhibiting expression from promoters dependent on their activity. Lastly, the stability of block-decoys *in vitro* was analysed by evaluating NFKB-RE block decoys in a longer-term cell culture process. As shown in Figure 4.6 decoys maintained significant inhibition of NFkB-RE mediated expression throughout a 4 day GFP production process. These data therefore indicate that block-decoys exhibit relatively long-term intracellular stability.

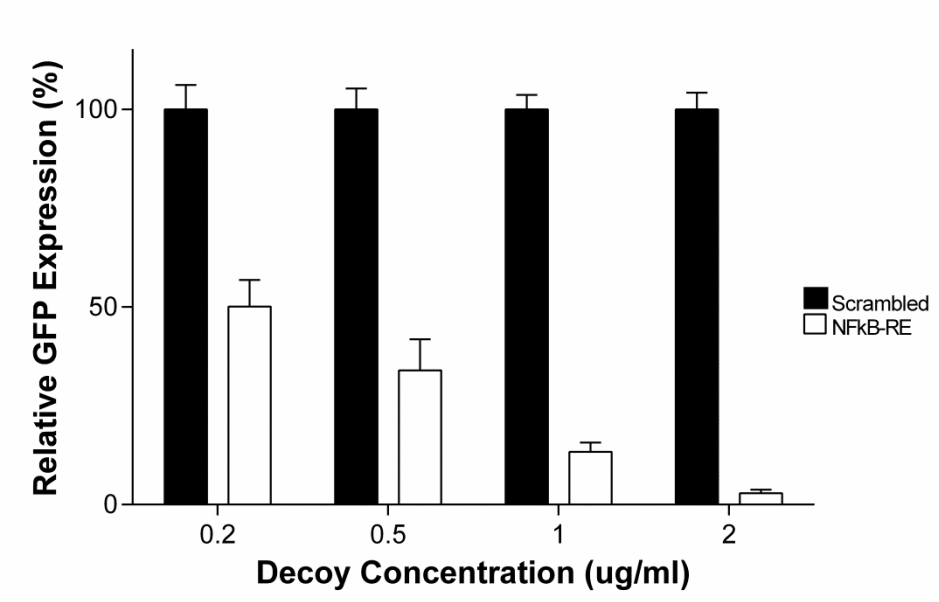


Figure 4.4: NFkB-RE block-decoys inhibit NFkB-RE mediated expression. CHO-S cells (2×10^5) were co-transfected with a NFkB-RE-dependent GFP reporter plasmid with either a NFkB-RE block decoy (white bars) or a scrambled NFkB-RE block decoy (black bars) at concentration of 0.2 – 2 μg/ml DNA per transfection. GFP expression was quantified 24 h post-transfection Data were normalised with respect to GFP expression in the presence of the scrambled decoy in each case. Bars represent the mean + SEM of three independent experiments each performed in triplicate.

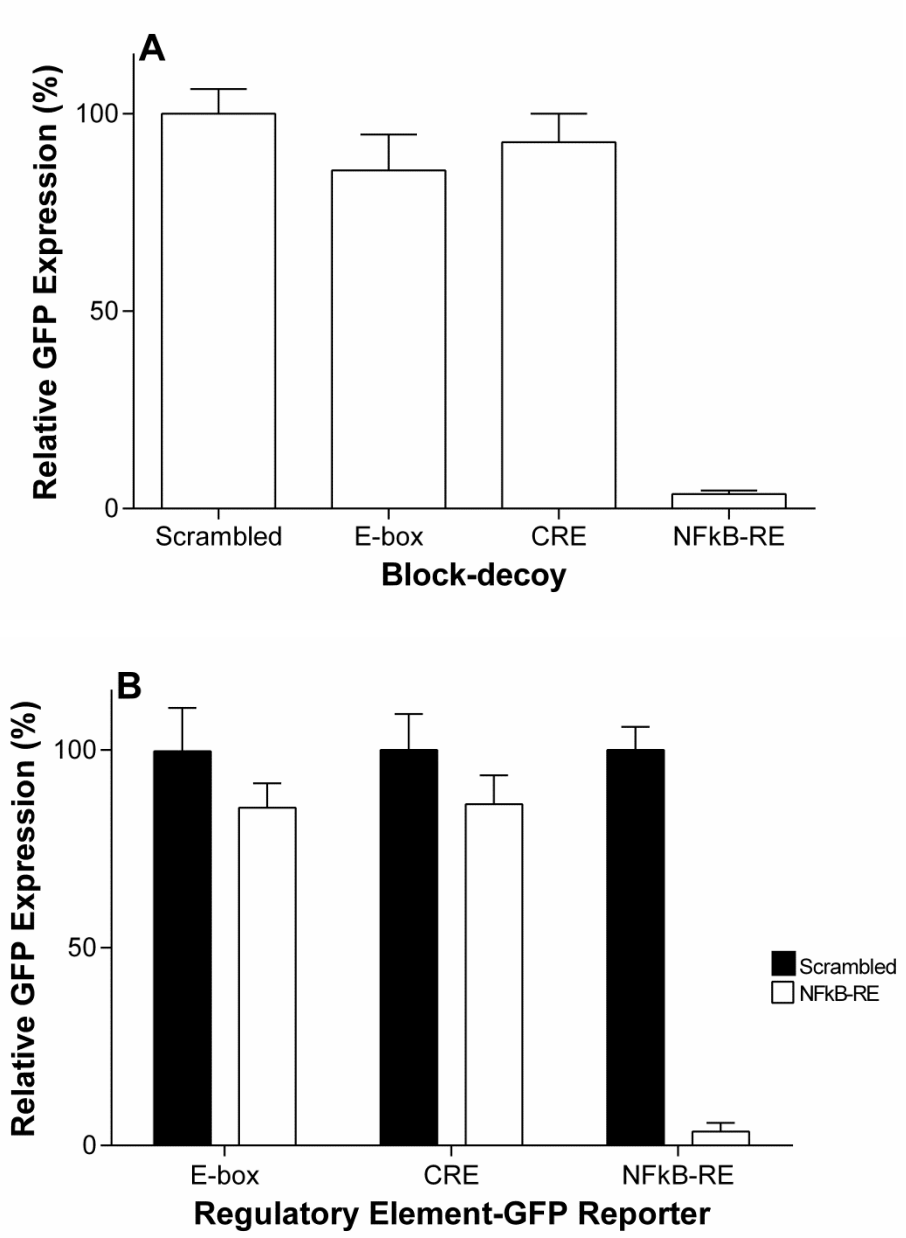


Figure 4.5: Circular block decoys specifically inhibit expression mediated by discrete regulatory elements. A) CHO-S cells (2 x 10^5) were co-transfected with a NFkB-RE-dependent GFP reporter plasmid with 2ug/ml scrambled NFkB-RE block decoy (control) or different regulatory element block decoys illustrating specific inhibition with NFkB-RE block decoy. **B)** Co-transfection of NFkB-RE block-decoy (white bars) or scrambled NFkB-RE block decoy (black bars) and different GFP reporter plasmids varying in transcription factor specificity (CRE, E-box or NFkB-RE) illustrating specific inhibition of NFkB-RE mediated reporter expression. GFP expression was quantified 24 h post-transfection. Data were normalised with respect to GFP expression in the presence of the scrambled decoy in each case. In A and B each bar represents the mean + SEM of three independent experiments each performed in triplicate.

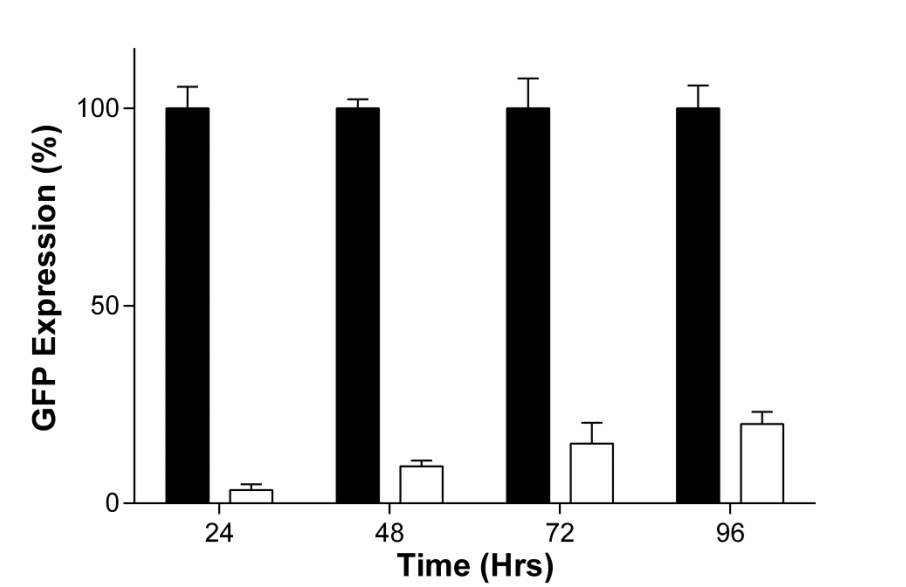


Figure 4.6: Block-decoys inhibit NFkB-RE mediated expression throughout a four day GFP production process. A) CHO-S cells (2 x 10^5) were co-transfected with a NFkB-RE-dependent GFP reporter plasmid with either a NFkB-RE block decoy (white bars) or a scrambled NFkB-RE block decoy (black bars) at concentration of 2 µg/ml DNA per transfection. GFP expression was quantified at varying timepoints post-transfection. Data were normalised with respect to GFP expression in the presence of the scrambled decoy in each case. Bars represents the mean + SEM of three independent experiments each performed in triplicate.

4.2.4 Constructing chimeric block decoys

A major advantage of the block-decoy strategy is that it can be utilised to construct stoichiometrically optimised chimeric decoys targeting multiple TFREs. In order to validate this functionality the specific inhibitory effects of two further TFRE-specific block-decoys were evaluated (E-box and CRE TFREs were selected as they exhibited the next highest transcriptional activities (after NFkB) in the CHO-S TFRE functional screen). As shown in Figure 4.7 both E-box and CRE block-decoys inhibited expression from corresponding TFRE-specific reporter plasmids in a dose-dependent manner. In both cases, TFRE-specific reporter expression was inhibited only by the corresponding block-decoy. Accordingly, chimeras specifically targeting multiple TFREs (NFKB-RE, CRE and E-box) could be constructed by ligating multiple discrete TFRE-blocks.

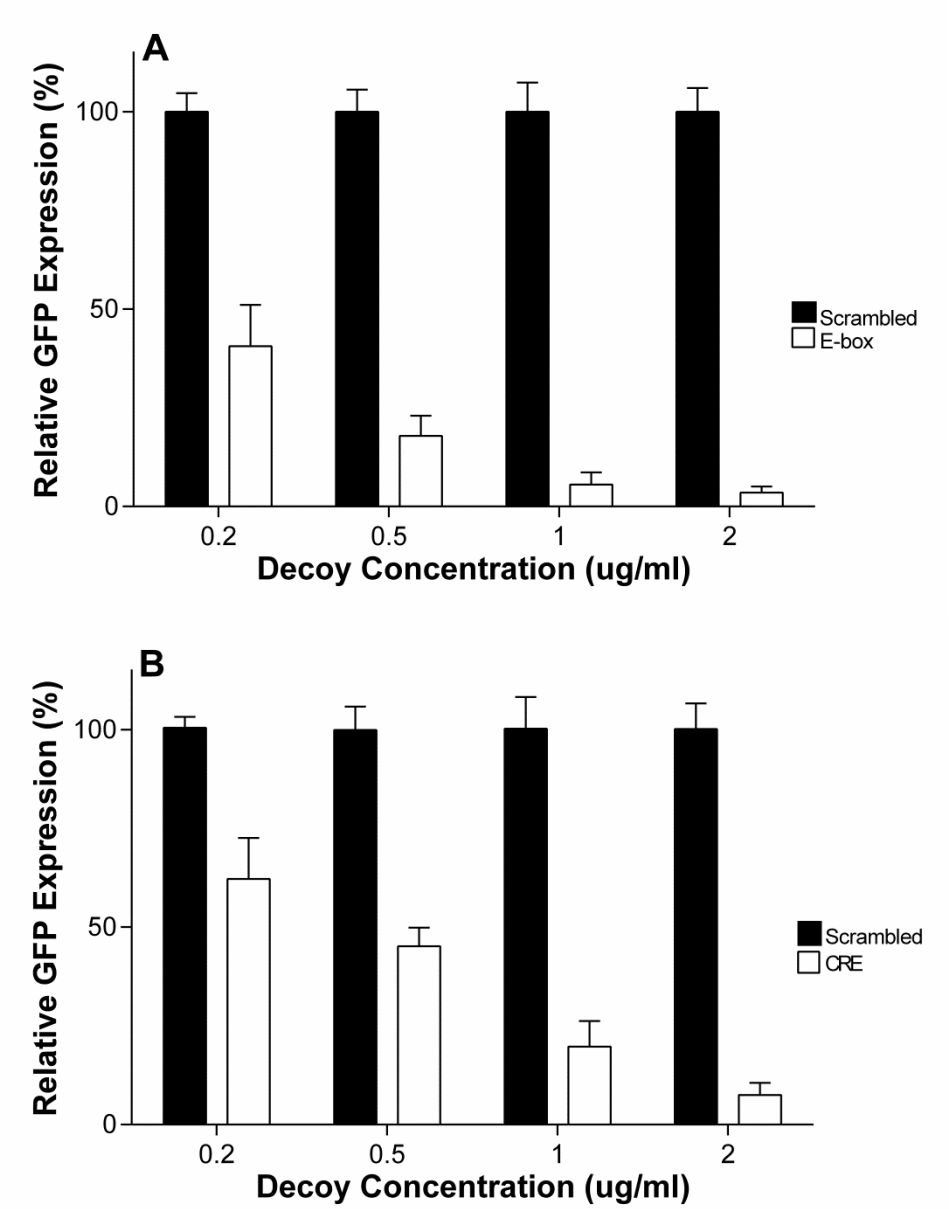


Figure 4.7: TFRE-specific block-decoys exhibit variable potency. A) CHO-S cells (2 x 10^5) were co-transfected with a E-box-dependent GFP reporter plasmid with either a E-box block decoy (white bars) or a scrambled E-box block decoy (black bars) at concentration of 0.2 – 2 µg/ml DNA per transfection. **B)** Co-transfection with a CRE-dependent GFP reporter plasmid with either a CRE block decoy (white bars) or a scrambled CRE block decoy (black bars). GFP expression was quantified 24 h post-transfection. Data were normalised with respect to GFP expression in the presence of the scrambled decoy in each case. Bars represent the mean + SEM of three independent experiments each performed in triplicate.

It is assumed that the relative extent to which each TFRE-specific block-decoy inhibits reporter expression from its corresponding TFRE-reporter plasmid is a function of block-decoy specific differences in (i) the relative intracellular abundance of TFs and (ii) TF-TFRE block binding kinetics. As shown in Figure 4.8 each TFRE-specific block decoy exhibited a characteristic inhibitory dose-response relationship, where at the highest concentrations expression from each corresponding TFRE-specific reporter was inhibited over 90%. Log transformation of block-decoy concentration data and nonlinear regression analysis enabled determination of the relative potency of each block-decoy, and revealed that their inhibitory potency occurred in the order: E-box > NFKB-RE > CRE, with a stoichiometry of E-box: 0.5: NFkB-RE: 0.8: CRE: 1.0 (calculated by interpolation to determine relative inhibitory concentrations). It was therefore concluded that equivalent inhibition of multiple TFREs would require stoichiometric tailoring of TFRE site copies within block-decoys according to the relative 'potency' of each TFRE-block.

To test the hypothesis that chimeric block-decoys could be constructed with controlled TFRE-block ratios, ligation reactions were evaluated where i) six distinct TFRE-blocks were ligated in an equimolar ratio, and ii) four TFRE-blocks were ligated in a 1.0 : 0.6 : 0.2 : 0.2 ratio. A "cloning-block" containing XhoI and KpnI sites was added to each ligation mix to enable subsequent linearisation. Linearised molecules were cloned into an acceptor plasmid and fifty clonally derived plasmids from each library were sequenced. The actual stoichiometric ratios of RE-blocks across the block-decoy libraries was approximately equal to the input ratios (1.18: 1.09: 1.03: 1.00: 0.96: 0.86 and 1.00: 0.57: 0.25: 0.22). It was therefore concluded that chimeric block-decoys containing stoichiometrically tailored quantities of TFRE-blocks could be constructed by controlling the molar ratio of TFRE-blocks in the ligation reaction, and therefore that cells transfected *in vitro* with chimeric molecules will contain TFRE-blocks at the specified concentrations across the cell population.

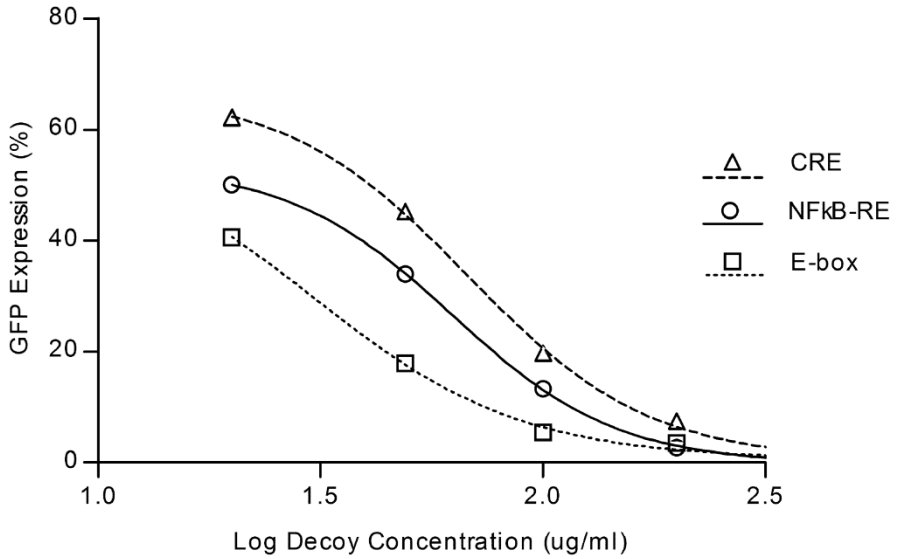


Figure 4.8: Stoichiometric optimisation of chimeric block-decoys targeting multiple regulatory elements. In order to determine the correct stoichiometry of different TFRE-blocks in chimeric decoys to achieve equivalent inhibition of each regulatory element, the relative ability of separate TFRE-blocks to inhibit TFRE-specific reporter expression was first quantified. CHO-S cells were separately co-transfected with three different TFRE-specific block decoys (CRE, NFkB-RE and E-box) or the corresponding scrambled block-decoy controls at varying block decoy concentration with the corresponding TFRE-specific GFP-reporter plasmids (at a ratio of decoy : reporter plasmid maintained at 1 : 1). GFP expression in block-decoy transfected cells is shown as a percentage of reporter expression in cells transfected with the same concentration of scrambled decoy control. Best fit curves obtained by non-linear regression analysis were utilised to determine the relative ratio of TFRE-specific blocks employed to construct chimeric decoys.

4.2.5 Chimeric block-decoys target multiple TFREs simultaneously

In order to demonstrate the novel capability of the block-decoy methodology to form stoichiometrically optimised chimeras, block decoys targeting multiple TFREs were constructed. In order to construct chimeric block decoys exhibiting maximal, equivalent inhibition of all TFRE-reporter plasmids the stoichiometry of TFRE-blocks in ligation reactions was adjusted according to the extent individual TFRE-specific block-decoys inhibited expression of the cognate TFRE-reporter (*i.e.* the relative 'potency' of each TFRE-block). Thus, to achieve concurrent inhibition of NFkB-RE and E-box to a similar extent using the block-decoy approach TFRE-

blocks were ligated in the stoichiometric molar ratio NFkB-RE 1.0 : E-box 0.62. Anticipating that chimeric decoys would require a greater concentration of decoy to be transfected to achieve a specific reduction in TFRE-reporter expression (as the number of copies of each TFRE-block is effectively reduced with an increase in the number of different TFRE-blocks utilised to construct a chimeric decoy) we (i) increased total TFRE-decoy DNA load per transfection and (ii) utilised alternative TFRE-SEAP reporter constructs to enable more sensitive detection of TFRE driven reporter expression. Preliminary experiments showed that a decoy concentration of 3.5 µg/ml decoy was the maximal decoy load that could be co-transfected with TFRE-reporter plasmid whilst still maintaining quantitation in the linear range from each TFRE-specific SEAP reporter plasmid (transfected at 2 µg/ml). Chimeric decoys were therefore transfected at this concentration. Figure 4.9 shows that the chimeric decoy constructed from NFkB-RE and E-box TFRE-blocks significantly inhibited expression from both TFREs at approximately equivalent levels. Moreover, it had no significant effect on GFP production from CRE-reporter plasmids, confirming specific inhibition of NFkB-RE and E-box mediated reporter expression.

In order to further demonstrate the functionality of block-decoys for investigating multi-TF mediated phenotypes in CHO cells, a chimeric decoy was constructed targeting all three TFREs at approximately equivalent levels by ligating TFRE-blocks in the stoichiometric molar ratio E-box 0.5 : NFkB-RE 0.8 : CRE 1.0. Utilising the maximal decoy concentration of 3.5 µg/ml, this equated to TFRE-block concentrations of 0.76 (E-box), 1.22 (NFkB-RE) and 1.52 µg/ml (CRE). Through interpolation of the single decoy data summarised in Figure 4.8 it was predicted that under these conditions chimeric decoys would inhibit expression from E-box, NFkB and CRE-SEAP reporter plasmids by 88%. Figure 4.9 shows that the chimeric decoy significantly inhibited expression from all three TFREs at approximately equivalent levels. E-Box, NFkB-RE and CRE dependent SEAP expression was reduced to 77%, 76% and 68% respectively, showing the chimeric decoy simultaneously sequestered a substantial proportion of the intracellular cognate TFs that bind to each of the three TFREs. The slight reduction in decoy potency compared to predicted values may be explained by (i) the reduced transfection efficiency associated with transfecting a higher concentration of DNA (resulting in fewer copies of each RE-block per cell) or (ii) TF-binding dynamics being affected by the presence of multiple RE-blocks. Nonetheless, the results show that three transcription factor binding pathways can be

inhibited simultaneously using a chimeric block-decoy containing stoichiometrically tailored quantities of each TFRE-block. This is the first time that a transcription factor decoy has been shown to target multiple elements by using an optimised number of copies of each binding site.

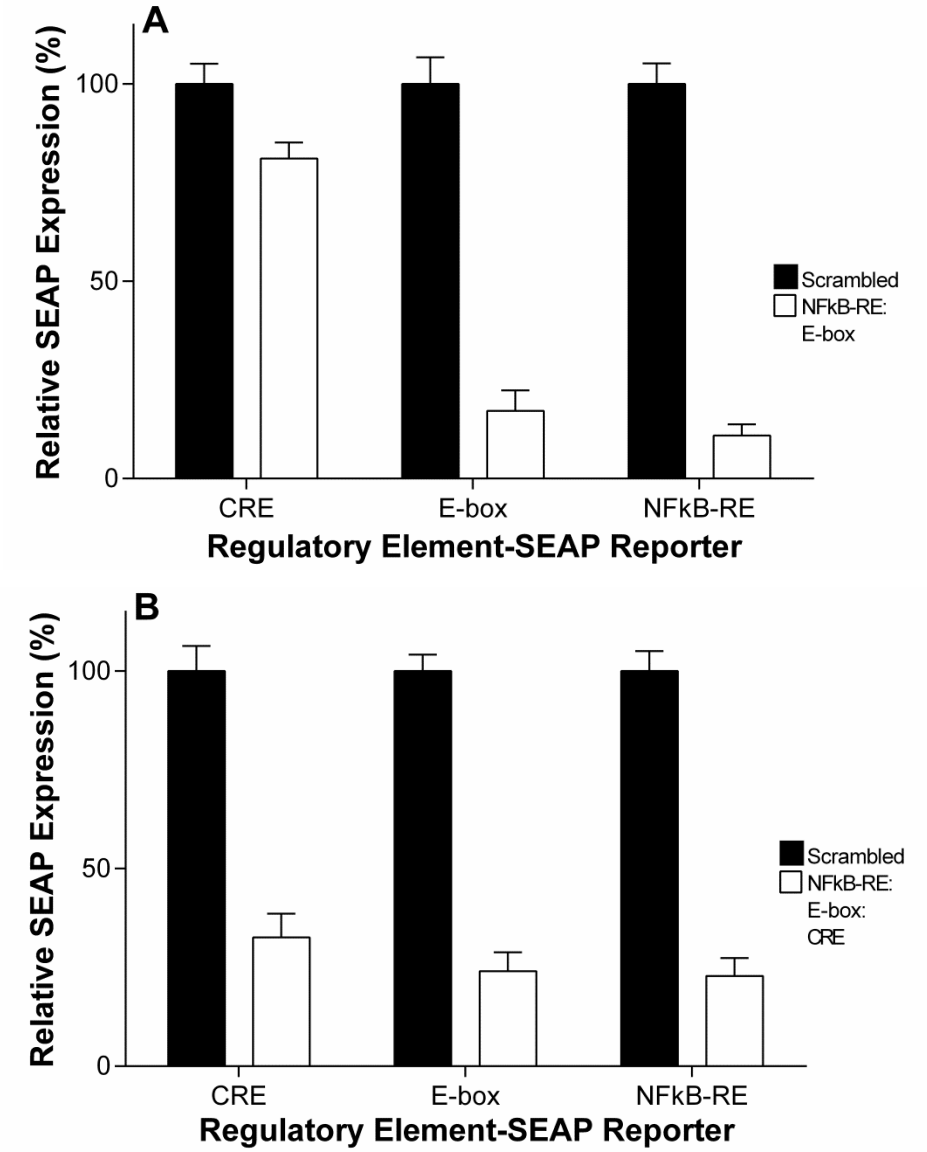


Figure 4.9: Chimeric block-decoys target multiple TFREs simultaneously. CHO-S cells (2×10^5) were co-transfected with 3.5 chimeric block decoys and 2 µg/ml of either CRE, E-box or NFkB-RE SEAP-reporter plasmid. Chimeric decoys were constructed using stoichiometric ratios of TFRE-blocks in the ratio **A)** NFkB-RE 1.0 : E-box 0.62 and **B)** CRE 1.0 : NFkB-RE 0.8 : E-box 0.5 (control scrambled chimeric decoys contained the same ratio of scrambled TFRE-blocks). SEAP expression was quantified 24 h post-transfection. Each bar shows SEAP expression in chimeric decoy treated cells relative to expression with the same concentration of scrambled decoy. In A and B each bar represents the mean + SEM of three independent experiments performed in triplicate.

4.3 Discussion

The work in this chapter describes a novel method for TF decoy formation that enables construction of chimeric decoys containing stoichiometrically optimised ratios of input TFRE-blocks. Exhibiting intracellular stability (resistance to exonuclease-degradation), chimeric block-decoys can target multiple TFREs simultaneously by specifically sequestering cognate TFRE-binding TFs. Specifically designed to enable characterisation of TF-TFRE interaction functionalities in CHO cells, block-decoys have significant advantages over existing decoy methods for studies requiring the simultaneous inhibition of multiple elements in defined combinations. Block-decoys can be applied to characterise any multi-transcription factor mediated CHO cell function (e.g. regulation of a specific promoter) or phenotype (e.g. regulation of cellular proliferation) and accordingly will facilitate both sophisticated transcriptional control and advanced cell engineering strategies.

Most studies utilising TF decoys in CHO cells are likely to require multiple decoys, targeting varying combinations of different binding motifs. For example, if a promoter of interest contains eight discrete TFREs there are over 200 possible unique chimeric combinations. By employing the block-decoy methodology to determine this promoter's functional regulation any of these chimeras could be rapidly constructed following the initial synthesis of eight TFRE-blocks. This compares to existing methods, where all chimeras would have to be designed and synthesised independently. The ability to rapidly construct hundreds of different chimeric combinations from a pool of input TFRE blocks provides substantial savings in both time and costs. Further, by utilising block-decoys the binding site ratios within each chimera could be adjusted to precisely control the relative extent to which each TFRE is inhibited in each transfected cell. This is a major advantage over the use of mixtures of single decoys, whose relative distribution within transfected cells is unpredictable. Moreover, adjustable control of TFRE-block ratios enables the optimal inhibition of each target element at any decoy concentration. Fine-tuning of binding site copy ratios enables more efficient inhibition, reducing the final decoy concentrations required and potentially facilitating more TFREs to be targeted simultaneously compared to existing methods. Block-decoys therefore offer significant advantages for concurrent inhibition of multiple TFREs *in vitro*, where

DNA load is often a restricting factor. Indeed, in this study where decoy concentrations were significantly limited by co-transfection with TFRE-reporters, block decoys enabled maximal equivalent inhibition (> 70%) of three TFREs simultaneously. Utilisation of increased decoy concentrations (as will be possible in the vast majority of block-decoy applications, which will not require vector co-transfection) will likely increase both the levels of inhibition and the number of elements that can targeted simultaneously.

Whilst the method's primary advantages are the ability to rapidly construct chimeric molecules and to control their binding site ratios, block-decoys have other potential benefits. Circular DNA has been associated with improved transfection efficiencies, compared to linear ODN (Chancham *et al*., 2001; Dhanoya *et al*., 2011). Further, multiple copies of the same binding site within a single decoy molecule may enhance TF sequestration (Gotea *et al*., 2010; Lee *et al*., 2012). It was previously shown that decoys with three site copies achieved stronger inhibition than those containing a single site (Gao *et al*., 2006). Therefore, the 7 – 20+ binding sites per block-decoy may enhance decoy function and efficiency.

Block-decoys can be utilised to determine the function of individual TFREs in the mechanistic regulation of discrete promoters in CHO cells. Given the current revolution in CHOmics (Datta *et al*., 2013) we are therefore in a position to i) determine CHO cell factory TF profiles throughout bioproduction processes and ii) identify the TFs required by specific promoters for transactivity. Accordingly, block-decoys will enable accurate prediction of discrete promoters' activities during bioprocesses and facilitate strategies (e.g. engineering of promoters, cell factories and/or bioreactor operating conditions) to optimise recombinant gene expression levels throughout production processes. For example, block-decoys could be employed to determine the mechanistic regulation of synthetic promoter 2/01 that was constructed in chapter 3. Intracellular levels of NFkB-RE, E-box, C/EBPα and GC-box binding TFs could then be specifically controlled/ monitored to ensure that sufficient quantities are maintained to facilitate maximal promoter activity during the production of both early stage products and easy to express proteins. Further, routine determination of synthetic promoters' functional regulation will facilitate CHO cell synthetic engineering strategies requiring the constitutive expression of several genes to be stoichiometrically balanced. Employing mulitple synthetic promoters to achieve precise gene expression ratios may be doomed to failure by promoter

intereferance – *i.e.* promoters exhibiting unpredictable activities due to 'competition' for availalbe TFs. For each synthetic promoter, block-decoys could be utilised to determine i) the TFREs required for activity, ii) how activity is affected by intracellular fluctations in TF abundances and iii) which TF-TFRE interactions are likely sources of activity variance. Compatible synthetic promoters could accordingly be selected for use in engineering attempts that i) individually display robust activity (e.g. activity not significantly impacted by minimal fluctations in TF abundances) and ii) collectively exhibit minimal promoter-promoter interference (e.g. promoters that do not rely on the same TF-TFRE interactions for activity). Lastly, an obvious application of block-decoys is in determining the TF-mediated regulation of critical functionalities such as proliferation, apoptosis, protein folding and glycosylation. Negative and positive effectors could accordingly be either knocked-out or over-expressed respectively in relatively simple cell engineering strategies. Moreover, neutral effectors could be utilised with their cognate TFs as inducible, controllable elements to construct regulated gene expression systems (switches, oscillators, synthetic regulatory cascades) for synthetic biology applications in CHO cells (Tigges and Fussenegger 2009; Weber and Fussenegger 2010).

Chapter 5: A mechanistic understanding of the CMV promoter's functional regulation in CHO cells.

In this chapter the mechanistic functionality of the hCMV-IE1 promoter is systematically deconstructed to identify the discrete TFREs that control its activity in CHO cells. In silico bioinformatics analysis (of both hCMV-IE1's TFRE composition and CHO cell's TF complement) identified a design space that was interrogated via TF sequestration (utilising block-decoys) and TFRE knockout (utilising synthetic CMV constructs with scrambled TFRE sequences). It was determined that i) the vast majority of CMV's activity within CHO cells is dependent on just two TFREs, NFkB and CRE, and ii) YY1 is a negative regulator of the CMV promoter in CHO. This mechanistic understanding of hCMV-IE1's functional regulation enables strategies to predictably control or improve its activity by engineering the promoter's TFRE composition or the cell factory's TF abundances.

5.1 Introduction

The human cytomegalovirus immediate early 1 (hCMV-IE1) promoter has been utilised for > 25 years to control recombinant gene transcription in CHO cells. hCMV-IE1 has been employed in the vast majority of regulatory approved production vectors, is currently used to drive expression of many biopharmaceutical products, and is the promoter of choice for transient gene expression (TGE) systems (Birch and Racher, 2006; Rita Costa *et al*., 2010). However, surprisingly given its long-term use, little is known about how it functions in the CHO cell and therefore strategies to precisely control or improve its transcriptional activity are not generally available. Determination of its mechanistic functionality in CHO cell factories would enable i) rational re-engineering to improve its bioproduction performance, ii) strategies to predictably control its activity during bioreactor processes and long-term sub-culture, and iii) identification of system-active 'parts' (both discrete TFREs and TFRE-combinations comprising cis-regulatory modules (CRMs)) that could be utilised to construct CHO-specific synthetic promoters (Li *et al*., 1999; Rajkumar and Maerkl, 2012; Blazeck *et al*., 2012; Brown *et al*., 2014). Reverse engineering of hCMV-IE1's functionality in CHO cells could therefore facilitate development of

improved gene expression control technologies for biopharmaceutical manufacturing.

In its natural context hCMV-IE1 drives expression of the major immediate early (IE) proteins IE1 and IE2 to initiate a hierarchical viral gene cascade (Isomura *et al*., 2004). Functioning as the master regulator of productive viral infection, hCMV-IE1 activity is the primary determinant of both cellular permissiveness and infection outcomes (*i.e.* abortive, lytic or latent) (Stinski and Isomura, 2008; Liu *et al*., 2013). Accordingly, in order to enable a broad host cell range, hCMV-IE1 is a highly complex element that has evolved to contain multiple discrete transcription factor regulatory elements (TFREs) cognate for common cellular transcription factors (TFs) (Stinksi and Isomura 2008). First characterised in the 1980s, its ability to deliver rapid, high levels of transcription in diverse cell types led to its establishment as the *de facto* choice for driving recombinant gene expression in mammalian cells *in vitro*. Still routinely employed in the vast majority of mammalian expression vectors hCMV-IE1 (hereafter referred to as CMV in this chapter) is perhaps the most recognisable and widely utilised genetic element in modern day molecular biology (Galle, 2013).

Unsurprisingly, the promiscuous strong activity of CMV has been utilised to drive transcription of biopharmaceutical products in CHO cells. Capable of driving constitutive high levels of target gene expression within CHO cell factories, it was a gold standard vector component for over 20 years (Hacker *et al*., 2009). However it has been associated with undesirable bioproduction characteristics including i) induction of cell stress and activation of apoptotic pathways, ii) cell cycle dependent activity, and most significantly iii) contribution to production instability via promoter silencing and gene deletion (de Boer *et al*., 2004; Dale 2006; Kim *et al.,* 2011). Accordingly, some contemporary production vectors employ alternative elements such as the Chinese hamster elongation factor-1α (CHEF-1α) promoter (Running Deer and Allison, 2004). Whilst endogenous regulatory elements offer some improved functionalities (e.g. enhanced expression stability), they are typically significantly less active than CMV and exhibit far less efficient transcriptional activity per unit DNA sequence (CHEF1α is > 15x larger than CMV). CMV therefore remains particularly attractive for use in situations where optimised productivity requires maximised transcription – for example in the production of easy to express proteins and in TGE (McLeod *et al*., 2011; Zhu *et al*., 2012). Its

functionality in the latter is likely to become increasingly important if TGE systems are developed for large-scale therapeutic protein manufacturing (Cain *et al*., 2013).

Numerous strategies have been employed in attempts to improve CMV functionality for biopharmaceutical production in CHO cells. For example, co-expression of the adenoviral E1 protein and heat shock at 42°C have both been shown to enhance CMV activity (Cockett *et al*., 1991; Pshenichkin *et al*., 2011). Further, compounds that promote euchromatic histone modifications (e.g. hyperacetylation and demethylation) such as sodium butyrate, 5-aza-2'deoxycitidine and dimethyl sulfoxide have been shown to increase CMV-driven transgene expression in CHO cells (Radhakrishnan *et al*., 2008; Choi *et al*., 2005). This is unsurprising given that the CMV promoter is known to recruit histone deacetylases (HDACs) and methyltransferases and is associated with heterochromatic histone modifications (*i.e.* CMV promoter silencing) during viral latency (Cuevas-Bennett and Shenk, 2008; Stinski and Isomura, 2008; Liu *et al*., 2013). Moreover, the same reagents have been shown to 're-activate' CMV in multiple human cell models, suggesting CMV may be silenced in CHO cells via commonly described mechanisms (Murphy *et al*., 2002; Grassi *et al*., 2003, Keller *et al*., 2007). Unfortunately these compounds are typically poorly compatible with bioreactor processes due to unpredictable and undesirable side-effects on critical cellular characteristics such as proliferation and apoptosis (Wang and Zhang, 2007; Jiang and Sharfstein, 2008). An alternative approach has utilised accessory elements from the CMV genome (unique region (UR), modulator, exonA, intronA) to increase protein production from CMV promoter-containing expression cassettes. This strategy has been employed by at least one well known contract manufacturer. Whilst employment of these four elements simultaneously has been shown to increase productivity (Mariati *et al*., 2010), multiple studies have indicated that the UR and modular components exhibit negative effects on CMV transcriptional activity (Huang *et al*., 1996, Chao *et al*., 2004; Lee *et al*., 2007). Enhanced productivity is likely a result of i) 5' UTR components (exonA and intronA) functioning via non-CMV specific, well-established mechanisms to increase mRNA abundance (increased RNA POL II processivity, enhanced mRNA stability and elevated 3'-end pre-mRNA processing), export and translation rates (Skoko *et al*., 2011; Bicknell *et al*., 2012), and/or ii) UR and modulator components exhibiting non-CMV specific, well established MAR-like insulator function in stable cell lines (Lashmit *et al*.,

2004; Stinski and Isomura 2008). Accordingly, these accessory viral regulatory elements are not considered to comprise a 'larger' CMV promoter (*i.e.* do not positively augment CMV's transcriptional activity), and are instead viewed as useful discrete expression cassette elements with non-promoter function (*i.e.* do not increase the rate of transcription initiation). Ultimately, CMV's functionality in bioreactor-scale protein production has not been improved in the past 25 years and tractable strategies to predictably control or rationally optimise its activity are currently unavailable.

The transcriptional activity of CMV is highly variable in different cell types. Cell-specific transactivational power is exhibited both in animal models *in vivo* (Mcgrew *et al*., 2004; Vasey *et al*., 2009; Mella-Alvarado *et al*., 2013) and in panels of mammalian cell lines *in vitro* (Qin *et al*., 2010; Schlabach *et al*., 2010*)*. As with any promoter, CMV's activity is a function of its constituent TFREs and the availability of cognate TFs within the cell (Coulon *et al*., 2013). CHO-specific mechanistic regulation of CMV activity is therefore a consequence of the functionality of one or more TF-TFRE interactions. The work in this chapter systematically deconstructs the complex CMV promoter to identify the discrete TFREs that control its activity in CHO cells. *In silico* bioinformatics analysis (of both the CMV promoter and CHO cells TF complement) and evaluation of the activity of discrete promoter regions identified a design space that was interrogated via TF sequestration (utilising block-decoys described in chapter 4) and TFRE knockout (utilising synthetic CMV constructs with scrambled TFRE sequences). Results indicated that the vast majority of CMVs activity within CHO cells is dependent on just two TFREs; NFkB and CRE. Further, YY1 was identified as a negative regulator of the CMV promoter in CHO. This knowledge will enable strategies to predictably control/ improve CMVs activity in CHO cells by engineering the promoter's TFRE composition and/or the factories' TF abundances.

5.2 Results

5.2.1. *In silico* identification of TFREs likely to regulate CMV promoter activity in CHO cells

In order to identify TFREs that are likely to functionally regulate CMV activity, the promoters TFRE composition was analysed *in silico* using online search tools that scan DNA sequences for TF binding sites. Transcription Element Search System (TESS) and Transcription Affinity Prediction tool (TRAP) were utilised, employing stringent search parameters to minimise false positives (Manke *et al*., 2008; Schug 2008). The CMV promoter is defined as the sequence spanning from -560 to +50 relative to the transcriptional start site of the IE1 gene within the viral genome, comprising three structurally distinct, synergistically functioning components - the distal, proximal and core promoters between -560 to -300, -300 to -50 and -50 to + 50 respectively (Stinski and Isomura, 2008). The bioinformatics analysis confirmed CMV to be an extremely complex genetic element, containing 42 separate binding sites distributed throughout the distal and proximal promoters at a frequency of one site per every 12.4 bp (Figure 5.1). A total of 12 discrete TFREs were identified at copy numbers ranging from 1 – 8, where the most abundant (> 3 sites) were YY1, NF1, RARE, GC-box, NFkB and CRE (comprising both CREB and AP-1 binding sites due to observed functional redundancy (van Dam and Castellazzi, 2001; Manna and Stocco, 2007).

In order to evaluate the likely functional relevance of the TFREs identified by bioinformatics analysis, a comprehensive literature review was carried out of previous studies investigating CMV's mechanistic regulation in other cell hosts. Given the CMV promoter's function as a master regulator of hCMV infection outcomes its activity has been extensively characterised in human cell lines. As shown in Table 5.1, 10 of the 12 TFREs have previously been identified as functional effectors of CMV activity in other cell systems (no additional TFREs were identified). Many of these TFREs have been differentially identified as neutral, negative or positive regulators of CMV activity, depending on cellular host and expression conditions (lytic viral infection, recovery from latency, or transient expression) (Galle, 2013). Previous studies therefore both i) highlighted the complex

and context-specific regulation of the CMV promoter, and ii) provided empirical evidence suggesting that the vast majority of elements identified *in silico* had the potential to regulate CMV's activity in CHO cells.

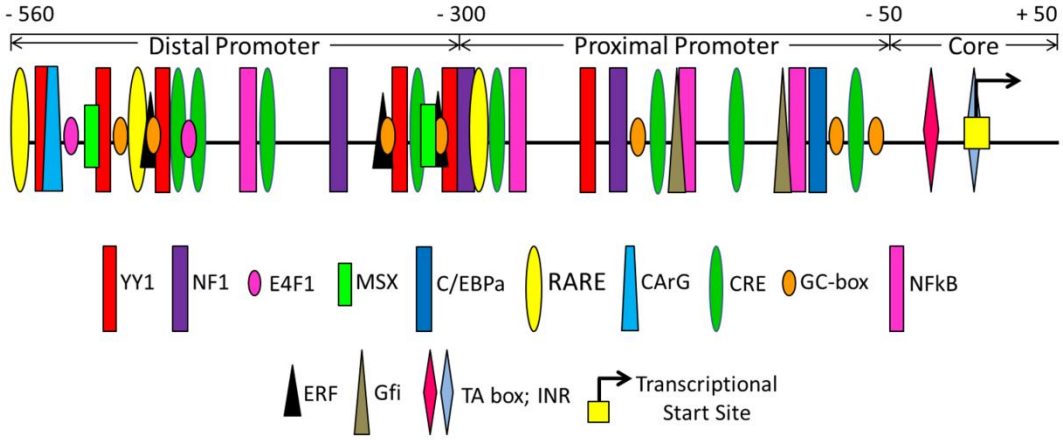


Figure 5.1: The CMV promoter contains multiple copies of numerous discrete transcription factor regulatory elements. The human cytomegalovirus immediate early 1 promoter was surveyed for the presence of discrete transcription factor regulatory elements using Transcription Element Search System (TESS) and Transcription Affinity Prediction (TRAP) algorithms using stringent search parameters to minimise false positives.

Table 5.1: Previous empirical evidence showing discrete TFREs functionally regulate CMV promoter activity. *In silico* analysis of the CMV promoter identified 12 discrete constituent TFREs. Literature databases were comprehensively searched to identify previous studies that have shown these TFREs to be functional regulators of CMV activity in any cell host. An example reference is given for each TFRE.

TFRE	**Evidence of effector function**
CArG	Caposio *et al*., 2010
C/EBP	Prosch *et al*., 2001
CRE	Lashmit *et al*., 2009
E4F1	X
ERF	Wright *et al*., 2004
GC-box	Isomura *et al*., 2005
GFI1	Zweidler-Mckay *et al*., 1996
MSX	X
NF1	Lashmit *et al*., 2009
NFkB	Liu *et al*., 2010a
RARE	Ghazal *et al*., 1992
YY1	Pizzorno, 2001

Previous work in this thesis (section 3.2.3) determined the relative transcriptional activity of 11 of the 12 (ERF was not included in the screen) CMV-constituent TFREs in CHO-S cells (Figure 5.2). Five of these elements could independently mediate recombinant gene expression in CHO-S cells using available TF activity. It was therefore hypothesised that these five CHO-active TFREs (NFkB, CRE, GC-box, E4F1, C/EBPα) were likely regulators of CMV activity in CHO. Whilst TFRE inactivity may have been a consequence of assay-design features (*i.e.* suboptimal TFRE sequences, TFs unable to drive transactivation independently) it could be evidence that cognate TFs for the six CMV-constituent CHO-inactive TFREs are not expressed in CHO cells. It is predicted that CHO cells may express a particularly narrow range of transactivators given their loss of specific functionalities over the course of long-term (> 50 years) culture *in vitro* (Lewis *et al.,* 2013; Xu *et al*., 2011). In order to evaluate the availability of cognate TFs for CMV-constituent TFREs within CHO cells, CHO genomic (Xu *et al*., 2011), transcriptomic (unpublished data available in-house; ~ 40 Gb of transcriptome sequence data from exponentially growing ECACC CHO-K1 cells cultured in CD-CHO) and proteomic data (Baycin-Hizal *et al*., 2012) were analysed. As shown in Table 5.2, the CHO cell TF complement contains potential cognate binding partners for 7 of the 12 CMV-constituent TFREs.

It is noted that i) unknown TFs may bind to TFREs, ii) TF expression may vary between different CHO-cell hosts, iii) TFs are particularly challenging to detect by proteomics (Hargrove *et al*., 1989; Baycin-Hizal *et al*., 2012) and iv) transcriptomic/ proteomic data sets are rarely exhaustive. However, the presence of available cognate TFs in two separate studies utilising two distinct CHO cell-lines clearly identifies TFREs that are likely to be regulators of CMV activity in CHO (C/EBP, CRE, GC-box, NF1, NFkB, RARE, YY1). Moreover, four TFREs for which cognate TFs were not identified (CArG, ERF, GFI, and MSX) were also inactive in the CHO-S TFRE functional screen and present at relatively low copy numbers (< 3 sites) in the CMV promoter. Therefore, based on the *in silico* analyses it was predicted that CMV's activity within CHO cells was a functional consequence of the interplay between one or more of the following eight TFREs: C/EBP, CRE, EF41, GC-box, NF1, NFkB, RARE, and YY1.

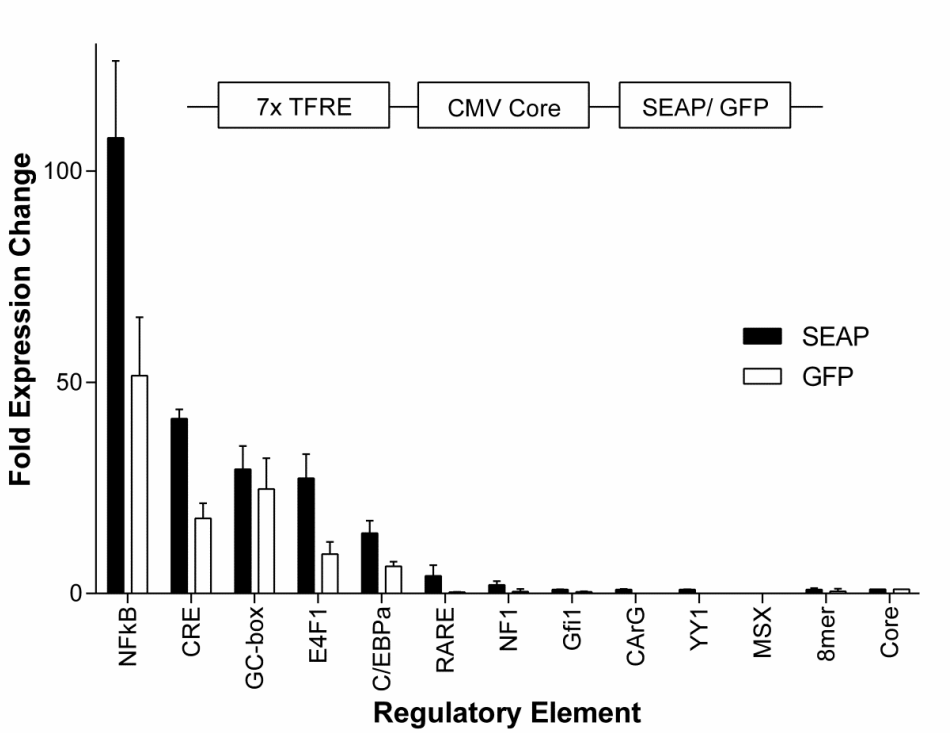


Figure 5.2: CMV-constituent TFREs exhibit variable activity in CHO-S cells. Seven copies of each TFRE were cloned in series upstream of a minimal CMV core promoter in reporter vectors encoding either GFP or SEAP reporters. CHO-S cells (2 x 10^5) in 24-well plates were transfected with 1 μg of SEAP (black bars) or GFP (white bars) TFRE reporter-vector. SEAP activity in cell culture supernatant and intracellular GFP were measured 24 h post-transfection. Data are expressed as a fold-change with respect to the activity of a vector containing only a CMV core promoter (Core). A random 8bp sequence with no known homology to TFRE sequences (8mer) was also used as a control. Bars represent the mean + SD of three independent experiments each performed in triplicate, using three clonally derived plasmids for each TFRE-reporter construct. Modified from Figure 3.7.

Table 5.2: The CHO cell TF complement contains potential cognate binding partners for multiple CMV-constituent TFREs. CHOmics data was analysed to evaluate the availability of cognate TFs for CMV-constituent TFREs within CHO cells. Genomic (Xu *et al*., 2011) and proteomic (Baycin-Hizal *et al*., 2012) data were accessed at CHOgenome.org. Unpublished transcriptomic data (~ 40 Gb of transcriptome sequence data from exponentially growing CHO-K1 cells) was available in-house. A maximum of four cognate TFs are listed for each TFRE within each dataset.

	Cognate TFs present		
TFRE	**Genome**	**Transcriptome**	**Proteome**
CArG	SRF	X	X
C/EBP	C/EBPγ; C/EBPδ; C/EBPε; C/EBPζ	C/EBPγ; C/EBPε	C/EBPζ
CRE	ATF1; CREB1; c-Fos; c-Jun	ATF1, CREB1, c-Jun	ATF1; CREB1; JunB
E4F1	E4F1	X	X
ERF	ERF-like	X	X
GC-box	EGR1; GCFC1-like; SP1-like; SP2-like	EGR2; GCFC1	GCFC1-like; SP3
GFI	GFI-1; GFI-1B	X	X
MSX	MSX1; MSX2; MSX3	X	X
NF1	NF1A; NF1B; NF1C; NF1X	NF1A; NF1B; NF1C; NF1X	NF1B; NF1X
NFKB	NFKB-p100; NFKB-p105; REL; RELA	NFkB-p105; RELA; RELB	REL; RELA; RELB
RARE	RARα; RARβ; RARγ; RXRα	RARα; RARβ; RARγ; RXRα	RXRβ
YY1	YY1-like	X	YY1

5.2.2. Functionally-redundant TFREs synergistically regulate the CMV promoter's activity in CHO cells

In order to evaluate the functional contribution of discrete CMV structural components to promoter activity, GFP reporter constructs containing either the distal or proximal promoter upstream of the CMV core were created. Their relative activity

compared to the full CMV promoter (distal + proximal) was determined by transient production as it provides both a model system of CMV's predominant modern-day function (*i.e.* driving transient protein production) and a direct readout of transactivation without potential interference from integration-specific effects or silencing. Whilst GFP production is not a direct measurement of transcriptional activity, assay conditions were optimised such that CMV-GFP reporter activity was in the centre of the linear assay range with respect to plasmid copy number (DNA load) and measured GFP output. Measurement of GFP production after transient transfection of CHO-S cells with each reporter plasmid is shown in Figure 5.3. This analysis identified that the distal and proximal promoters exhibited 49% and 76% of the activity of the full CMV promoter, where minimal reporter expression was observed with the core promoter alone (< 1% of reporter activity of CMV-GFP). The data therefore indicated that whilst both components exhibited significant activity (*i.e.* both contained CHO-active TFREs), the proximal promoter TFRE composition contained greater transactivation potential. Further, given that the cumulative activity from both components in isolation was greater than that of the full promoter, the analysis suggested that CMV contains functionally redundant TFREs.

In order to both i) further evaluate the functional redundancy of CMV-constituent TFREs and ii) determine whether transcriptional activity could be pinpointed to smaller discrete promoter regions, six cis-regulatory modules (CRMS; 100bp regions containing > 5 TF binding sites) from within the CMV promoter were cloned upstream of a CMV core in GFP reporter constructs (Figure 5.4). The relative transcriptional activity of CMV-CRMs is shown in Figure 5.4. The majority of CMV-CRMs had minimal activity compared to the full CMV promoter. The data showed that CMV was 'greater than the sum of its parts' as the cumulative activities of the six CRMs (that collectively contained the promoter's entire TFRE composition) totalled only ~ 60% of that of full CMV promoter. It was therefore inferred that CMV-constituent TFREs function synergistically to regulate the promoter's activity in CHO cells. Whilst CMV apparently contains functionally redundant TFREs, the CRM data indicated that positive effectors of activity are widely distributed throughout the sequence and further highlighted that they are relatively more abundant within the proximal, as compared to the distal, promoter.

Whilst significant activity could not be localised to smaller discrete promoter regions, preventing a simplified analysis of mechanistic regulation, comparison of

CRM's relative activities and TFRE compositions enabled determination of putative TFRE functionalities. For example, the two most active CRMs (CRM 5 and 6) had very similar TFRE compositions, where NFkB, CRE and GC-box elements accounted for 5/6 of the constituent TFREs. The data therefore indicated that one or more of these three TFREs were positive regulators of CMV activity in CHO. This was unsurprising given that previous work in this thesis determined NFkB, CRE and GC-box elements to be functional effectors of CHO-specific synthetic promoter activities. However, CRMs 3 and 4 both contained multiple copies of these TFREs yet exhibited virtually no transcriptional activity. Whilst more complex design rules may have governed NFkB, CRE and GC-box function within CRM constructs (e.g. spatial effects, combinatorial interactions), it was predicted that the inactivity of CRMs 3 and 4 was a consequence of negative effector TFREs. Both constructs contained multiple NF1 and YY1 sites, indicating that one or both of these elements may function to negatively regulate CMV activity in CHO. It was therefore concluded that the mechanistic regulation of CMV promoter activity in CHO cells likely involves TFs binding at discrete TFREs to repress transcriptional activity.

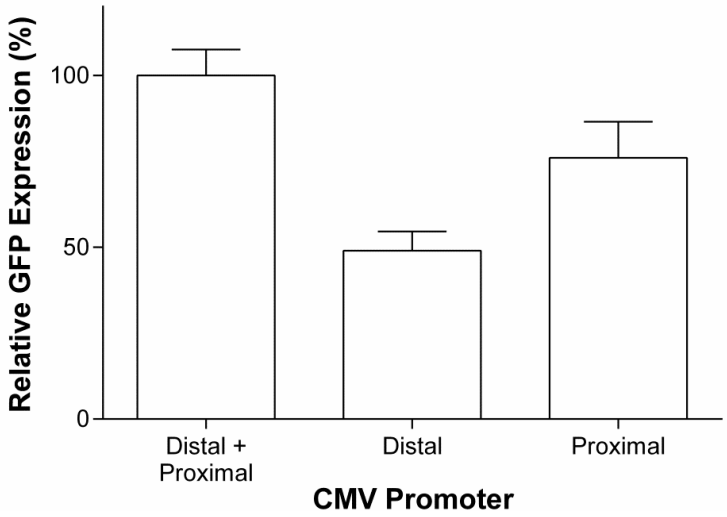


Figure 5.3: Discrete CMV promoter structural components exhibit differential activity in CHO-S cells. The CMV promoter comprises two structurally distinct, synergistically functioning components (distal and proximal promoters; see Figure 5.1 for relative TFRE compositions). CMV distal and proximal promoters were cloned upstream of the CMV core promoter in GFP reporter plasmids. CHO-S cells (2×10^5) in 24-well plates were transfected with 300ng of CMV reporter-vector and GFP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by the full CMV (distal + proximal) promoter. Bars represent the mean + SD of three independent experiments each performed in triplicate.

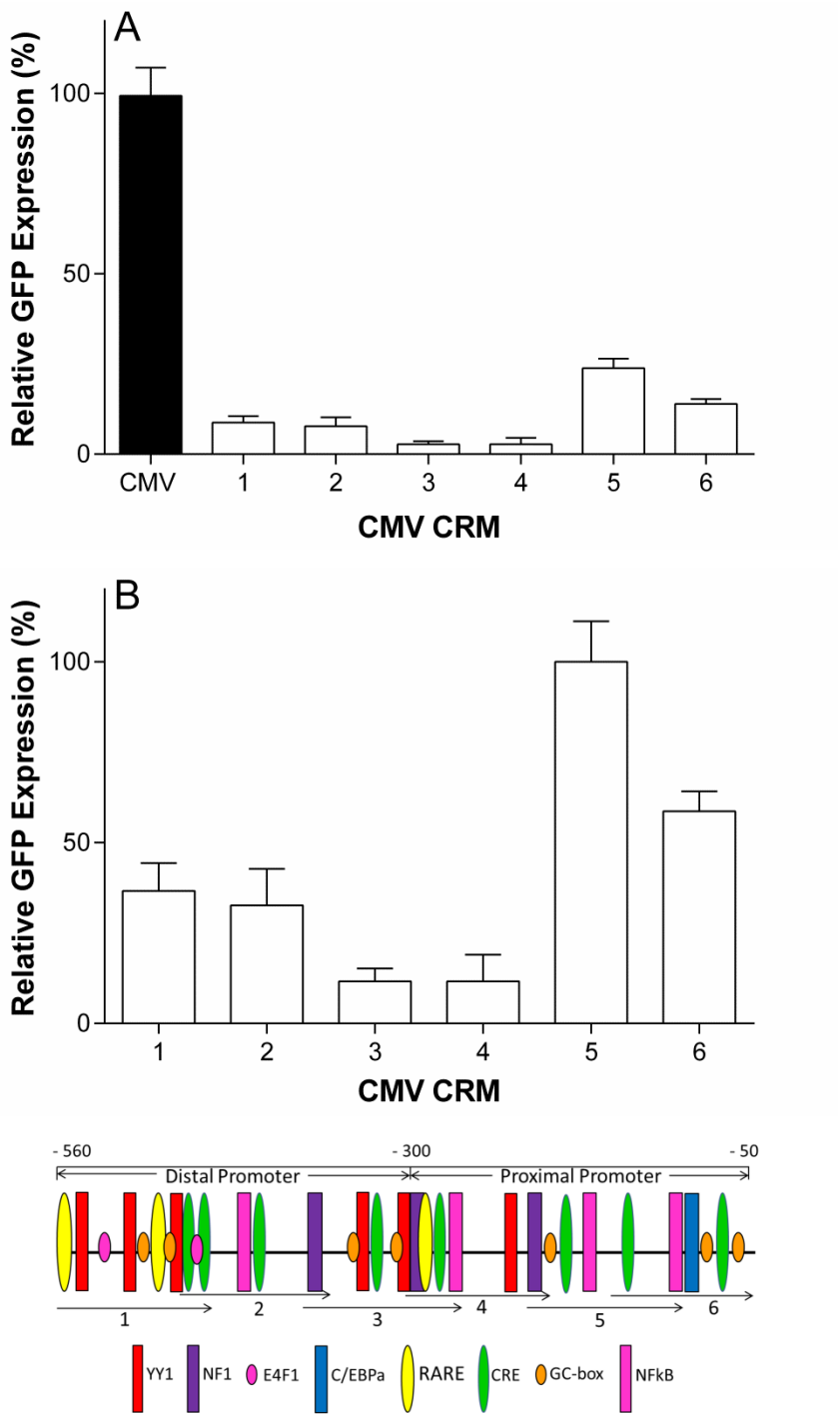


Figure 5.4: Discrete cis-regulatory modules from within the CMV promoter exhibit differential activity in CHO-S cells. Cis-regulatory modules (CRMs) from within the CMV promoter were cloned upstream of the CMV core in GFP reporters. CRM TFRE compositions are shown. CHO-S cells (2 x 10^5) in 24-well plates were transfected with 300ng of CMV-CRM reporter-vector and GFP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by either **A)** the full CMV promoter or **B)** the most active CRM. Bars represent the mean + SD of three independent experiments each performed in triplicate.

5.2.3. YY1, NFkB and CRE are functional regulators of CMV activity in CHO cells

In order to specifically determine whether discrete TFREs functionally regulated CMV activity in CHO cells, TF decoy molecules targeting the eight TFREs identified *in silico* as likely effectors of promoter activity were constructed. TF decoys, short synthetic oligodeoxynucleotides containing a specific TF binding element, compete for available intracellular TFs and prevent their association at target promoters. This site-specific TF sequestration makes them an ideal method for determining the functional contribution of individual TFREs to a promoter's activity. The block-decoy methodology of decoy formation that was developed previously in this thesis and designed specifically for characterising TF-TFRE interaction functionalities in CHO cells was utilised (Brown *et al*., 2013). In order to facilitate the use of maximal decoy concentrations, CMV-SEAP reporter constructs were created to enable more sensitive detection of CMV driven reporter expression (*i.e.* to reduce reporter-plasmid DNA load). It was previously shown that three different TFRE-specific block-decoys were able to specifically inhibit > 90% of expression from promoters dependent on their target TFREs at decoy concentrations of 2 μg/ml (Brown *et al*., 2013). Accordingly, to ensure maximal inhibition of the transactivation mediated by each CMV-constituent TFRE, decoy concentrations of 4 μg/ml were employed in this study. Preliminary experiments confirmed that SEAP quantitation was in the linear assay range with respect to plasmid copy number and measured SEAP output when CHO-S cells were co-transfected with 4 μg/ml decoy and 1 μg/ml CMV-SEAP-reporter (moreover, experiments within this laboratory have confirmed that SEAP activity in cell culture supernatant is linearly correlated with SEAP mRNA levels post-transfection, providing a direct measurement of transcriptional activity).

Measurement of SEAP production after transient co-transfection of CHO-S cells with CMV-SEAP-reporter and each TFRE-specific decoy is shown in Figure 5.5. This analysis identified that CMV activity was increased > 1.5 fold by YY1 decoys, suggesting that YY1 binds to cognate sites within the CMV promoter to negatively regulate its transactivation. In contrast, CMV activity was reduced by 38% and 49% by CRE and NFkB decoys respectively. The data therefore indicated that CMV activity in CHO was highly dependent on TFs binding at these two

discrete elements. Although CMV activity was not significantly affected by the other five TFRE-specific decoys (C/EBP, E4F1, GC-box, NF-1, RARE) it is assumed that the relative extent to which each TFRE-specific block-decoy inhibits expression from its target TFRE is a function of block-decoy specific differences in both the relative intracellular abundance of TFs and TF-TFRE block binding kinetics. Therefore, it is noted that the block-decoy approach does not provide definitive proof of neutral effector function (*i.e.* false-negatives could have occurred). Moreover, as CMV apparently contains functionally redundant TFREs (Figure 5.3), it was hypothesised that identified neutral effectors may have regulatory function in the absence of NFkB and CRE-mediated transactivation. In order to both evaluate this hypothesis and determine if the two identified positive regulatory elements (NFkB and CRE) functioned synergistically to transactivate CMV, a chimeric decoy targeting both CRE and NFkB was created by ligating NFkB and CRE TFRE-blocks at a 1: 1 stoichiometric molar ratio. In order to maintain decoy loads and TFRE block copy numbers, single target-decoys (NFkB or CRE) were constructed by ligating consensus and scrambled TFRE blocks at a 1: 1 molar ratio (e.g. CRE decoy constructed by ligating CRE-consensus and NFkB-scrambled TFRE blocks). Anticipating that chimeric decoys would require a greater concentration of decoy to be transfected to achieve a specific reduction in TFRE-mediated expression (as the number of copies of each TFRE-block is effectively halved per decoy molecule) an increased TFRE-decoy DNA load was utilised per transfection. Preliminary experiments showed that a decoy concentration of 6 µg/ml was the maximal decoy load that could be co-transfected whilst still maintaining SEAP quantitation in the linear range from the CMV-SEAP reporter plasmid (transfected at 1 µg/ml). As shown in Figure 5.5, CMV-driven SEAP expression was reduced by 77% by the chimeric decoy, where decoys targeting either CRE or NFkB individually reduced CMV activity by 34% and 46% respectively. The analysis therefore indicated that i) the vast majority of CMVs activity in CHO cells is a functional consequence of TF-TFRE interactions at NFkB and CRE sites and ii) a large proportion of the CMV sequence is redundant in CHO, comprising multiple TFREs (C/EBP, E4F1, GC-box, NF-1, RARE) that exhibit minimal transactivation potential.

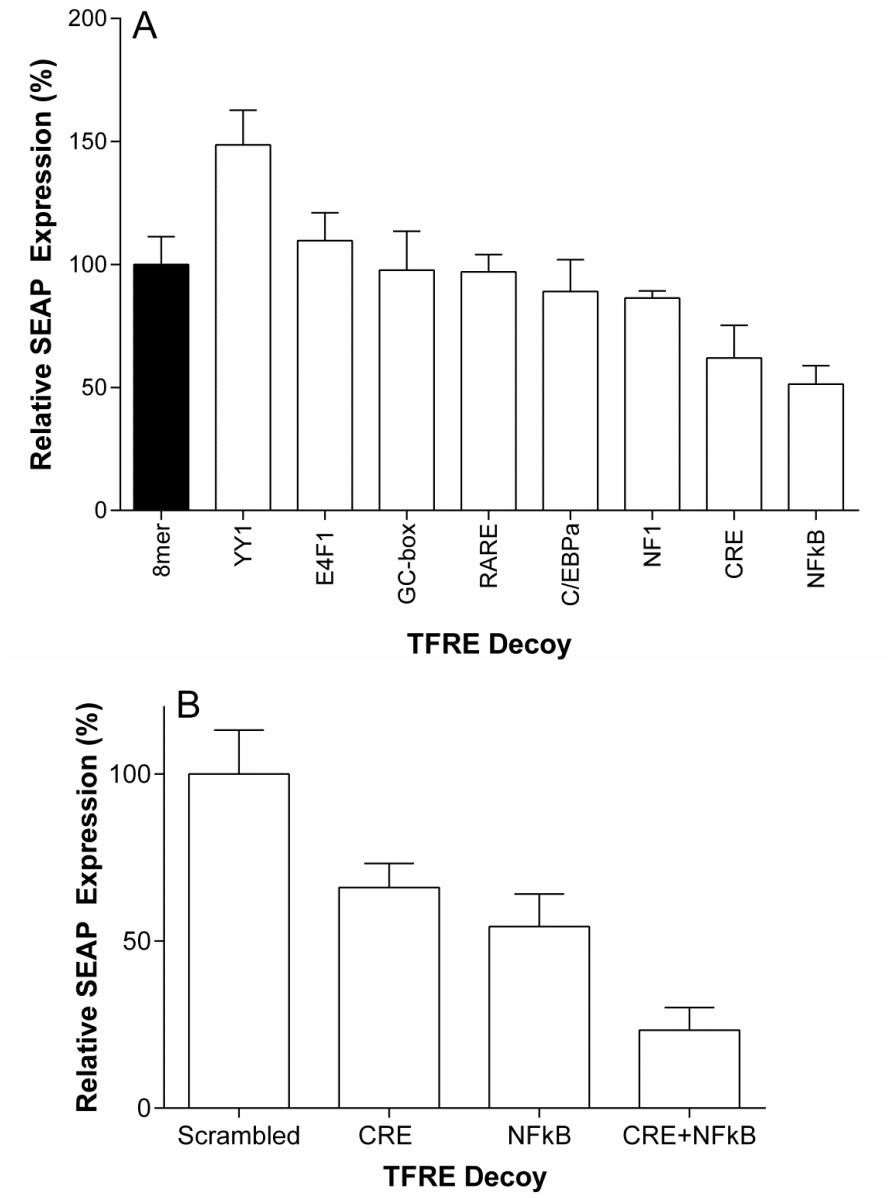


Figure 5.5: YY1, NFkB and CRE are functional regulators of CMV activity in CHO-S cells. A) CHO-S cells (2 x 10^5) were co-transfected with 1 µg/ml CMV promoter-SEAP reporter plasmid and 4 µg/ml block-decoys targeting different transcription factor regulatory elements. **B)** CHO-S cells (2 x 10^5) were co-transfected with 6 µg/ml chimeric block decoys and 1 µg/ml CMV-promoter reporter plasmid. Chimeric block decoys targeting both NFkB and CRE (NFkB + CRE) were constructed by ligating CRE and NFkB TFRE-blocks at a stoichiometric molar ratio of 1: 1 (control scrambled chimeric decoys contained the same ratio of scrambled TFRE-blocks). Chimeric decoys targeting CRE or NFkB were constructed by ligating consensus and scrambled TFRE blocks at a 1: 1 molar ratio (e.g. CRE decoy constructed by ligating CRE-consensus and NFkB-scrambled TFRE blocks). SEAP expression was quantified 24 h post-transfection. Each bar shows SEAP expression in decoy treated cells relative to expression with the same concentration of either A) block-decoys containing a random 8bp sequence with no known homology to TFRE sequences (8mer control) or B) scrambled decoy control. In A and B each bar represents the mean + SD of three independent experiments performed in triplicate.

5.2.4. A CMV promoter devoid of NFkB and CRE sites exhibits minimal activity in divergent CHO cell hosts

Decoy-mediated physical sequestration of YY1, NFkB and CRE-binding TFs could theoretically have reduced transient protein production via non-CMV specific mechanisms by affecting the regulation of multiple endogenous genes. Therefore, in order to confirm that TF-TFRE interactions at these elements specifically regulated CMV activity in CHO cells, synthetic CMV constructs were created with YY1, NFkB and CRE binding sites 'deleted'. NFkB (all 4 NFkB sites scrambled to AATCGCAAGT), CRE (all 8 CRE sites scrambled to CTACTGTG) and NFkB + CRE (all NFkB and CRE sites scrambled) knockouts (KO) were synthesised and inserted into SEAP reporter vectors. Construction of a YY1 knockout was more complex as the YY1 consensus sequence is extremely degenerate and has a very short core sequence (CCAT) (Golebiowski *et al*., 2012). YY1 sites are accordingly challenging to identify *in silico*, particularly when stringent search parameters are employed to minimise false positives (as they were in the CMV bioinformatic survey) (Schug, 2008). It was therefore predicted that the CMV promoter may contain additional potential YY1 binding sites that were not identified by the initial *in silico* analysis. Utilising the online Regulatory Sequence Analysis (RSA) oligo-analysis tool (van Helden *et al*., 1998) it was determined that the core YY1 sequence CCAT (or the reverse orientation ATGG) occurs in 16 discrete occurrences and is the most abundant 4mer within the CMV promoter (Figure 5.6). Therefore as i) YY1 core motifs are apparently evolutionarily maintained in the CMV promoter, suggesting functional relevance and ii) empirical evidence indicates YY1 can potentially bind to any site containing the CCAT core sequence (Golebiowski *et al*., 2012), all 16 CCAT motifs were scrambled (to TGTC) to ensure deletion of every possible YY1-binding site. However, as shown in Figure 5.6, due to binding site overlap this potentially affected the functionality of multiple putative positive effector sites (*i.e.* scrambling YY1 elements disrupted multiple CRE and NFkB consensus sequences). Accordingly, it was hypothesised that the YY1-KO-CMV promoter would exhibit a reduction in the binding of both transcriptional repressors and activators.

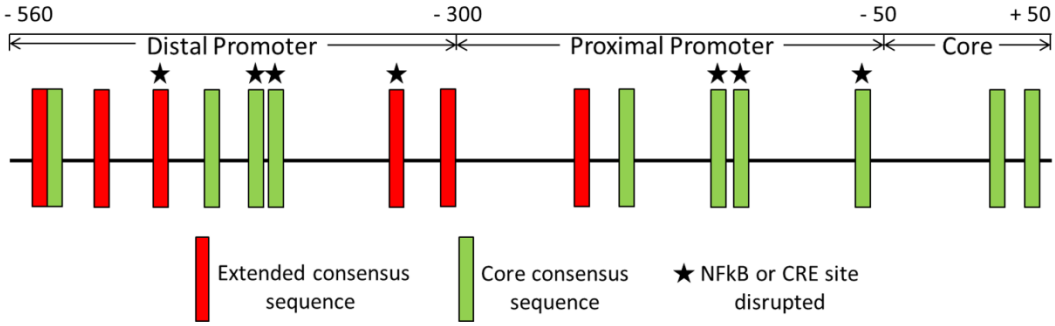


Figure 5.6: The CMV promoter contains 16 putative YY1 binding sites. The human cytomegalovirus immediate early 1 promoter was surveyed for the presence of extended ((C/G/A)(G/T)CCATN(T/A)(T/G/C)) and core (CCAT) YY1 consensus sequences. Analysis with the RSA oligo-analysis tool determined that the YY1 core consensus sequence is the most abundant 4mer within the CMV promoter, indicating evolutionary maintenance and functional relevance. A CMV-YY1-Knockout(KO) was constructed by scrambling all 16 potential YY1 binding sites (measurement of its activity in divergent CHO cell hosts is shown in Figure 5.7). Due to TFRE overlap, sequence scrambling at 7 discrete YY1 sites simultaneously disrupted NFkB or CRE sites within the CMV-YY1-KO construct (as indicated).

As it was hypothesised that CMV's mechanistic regulation may vary in different CHO cell lines due to varying TF complements, the activity of synthetic CMV promoters was evaluated in two commonly utilised hosts. Figure 5.7 shows the transient SEAP production from all synthetic CMV promoters compared to a control wild-type CMV (no TFREs scrambled) in CHO-S and CHO-K1. The data showed that relative promoter activities were approximately maintained in both CHO hosts. YY1-KO, NFkB-KO and CRE-KO promoters exhibited 38 - 42% (CHO-K1 – CHO-S respectively), 48 - 30% and 92 - 85% of wild-type promoter activity respectively. The analysis therefore identified that whilst deletion of NFkB elements significantly reduced CMV activity, removal of CRE sites had minimal effect on promoter strength. However, when both NFkB and CRE sites were deleted simultaneously (NFkB+CRE-KO) promoter activity was reduced to 15-8% compared to wild-type, suggesting that CMV-NFKB-KO mediated expression was dependent on CRE-mediated transcriptional activation. The data therefore indicates that CRE sites within CMV exhibit significant transactivation potential, but are largely functionally

redundant in the presence of NFkB sites. Further, the analysis verified the findings of the block-decoy study in showing that the vast majority of CMV promoter activity (> 80%) in CHO is a functional consequence of TF-TFRE interactions at NFkB and CRE regulatory elements. Accordingly, it was perhaps unsurprising that CMV-YY1-KO exhibited reduced activity compared to wild-type CMV (despite the finding that YY1 block-decoys increased CMV activity > 1.5 fold) given that multiple CRE and NFkB sites were disrupted in this construct (Figure 5.6). Whilst it therefore proved intractable to conclusively determine if YY1 sites are negative regulators of CMV activity, the TFRE deletion analysis definitively confirmed that CMV activity in CHO cells is predominantly dependent on just two discrete positive regulator TFREs; NFkB and CRE.

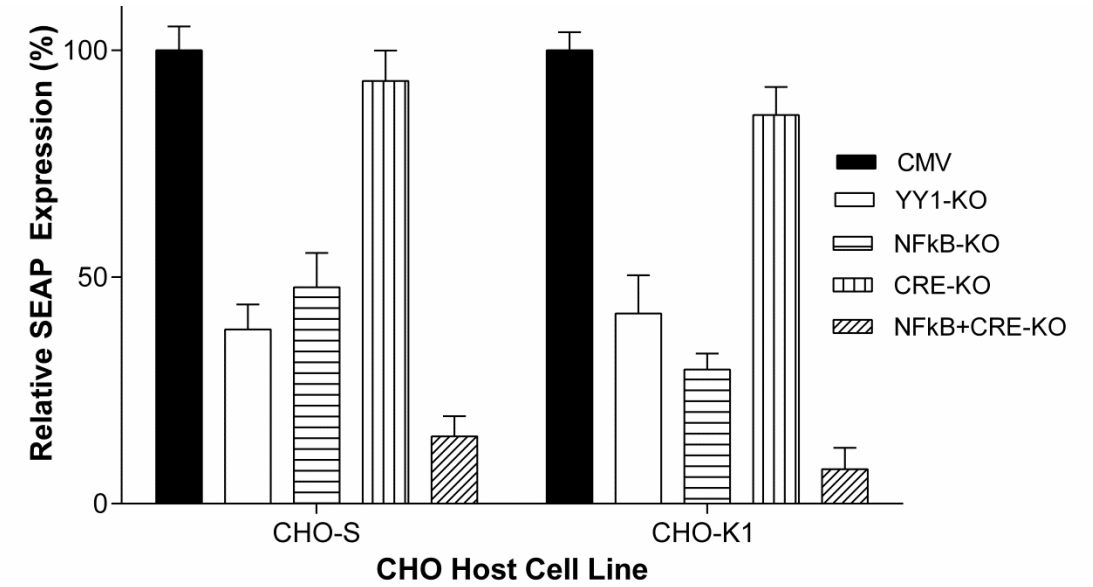


Figure 5.7: Deletion of NFkB and CRE sites abolishes the vast majority of CMV promoter activity in CHO cells. Synthetic CMV promoters with varying TFREs knocked out (*i.e.* scrambled) were synthesised and cloned into SEAP reporter vectors. The relative activity of each synthetic CMV construct was determined in CHO-S and CHO-K1 cells. Cells (2×10^5) were transfected with 300 ng SEAP-reporter vector, and SEAP production was quantified 24 h post-transfection. Data are expressed as a percentage of the activity of the wild-type CMV promoter in each cell line. Bars represent the mean + SD of three independent experiments each performed in triplicate.

5.3 Discussion

The work in this chapter shows that CMV promoter activity in CHO cell factories is predominantly mediated via CRE and NFkB TFREs. Physical prevention of TF-TFRE interactions at these sites, by either intracellular TF sequestration or TFRE deletion, independently reduced CMV activity by approximately 75% and 85% respectively. The involvement of these two elements in CMV regulation is unsurprising given that they i) are present at relatively high copy numbers in the promoter, ii) have cognate TFs available within CHO cells (Baycin-Hizal *et al*., 2012), iii) have previously been shown to be transcriptionally active in CHO cells (Brown *et al*., 2014) and iv) have commonly been identified as regulators of CMV activity in divergent cell types (Hunninghake *et al*., 1989; He and Weber, 2004; Lashmit *et al*., 2009; Liu *et al*., 2010a). However, it was unexpected that the remaining CMV-constituent TFREs were largely unable to positively regulate promoter activity, suggesting that the majority of the CMV sequence is functionally redundant in CHO. Despite CMV's highly complex TFRE composition (12 discrete TFREs, 42 binding sites) its mechanistic regulation in CHO cells is therefore apparently relatively simplistic.

Disruption of NFkB-mediated promoter transactivation via either TFRE-deletion or TF-sequestration reduced CMV activity by approximately 50%. Additional, simultaneous disruption of CRE-mediated transactivation (by either method) further reduced CMV activity to approximately 80% of wild-type levels. However, in isolation, physical sequestration of CRE-binding TFs significantly reduced CMV activity whilst CRE site deletion had minimal effect. These data suggest that CRE-binding TFs can regulate CMV activity both directly (*i.e.* binding at CRE sites within the promoter) and indirectly (for example by either interacting with, or regulating the expression of, other TFs). With respect to the latter, it was recently shown that CRE-binding protein 1 regulated CMV activity in 293T cells without directly binding to the promoter (Chia *et al*., 2014). The analyses therefore indicated that whilst NFkB sites could almost completely compensate for CRE site deletion, CRE sites could only partially compensate for NFkB site deletion. Accordingly, CMV promoter function in CHO cells can largely be explained by the following simple design rules: 1) activity is primarily dependent on NFkb sites

(mirroring previous work in this thesis that showed CHO-specific synthetic promoter activities were primarily a function of NFkB site copy numbers), 2) when NFkB-mediated transactivation is disrupted (as may occur under certain cellular conditions) activity is predominantly dependent on CRE sites, and 3) if NFkB and CRE-mediated transactivation are simultaneously disrupted the vast majority of promoter activity is abolished.

A mechanistic understanding of the CMV promoter's functional regulation in CHO cell factories will enable strategies to precisely control or improve its transcriptional activity. For example, it has been shown in multiple cell models that the CMV promoter can be 'de-silenced' by increasing the intracellular abundances of active positive effector TFs (Keller *et al*., 2007; Stinski and Isomura, 2008; Liu *et al*., 2013). Indeed, in N-Tera2 cells CMV activity can be immediately de-silenced by phorbol ester stimulation via a mechanism that is entirely dependent on TF-TFRE interactions at NFkB and CRE sites (Liu *et al*., 2010a). Accordingly, production cell lines suffering productivity loss as a result of CMV silencing could potentially be 'rescued' by either over-expressing, or activating, NFkB and CRE-binding TFs. Moreover, utilisation of the same strategies in the development of future producer cell lines may both increase production and help to prevent the occurrence of silencing. Overexpression of NFkB and CRE-binding TFs and/or the development of production processes to specifically control/optimise their intracellular abundances could also enable significant increases in CMV-driven transient protein production. Given that CMV is currently the promoter of choice for CHO TGE systems this could both optimise the production of early stage products (*i.e.* development material for toxicology and clinical trials testing (Daramola *et al*., 2013)) and help facilitate exploitation of TGE as a method for large-scale biopharmaceutical manufacturing (De Jesus and Wurm, 2011).

Disrupting YY1-mediated regulation of CMV activity may provide additional opportunities to improve promoter performance. Physical sequestration of YY1 with block-decoys increased CMV activity to approximately 150% of wild-type levels, indicating that YY1 is a negative regulator of CMV in CHO. YY1 is a well-known repressor of CMV activity in other cell types and can repress transcription by competing with activators for promoter binding sites and/or recruiting co-repressor complexes (Liu *et al*., 1994; Pizzorno, 2001, Liu *et al*., 2010b). With respect to the former, many of the putative YY1 sites in CMV overlap

with, or are in close proximity to, NFkB and CRE elements. With respect to the latter, YY1 can recruit histone deacetyltransferases and methyltransferases to promoters (Zhang *et al*., 2011), and therefore may play a key role in mediating CMV silencing in stable CHO cell lines (Liu *et al.*, 2013). However, in contrast to block-decoy-mediated disruption of YY1 regulation, deletion of all putative YY1 binding sites reduced CMV activity by approximately 60%. This is likely explained by the fact that deletion of YY1 sites simultaneously disrupted multiple NFkB and CRE elements. Construction of CHO-specific synthetic 'CMV-like' promoters devoid of YY1 binding sites and containing fully optimised CRE and NFkB compositions (*i.e.* optimised copy numbers, orientation, spacing, location and affinities) (Sharon *et al*., 2012) may provide significantly increased activities. Such promoters could potentially be used to maximise recombinant gene transcription levels in transient protein production. Alternatively, TGE yields may be optimised by utilising the wild-type CMV promoter in conjunction with YY1-specific block-decoys.

Chapter 6: Conclusions and future work

6.1 Conclusions

The tools developed in this thesis enable sophisticated, next-generation transcriptional control in CHO cell factories. CHO cell engineers can now for the first time i) precisely control recombinant gene expression over broad dynamic ranges, ii) rapidly determine/ control the mechanistic regulation of any multi-transcription factor mediated cell function (e.g. regulation of a specific promoter) or phenotype (e.g. regulation of cellular proliferation), and iii) implement strategies to predictably control and improve the activity of the most commonly utilised genetic component, the hCMV-IE1 promoter. Accordingly, functionally ill-defined and uncontrollable genetic elements exhibiting suboptimal performance can now be replaced with bespoke control systems offering predictable, precise and optimised transcriptional activity. The novel suite of tools presented in this thesis can therefore facilitate development of next-generation biopharmaceutical manufacturing systems via the following specific applications:

1. **CHO cell factory engineering.** The provision of 140 discrete promoter activities, covering over two orders of magnitude, will enable CHO cell engineers to precisely vary the transcriptional activity of many functional genes simultaneously in order to obtain desired phenotypes. Critically, by utilising block-decoys to 'stress-test' synthetic promoters (*i.e.* determine how discrete promoters' activities are affected by fluctuations in intracellular TF abundances) elements can be selected that do not exhibit promoter-promoter interference, allowing construction of multi-gene engineering systems with robust, reliable and predictable performances. The two technologies can therefore be utilised in tandem to enable creation of bespoke, synthetic mammalian cell factories harbouring multiple genetic components operating at an optimal, designed stoichiometry. By facilitating optimisation of key bioproduction functionalities such as protein folding and glycosylation this will potentially enable both (i) re-engineering of existing stable cell factories to render failed products manufacturable and (ii) forward engineering of new cell factories with predictable manufacturing properties. This will eliminate the current concept

that some products are ‘difficult-to-express’ (DTE), and replace it with the concept that ‘designer products’ are easy to express in ‘designer cell factories’.

2. Product-gene transcription rates. The precise control of recombinant gene expression enabled by the synthetic promoter libraries can be utilised to i) maximise transcription levels of easy-to-express proteins, ii) optimise transcription levels of DTE proteins, such that activity is kinetically coordinated with polypeptide-specific folding and assembly rates, and iii) achieve optimal mAb-specific light chain: heavy chain expression ratios (block-decoys could again be employed to select promoter pairs that do not exhibit promoter-promoter interference). In the long-term, this will eliminate the productivity-limiting one-promoter-fits-all approach and replace it with the concept that each specific product requires a product-specific promoter (*i.e.* protein-promoter matchmaking) in order to both optimise factory performance and maximise productivity. In the short-term, whilst synthetic promoter technology is being further validated, the identified hCMV-IE1 control strategies will facilitate optimisation of the incumbent manufacturing component. Over-expression/activation of NFkB and CREB and/or silencing/inactivation of YY1 could both i) ‘rescue’ silenced stable cell factories, ii) prevent silencing and enhance productivity in new cell factories and iii) significantly increase transient protein production yields.

6.2 Future work

Future work is required to both further validate and improve the tools developed in this thesis. Whilst many future studies may utilise block-decoys (e.g. to determine the mechanistic regulation of discrete promoters or cell phenotypes) direct follow-on work will predominantly focus on hCMV-IE1-control strategies and synthetic promoter development.

6.2.1 Strategies to improve hCMV-IE1 function – reduction to practice

Whilst multiple potential strategies to improve hCMV-IE1 functionality have been proposed, follow-on work should focus on testing the following two approaches as they offer the greatest potential industrial benefits in both the short and longer term:

1. Re-activation of silenced stable cell lines. It is hypothesised that transiently increasing the intracellular abundance of NFkB and CREB and/or decreasing the abundance of YY1 will enable hCMV-IE1 'de-silencing'. This hypothesis could be interrogated in stable producing cell lines that have suffered productivity loss as a result of hCMV-IE1 silencing. These cell lines could be transiently transfected with vectors encoding discrete combinations of the NFkB gene, CREB1 gene and a YY1-targeting short interfering RNA (siRNA). Subsequent measurement of hCMV-IE1 activity (*i.e.* protein production) over short-term batch culture and during bioreactor-scale production processes would determine the efficacy of this approach to 'reactivate' silenced cell lines. This work could potentially identify a simple method to routinely 'de-silence' stable cell lines that utilise the hCMV-IE1 promoter, with minimal effect on cell factory performance (e.g. cell growth and viability), by transient transfection of an optimised 'CMV-reactivation vector'.

2. Optimisation of transient gene expression (TGE) systems. The work in chapter 5 showed that transient hCMV-IE1-driven protein production was increased by disruption of YY1-mediated regulation. The potential to increase TGE yields by inhibiting YY-1 activity could be determined in longer-term larger-scale transient production processes utilising either YY1-specific block-decoys or YY1-targerting siRNAs. Utilisation of reporter-vectors harbouring different promoter elements (*i.e.* testing YY1-inactivation in combination with a range of discrete promoters) would determine whether YY1-knockdown specifically increases hCMV-IE1 activity or generically improves CHO cell functionality. Measurement of key factory performance indicators (e.g. cell viability), cellular productivity and protein yields would accordingly confirm whether TGE systems can be simply optimised by YY1-inactivation.

6.2.2 Synthetic promoter library development – reduction to practice

Given the availability of expression stability enhancing strategies (such as recombination mediated cassette exchange and matrix attachment regions (MARs)) it is anticipated that synthetic promoters will maintain their relative activities in stable expression systems. However, clearly, it will be imperative to rigorously test this hypothesis. This could be achieved by cloning a panel of functionally diverse

(with respect to expression range) synthetic promoters into reporter-protein encoding vectors containing the genetic elements necessary for producing stable CHO cell line transfectants (e.g. glutamine synthetase selection marker, MARs, introns). Each synthetic promoter-reporter vector could then be utilised to create triplicate stable transfectant pools, prior to monitoring of reporter production and cell proliferation over long-term (> 100 generations) batch culture. Evidence of robust and predictable performance in stable CHO cell factories would be a significant step towards integrating synthetic promoter technology into biomanufacturing systems.

Whilst TFRE-reporter vectors provided a robust measurement of TFRE activities they were limited to assaying a relatively narrow range of TFREs (< 35), restricting the subsequent synthetic promoter design space. Utilisation of a high-throughput assay system, such as the recently described massively parallel reporter assay (MPRA) (Melnikov *et al*., 2012; Kheradpour *et al*., 2013), would enable functional screening of all known TFREs (> 200). In addition to the screen used in this study (short-term transient expression in multiple CHO hosts) the MPRA could be employed to determine the relative activity of each TFRE in bioprocess-relevant production conditions through a fed-batch culture, utilising stable pools. TFREs that are active in all CHO cell types, and maintain high activity during extended stationary phase culture could then be selected to construct next-generation synthetic promoters specifically designed to operate within specific process parameters, enabling optimal control of cell specific production rate during bioreactor operations. Moreover, identification of production phase specific TFRE functionalities would enable promoter design that permits biphasic cell biomass accumulation and product synthesis.

Finally, it will be necessary to exemplify that next-generation CHO cell factories can be created by employing synthetic promoters to improve factory performance via controlled and predictable multigene engineering. This could be achieved by using a combination of design-of-experiments response surface modelling and multigene expression to construct engineered synthetic cell factories capable of significantly improved DTE protein synthesis. For example, DTE product genes (such as tissue plasminogen activator and erythropoietin) could be co-expressed with discrete combinations of functional genes that have previously been utilised singly for CHO cell engineering (e.g. chaperones, redox proteins, unfolded protein response transactivators, vesicle trafficking components, *etc.*). Statistical

modelling could then be employed to determine optimal functional gene stoichiometries that could subsequently be achieved in novel synthetic factories by utilising appropriate synthetic promoters (*i.e.* with appropriate relative activities) to construct multigene engineering vectors. Creation of stable cell clones expressing the DTE protein, and monitoring of their growth and productivity, would then determine whether these bespoke CHO cells were able to increase productivity via protein-specific design solutions.

Appendix A: Synthetic promoter TFRE-block compositions

Supplementary Table 1: Sequence and relative activity of first generation synthetic promoters. N = NFκB-RE, E = E-box, G = GC-box, B = C/EBPα-RE, C = CRE, F = E4F1-RE. Reverse orientation (3' to 5' with respect to the SEAP reporter) is indicated by an apostrophe.

Promoter Sequence	Relative Activity (%)
1/01 N E E N G B C E' B' N N E G' N	100
1/02 E' F F' B' E G' N' E' B N N' N N' G N	84.46
1/03 E E G' C N' G' G G' C' E N N' B' E' B E'	67.20
1/04 B' C' E' G C B' N N' N' G F' B N'	58.3
1/05 G F N G N B' N' B C E' N'	56.87
1/06 C E E' G' E N E C B' G' C' N' G'	54.89
1/07 B' E B' G' N B B E C' N N	54.71
1/08 E E' N' C B' E' E G B E N	52.64
1/09 B G' N B' C' N' E E' C' G E'	51.31
1/10 B' E N' E E' B C' G G N	48.83
1/11 C' E' N B' E G' F'	47.64
1/12 B' N N E E G' C E N' B' B	47.18
1/13 G B E' N' G E N C' B' N E' E' C E'	44.7
1/14 C G G N C B B' N' N C' F E' N	38.43
1/15 N' B F' C' N' B E' N C' N'	35.94
1/16 C F' N N B' C' E E' F'	34.31
1/17 G' G G B' E N B N' B' C B' E G' G	34.27
1/18 N N' E B' C G' E B' C'	34.03
1/19 E E C E C N N' F C N B N' E'	33.77
1/20 F' G E E E' N' F E B G' E' E	33.35
1/21 N E N' E' N N N'	32.57
1/22 F G' B' E N' E' B' E' G G B	31.73
1/23 C' N E' B C B N G' E	30.32
1/24 C N' B B E' F N E' G N' C	28.25
1/25 G G' B' G' B' N' C N N B' E' C C E' B'	28.13
1/26 N G' B' N E E E' F' N B' N' B	27.33
1/27 B' C E F' B' E F' N G N N	26.87
1/28 E' F N' N E E C N E' F' N G' N G E' F N G' N E' F'	25.72

1/29 N N N N N N	25.69
1/30 C B’ N E’ E’ G N’	24.68
1/31 B’ G B E’ F’ N’ C B N’ E’ B G	23.33
1/32 E’ F G’ N G E E N F N C G B’ G G B’	20.09
1/33 C N’ F N G’ G’ G’ N’ F F C’ N E’ N’	19.46
1/34 G G’ N’ E E F C E B’ E’ C’ N’ F G’	15.22
1/35 C B C’ E B G’ N’ B’ E’ G N’ G G	15.16
1/36 E C’ N G F’ F G F G’ N E B C’	14.48
1/37 N F’ E E F F’ B N B N F G F’ C C’ G N’ E C N’	14.44
1/38 N’ C N C C E’ G C B’ N N	13.75
1/39 C’ E’ C’ E’ F B G’ G’ N’ G E’ E’ C C	13.59
1/40 E’ F G C E F F’ G’ G F’ G F G E G B’ N B’ E N F B G’	13.57
1/41 B’ B’ E G’ F’ G’ B’ F’ E N’ B C	12.38
1/42 F’ B’ G’ N B B F’ E’ E B C’ N B E’ N G’ B’ C N’ G	12.23
1/43 G E’ F N C B C’ E E	12.1
1/44 G’ E E F C’ C’ B’ N E’	12.04
1/45 F’ B N’ C’ B C’ G N’ E’ G E C	12.01
1/46 N G E’ C’ G’ E B E’ E C F C’ E’ E F E G C	11.56
1/47 E E E E E E E	10.78
1/48 E’ G’ F G’ E’ C F G’ E’ G B’ E’ B B C	9.11
1/49 B’ E’ F G’ N F C G G’ B’ N’ B’ E C	8.94
1/50 F G’ E’ E’ C’ B’	8.74
1/51 F C’ G’ N F’ C G N G C’ B E G’	8.53
1/52 C N’ C’ N C’ G’ N’ E’ N’ F C C N’ B’ C’ C N C’ N’ F G	8.53
1/53 G G’ F’ N B’ N’ F’ E’ B’ N	8.44
1/54 C C C C C C C	8.43
1/55 G’ C C’ C N’ E N	8.03
1/56 G’ B’ C’ N E’ G E F’ F	7.38
1/57 N’ B’ B G’ F G’ F N E	6.89
1/58 N G C’ C	7.74
1/59 C’ G’ G E’ E’ G’ B N B N F’ F E F’	6.06
1/60 G G G G G G G	6.03
1/61 G’ B’ C C N’ B’ E E B’	5.94
1/62 F F F F F F	4.7
1/63 E’ E G G C’ N’ G B’ C G F’ F’ B E E B F’	4.62

1/64 E B B C’ F’ C B E’ G’	4.29
1/65 F’ B G’ B’ C E E’ F’ B	4.04
1/66 E’ E C G N C’ F’ E C G’ E F’ N C	4
1/67 E’ G’ G’ N F E’ F’ N N’ G F F’	3.94
1/68 G’ G’ N’ F’ N’ C B’ E’ C’ E B F E’ F N’ G C’ G C’ N G’ C B C F N’ B E F B’ G	3.57
1/69 E G B C F’ B B G N C	3.43
1/70 E’ N’ G F N E’ G’ C F E’ C’ G G C B’	3.35
1/71 E’ N’ F G’ C F E B C’ C N	3.24
1/72 F F’ C G’ C’ B’ C’ N’ C G B B F’ C N’ F E’ E B’ F’ G’ G B’	3.14
1/73 B B B B B B B	3.13
1/74 C’ G’ C’ B’ B E B’ C’ B B B	3.1
1/75 G G N’ N’ F G F C’ B G’	2.57
1/76 B F G G G G F B C’ E F	2.52
1/77 E F E’ N F G’ N F N G’ F C’ F G’ G	2.46
1/78 F’ G G C’ G G’ F N’ B’	2.39
1/79 F F F’ E’ F’ C’ F B C’ N B E B’	2.22
1/80 E’ G C E F’ F’ B’ G E’ G	2.11
1/81 B’ G C B’ G F G F’	2.02
1/82 B’ E’ B’ B G’ B F’ C’ B’ N C’ C’ G’	2.01
1/83 F’ C’ F’ N B G’ N G’	2
1/84 B’ C’ B’ F’ G’ G E’ B C G C	1.94
1/85 C’ E E’ E B G F G F’ F F’ E G	1.87
1/86 G’ E’ N F E B’ F C’ E’ F’ C C’	1.71
1/87 F’ C B’ G F G G E F’ C’ E	1.57
1/88 C’ N’ G E’ C’ G C N’ G’ F F	1.35
1/89 C’ G G B’ C C B’	1.34
1/90 F C’ G’ B’ F E’ F’ G’ F G C	1.33
1/91 E’ B’ C E F’ F’ F B B G’ B’ C C B C’ G	1.33
1/92 C’ E’ F C F’ F’ E F’ E B	1.26
1/93 B F’ G E G’ G N F C F	1.23
1/94 B’ E G N’ N F C G F N F C’ F	0.66
1/95 G’ G’ N’ F’ F’ E’ C’ B’ C C’	0.38
1/96 E B E’ B’ E E B’ F G C’ F F B	0.36

Supplementary Table 2: Sequence and relative activity of second generation synthetic promoters. N = NFκB-RE, E = E-box, G = GC-box, B = C/EBPα-RE. Reverse orientation (3' to 5' with respect to the SEAP reporter) is indicated by an apostrophe.

Promoter Sequence	**Relative Activity (as a % of hCMV-IE1 activity)**
2/01 N G E E' N E' E' N N N E' E N E'	216.66
2/02 E N' N' N' E E E' N' E' B' N' E' N N E'	175.94
2/03 E E N' G' E N' E' N' N' E' N' N' N B	174.82
2/04 N E N E N N' N' E N	169.04
2/05 N' N G N B' E' E' N B N E' N N'	166.6
2/06 N' N' N N E' N G' E' N' E B'	157.92
2/07 N' N' E' N G' N' N' N' B N	155.3
2/08 E' E' N' B' N' N' G' N N N' E	154.47
2/09 N' N' E N B N N' E' N G E' E	151.37
2/10 E' N B' N' N E' N B N N N' G' N N N'	150.63
2/11 E N' E N N' N' N' N E' N' N' N' N' B N' E' N E' N E'	150.17
2/12 N' E' N N B' N' N N E' B E	148.93
2/13 E B B' E B' N N E' N N' E E G' N G N E B N	143.62
2/14 E N' N' N G N N' G' B N' G N	140.87
2/15 N G' E' N' N' N N' E' N	140.87
2/16 N G' E N N' B' N N G' G	140.77
2/17 E' N G' G' E' E N N N E' N N' G' N' B' N E E' N N N	138.95
2/18 N' N' N N' G' N' N G' E' N B E	138.83
2/19 E' N' N B' N N E' G N N E B' N E	137.6
2/20 E E N' E' E' E N' G' E' N E' N G'	137.53
2/21 N N E' E N' N G E' N' E'	132.61
2/22 N E' G' E N' B E' N' E N	117.77
2/23 N' E' N' N G N' N N G E'	117.22
2/24 B' E N' N' E N' G' N E' G N B' N	113.38
2/25 N G' N N N N' N' B N' N	108.35
2/26 B E' G' N G' N' E G E' N' G'	99.36
2/27 N' N E' N' N' E E B' E G' E'	98.83
2/28 E N' N' N N' N N' B	97.9
2/29 B' N' G' G' E' N' E' N E N' B N	94.66
2/30 E N N' N E N B N' N'	94.36

2/31 E’ G’ G’ N’ G’ E B N’ E N’ N	90.93
2/32 N N G’ E E’ N N’ G’ E E	90.49
2/33 N’ E G N N’ E E’ B’ N’	87.38
2/34 B E G’ N’ N N G E	82.11
2/35 N’ N’ G B G’ B’ N N’ N N’ B’	76.61
2/36 N’ E N’ G N N B’ E G	75.11
2/37 G’ N N N’ N E E G’ G’	70.59
2/38 G E B N N E’ E’ N’	70.24
2/39 N N G E’ B E’ N	63.19
2/40 G’ N’ N E’ G N’ N’ B G	62.21
2/41 E B’ E’ E B’ N’ N’ E’ B’	57.26
2/42 N’ G N’ N E G B’ B’	41.65
2/43 G E’ B’ N G’ N G B’ N	36.61
2/44 B E E’ E B E’ B’ B E’ N	32.3

Appendix B: Synthetic promoter library analysis report

A synthetic promoter library analysis R script (synpro.anal.R) was developed that automatically generates a word document analysis report from a CSV file detailing promoter activities and sequence compositions. The data provided within the report enables rapid identification of optimal design spaces for next generation promoter library construction. This appendix shows an example of the full analysis report generated from synthetic promoter library 1 data.

Synpro.anal.R can be downloaded at the following site: https://sourceforge.net/projects/syntheticpromoteranalysis/files/

Synthetic Promoter Library Analysis Report

This analysis report was automatically generated, by executing the R script synpro.anal.R, from a CSV file detailing synthetic promoter activities and TFRE-block composition. The data within this report enables simple, rapid identification of an optimal design space for next-generation promoter library construction. **Input/ response is required where text is underlined and bold.**

Reporter protein used:

Cell line:

Date:

The analysis first converts the TFRE blocks into letter identifiers (TRREs sorted alphabetically). Within this analysis letter identifiers correspond to the following TFREs:

A:

B:

C:

D:

E:

F:

Section 1: Key Library Statistics

The first stage of the analysis checks key library statistics and evaluates three basic assumptions. Output 1 provides a summary of the dataframe statistics, showing the distribution of each TFRE block across the promoter library.

```
       A                B                C                D                E                F
 Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
 1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:0.000   1st Qu.:1.000   1st Qu.:1.000
 Median :2.000   Median :2.000   Median :2.000   Median :1.000   Median :2.000   Median :2.000
 Mean   :1.897   Mean   :1.814   Mean   :2.237   Mean   :1.629   Mean   :1.969   Mean   :2.072
 3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:3.000
 Max.   :7.000   Max.   :7.000   Max.   :7.000   Max.   :6.000   Max.   :7.000   Max.   :7.000
   Expression
 Min.   :  0.00
 1st Qu.:  3.13
 Median :  8.94
 Mean   : 17.88
 3rd Qu.: 28.13
 Max.   :100.00
```

Output 1: Relative abundance at which discrete TFRE blocks occurred per promoter.

Assumption 1: The relative abundance of TFRE blocks across the promoter library is as designed = TRUE/FALSE

The range of promoter activities across the library is shown in Figure 1. If the range of activities is undesirable then the library may need to be re-constructed with different design criteria.

Assumption 2: Promoter activities cover the desired/ expected expression range. = TRUE/FALSE

Figure 2 shows the correlation between total number of sites per promoter and relative transcriptional activity. Promoter length is not expected to be a key determinant of promoter activities. If this assumption is false then library 2 will need to be tailored accordingly.

Assumption 3: Relative promoter activity is not simply a function of promoter length. = TRUE/FALSE

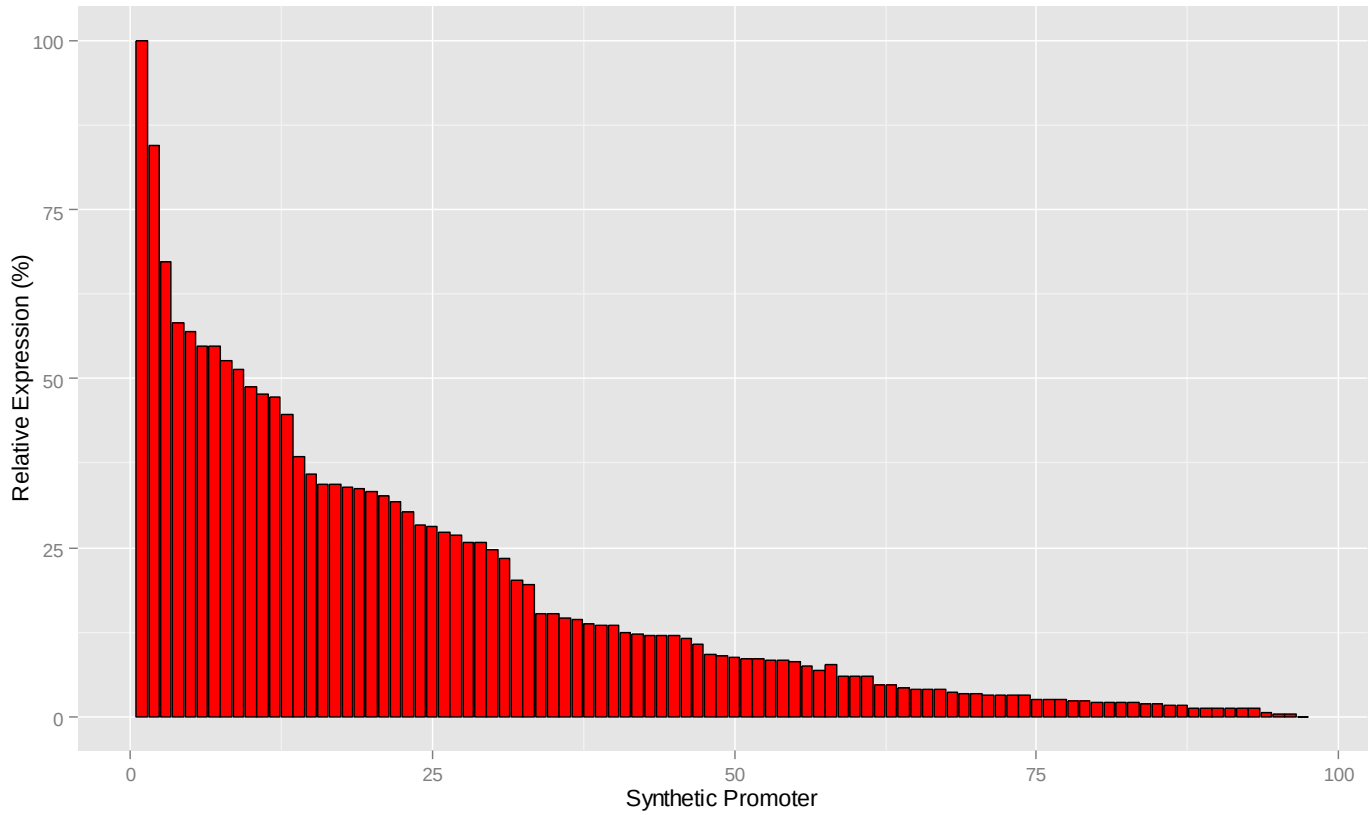


Figure 1: Range of promoter activities across the library

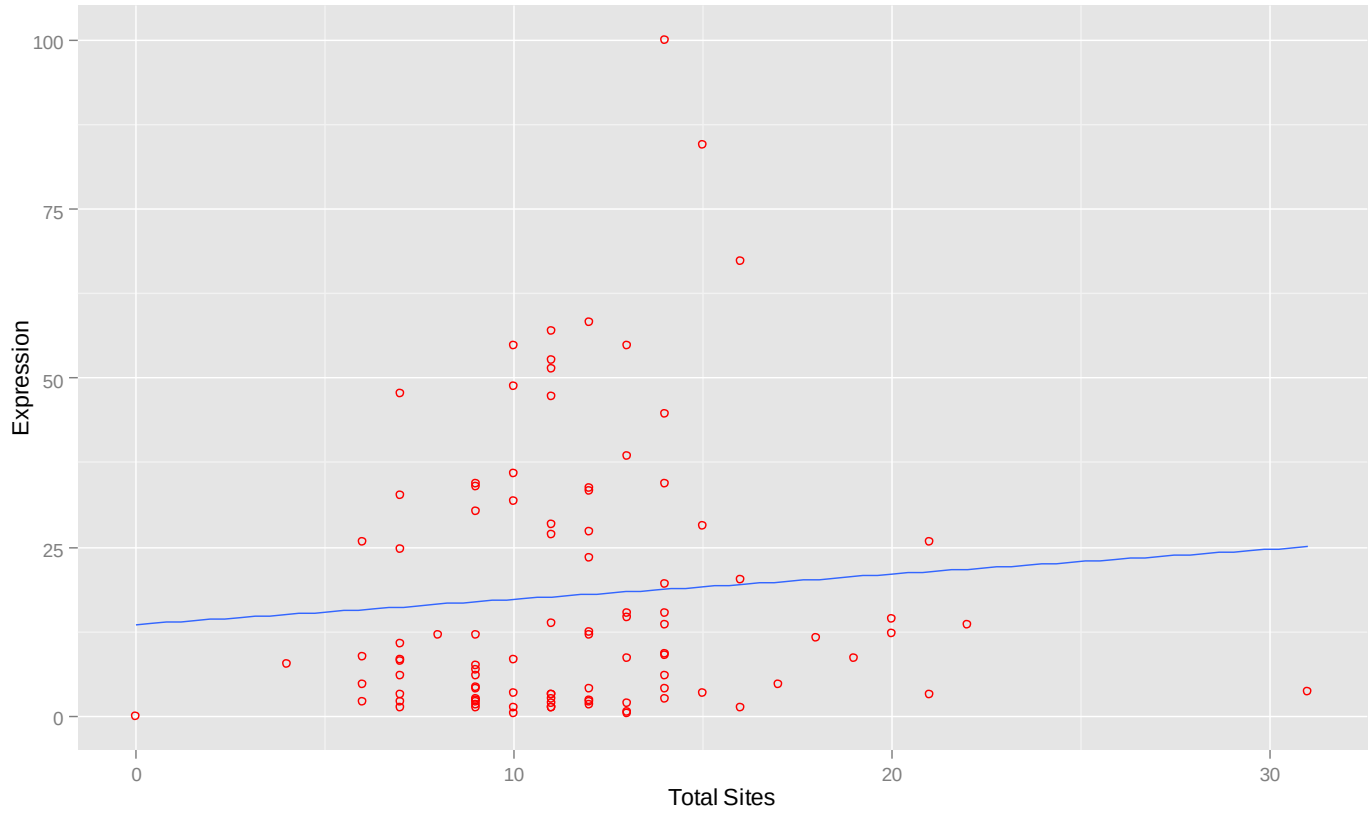


Figure 2: Correlation between total TFRE sites and promoter strength

Section 2: Identification of TFREs that are positive, neutral and negative regulators of promoter activity.

The second analysis section provides the data required to evaluate the contribution discrete TFREs make to promoter activities. Three figures are generated for each TFRE that enables blocks to be designated as positive, negative or neutral effectors of promoter strength:

A) The number of the TFRE block in each synthetic promoter is plotted against relative activity of that promoter. The linear regression line is shown, where the slope of the line indicates the extent to which the TFRE occurs in promoters of varying activity.

B) The mean number of the TFRE block in higher or lower activity promoters (over or under mean promoter activity).

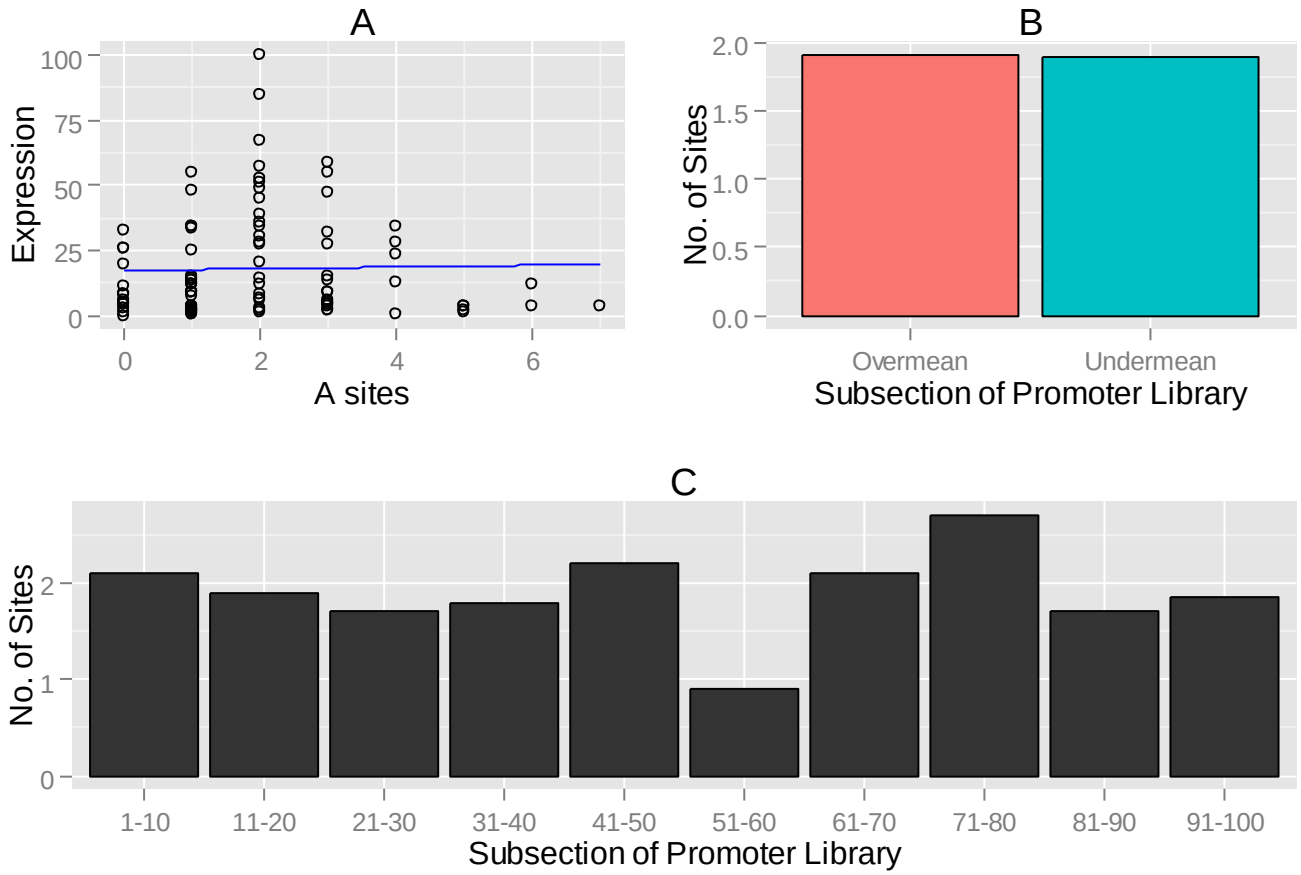
C) The mean number of the TFRE block in promoters within discrete library subsections. Subsection 1-10 contains the top 10% of promoters (ranked by activity).

Determination of each TFRE block's functionality facilitates binary (yes/ no) decisions regarding formation of next-generation design spaces. **Comments regarding each sites regulatory function can be made in Table 1.**

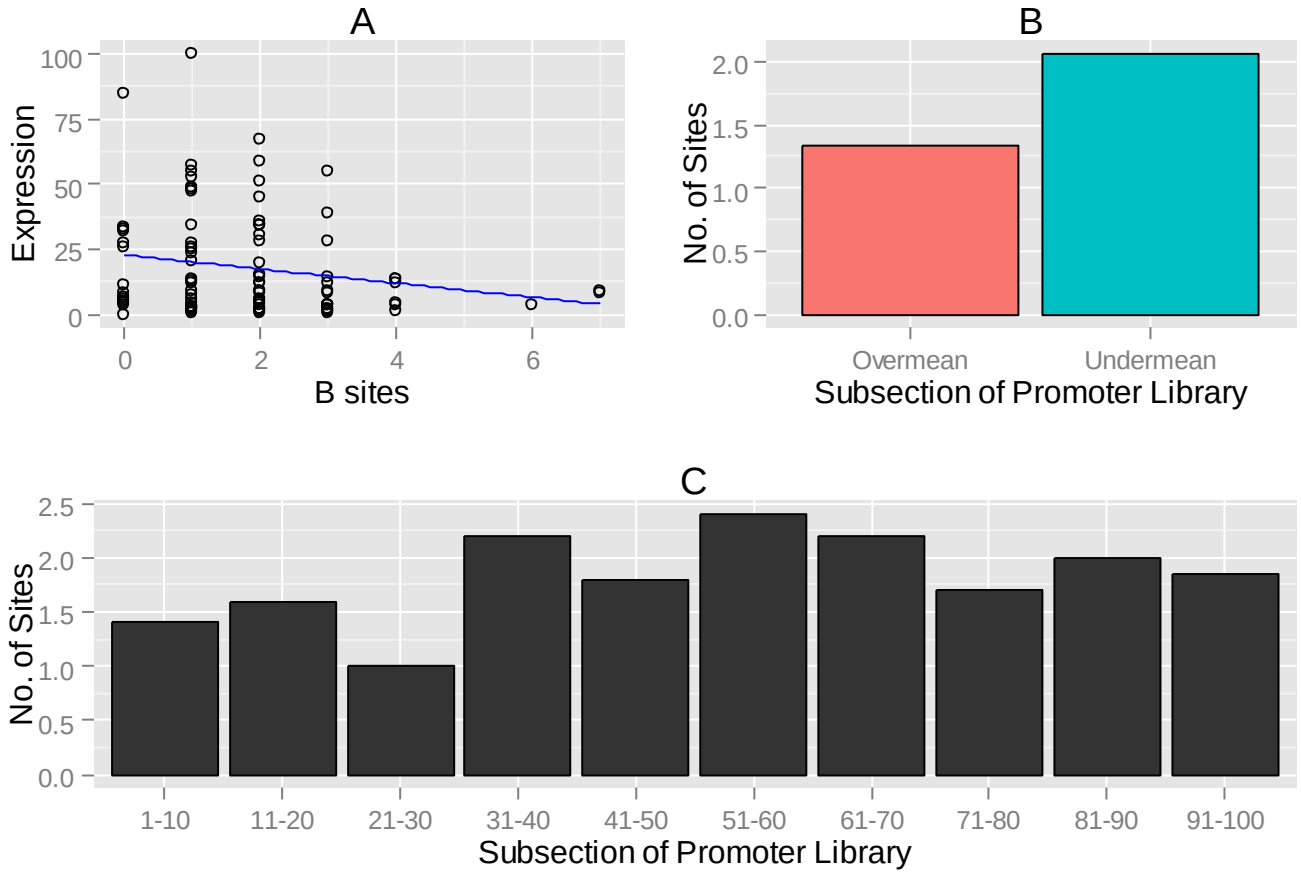
TFRE	Designation	Comments
A	Positive/Negative/Neutral	Comments about each site
B	x	x
C	x	x
D	x	x
E	x	x
F	x	x

Table 1 Characteristics of each individual site

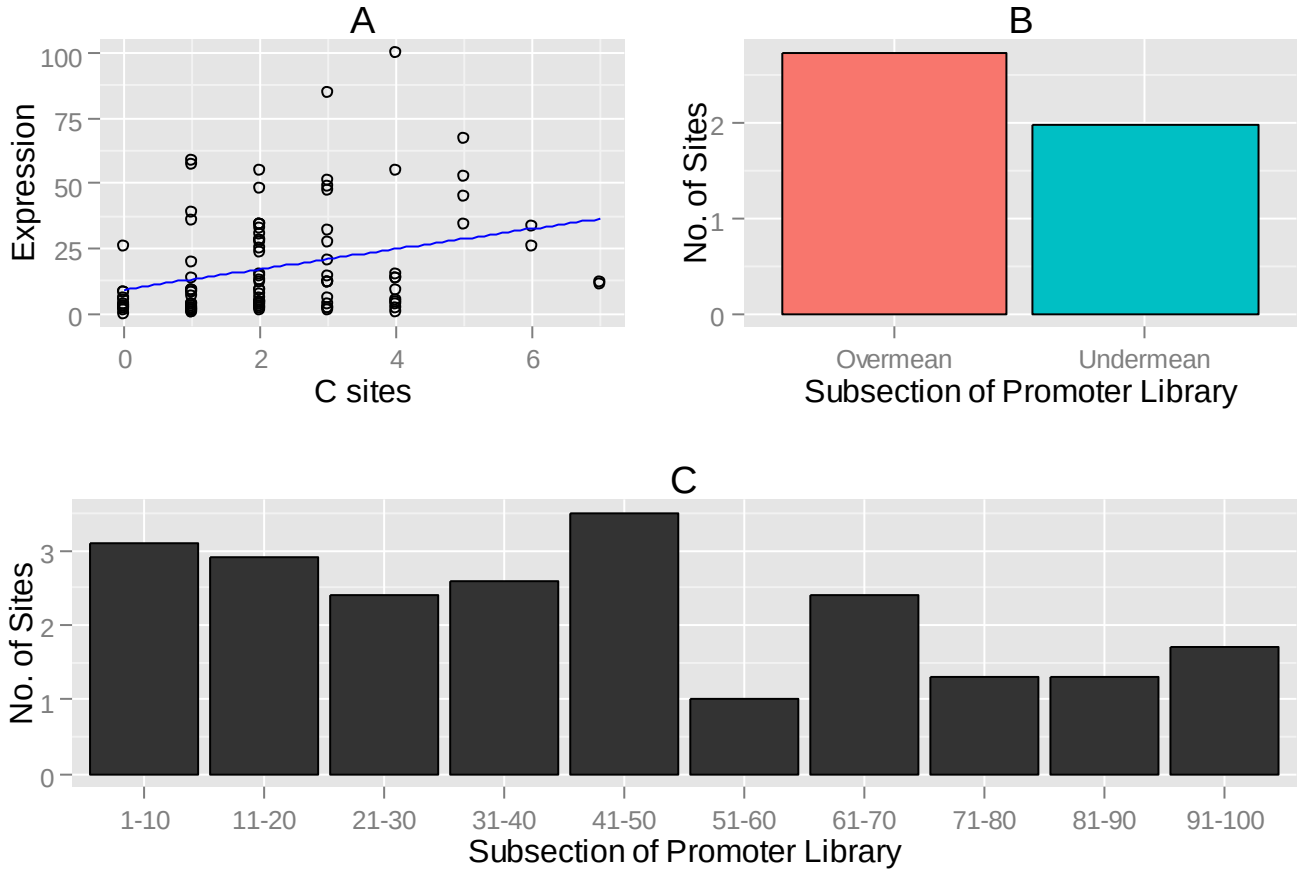
Statistics for site A



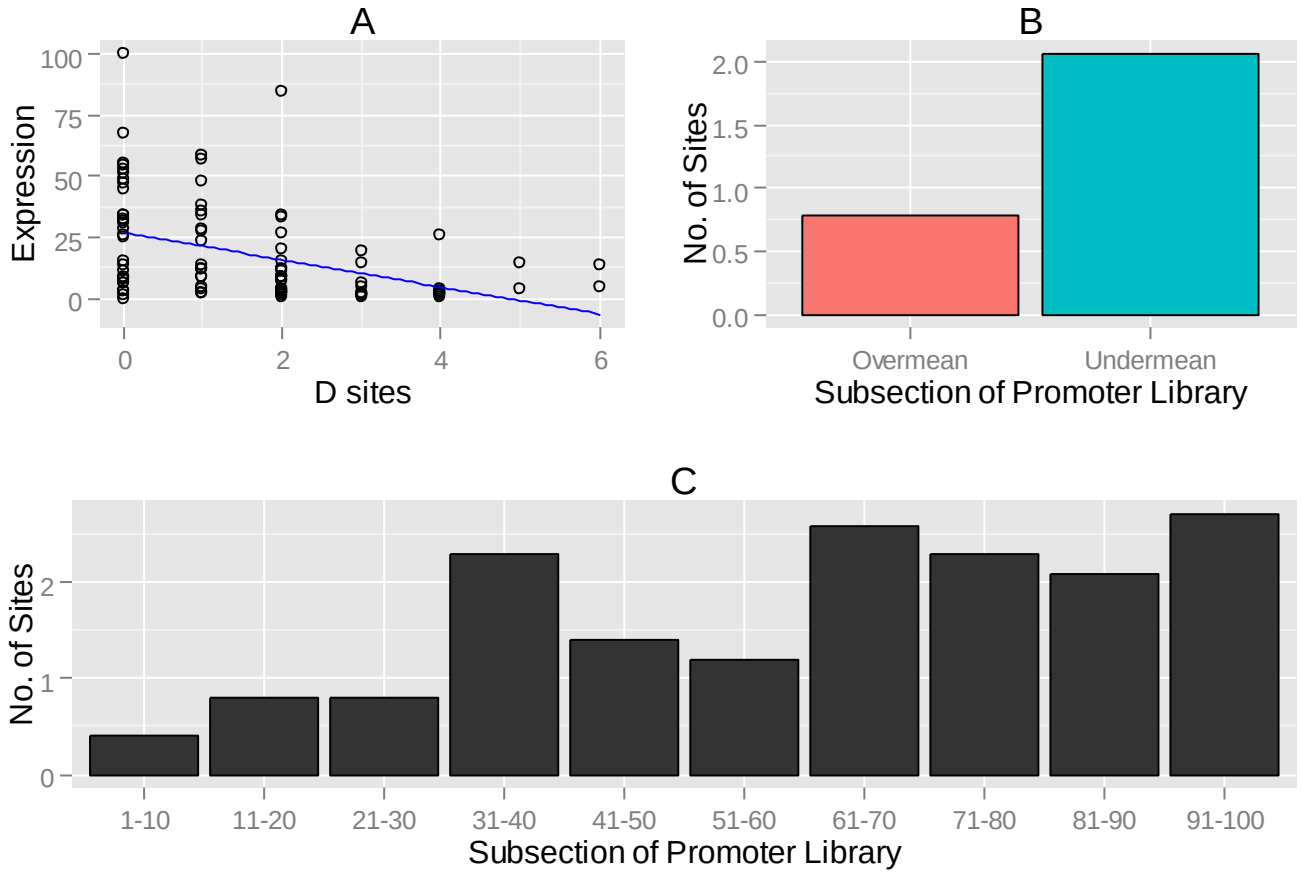
Statistics for site B



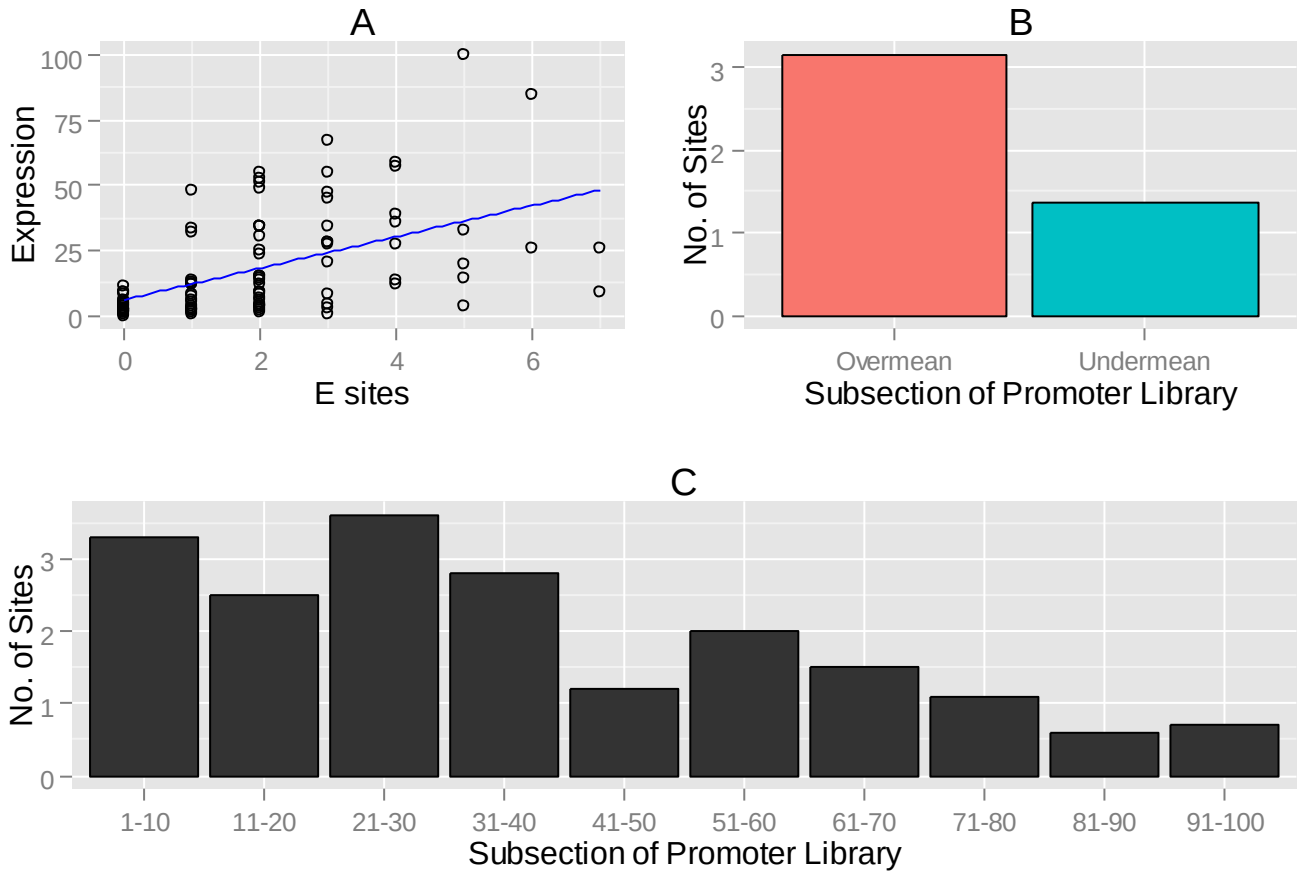
Statistics for site C



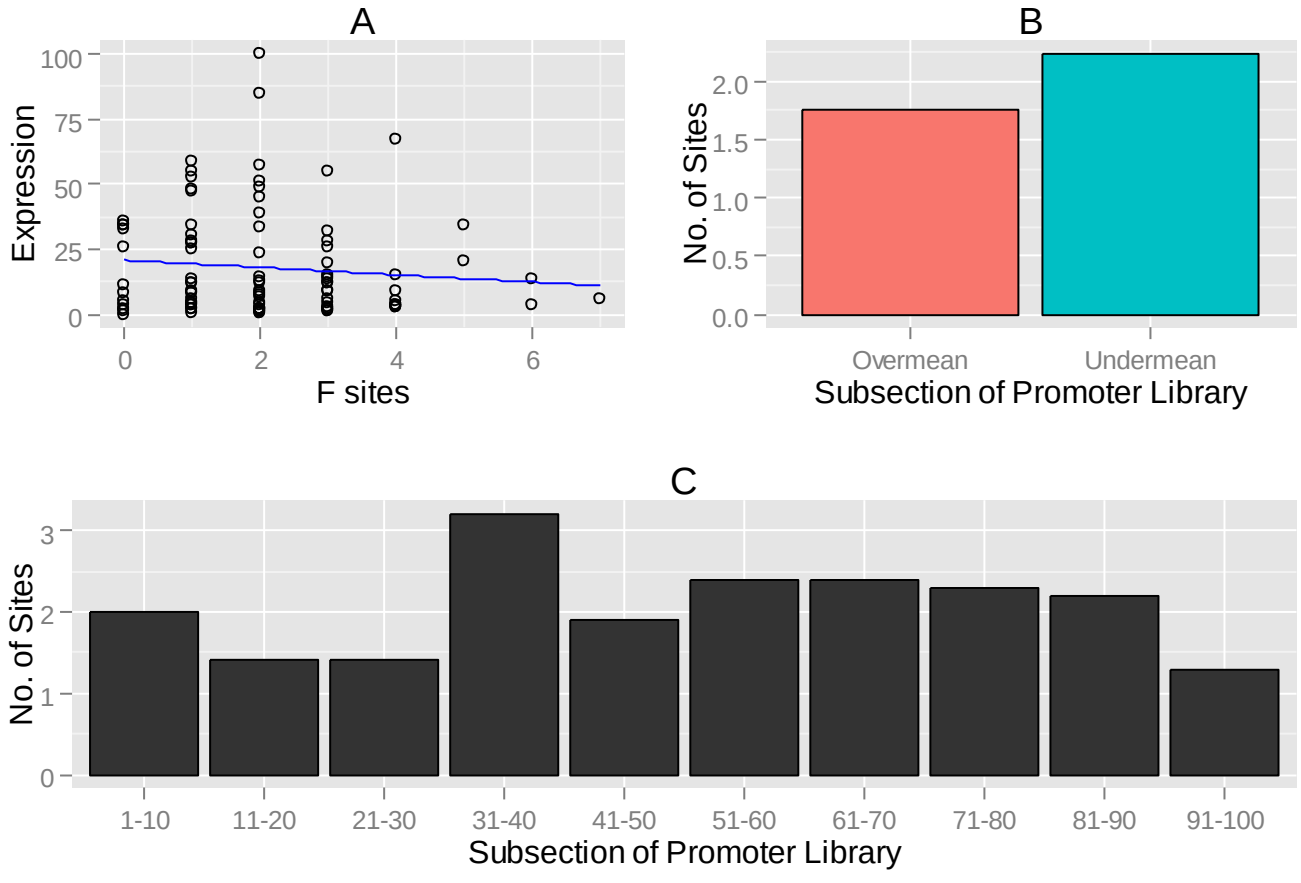
Statistics for site D



Statistics for site E



Statistics for site F



Section 3: Multiple linear regression analysis

The previous section identified TFREs to be excluded from/ included in second generation library construction. The third analysis section provides the data required to manipulate this design space by identifying optimal TFRE block stoichiometries that can be employed in next-generation library construction. Every possible multiple linear regression model (*i.e.* each combination of TFRE block variables; e.g. A, A + B, A + B + C, *etc.*) explaining promoter activity is shown in Table 2.

Table 2 Every possible linear regression model explaining promoter activity.

	A	B	C	D	E	F	C_p
1	0	0	0	0	1	0	61.4384544023012
1	0	0	0	1	0	0	78.5724635748259
1	0	0	1	0	0	0	95.1419454497566
1	0	1	0	0	0	0	107.141473786276
1	0	0	0	0	0	1	112.341919802167
1	1	0	0	0	0	0	114.350229719397
2	0	0	0	1	1	0	28.2756569589081
2	0	0	1	0	1	0	50.221521871889
2	0	1	0	0	1	0	51.8097338126759
2	0	0	1	1	0	0	55.478137419795
2	0	0	0	0	1	1	59.5660754377266
2	1	0	0	0	1	0	62.5587845193788
2	0	1	0	1	0	0	73.9041066417525
2	0	0	0	1	0	1	80.4794264130189
2	1	0	0	1	0	0	80.5677652297131
2	0	1	1	0	0	0	89.8836699216804
3	0	0	1	1	1	0	12.2380702117546
3	0	1	0	1	1	0	19.5213468747679
3	1	0	0	1	1	0	29.8616073813271
3	0	0	0	1	1	1	30.2206407123572
3	0	1	1	0	1	0	40.9277548746367
3	0	0	1	0	1	1	46.2356466543471
3	0	1	1	1	0	0	50.9522038289653
3	0	1	0	0	1	1	51.1155842369699
3	1	0	1	0	1	0	51.4097393822449
3	1	1	0	0	1	0	52.3550181535026
4	0	1	1	1	1	0	3.92181845593615
4	0	0	1	1	1	1	13.7873763075448
4	1	0	1	1	1	0	13.9029502237848
4	1	1	0	1	1	0	20.7028328390454
4	0	1	0	1	1	1	21.5149585219501
4	1	0	0	1	1	1	31.7542490162291
4	0	1	1	0	1	1	38.4363414380822
4	1	1	1	0	1	0	41.5687626775507
4	1	0	1	0	1	1	46.7979178478901
4	1	1	0	0	1	1	51.1591808761318
5	1	1	1	1	1	0	5.22380260258146
5	0	1	1	1	1	1	5.79330319039121
5	1	0	1	1	1	1	15.3267308461078
5	1	1	0	1	1	1	22.7011585925734
5	1	1	1	0	1	1	38.4460666837745
5	1	1	1	1	0	1	54.904821935078

In table 2 block inclusion/ exclusion within a model is indicated by a 1 or a 0 respectively. The total number of parameters in the model (including the intercept) and the models Cp value are shown in the first and last columns respectively. The better the models' explanation of the data, the closer its Cp value will be to the number of model parameters. Figure 3 shows the Cp of each model against its number of parameters, indicating the ability of different numbers of paramaters to significantly explain promoter strength. The best fitting model for each possible number of parameters is shown in Table 3. The most parsimonious model is then reported with associated key statistics (model summary, analysis of variance table, the contribution of each variable to $r^{2)}$ in outputs 2-5 and Figure 4.

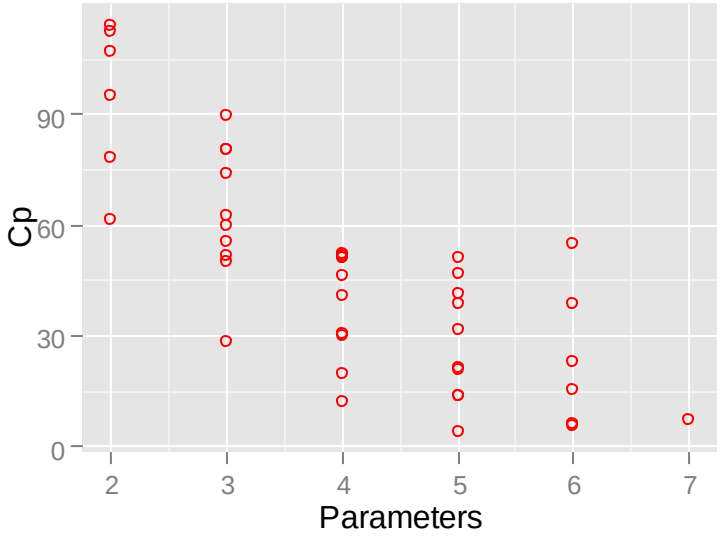


Figure 3: Model size vs. Cp

	A	B	C	D	E	F		r2
1	0	0	0	0	1	0	61.4384544023012	0.255650632826516
2	0	0	0	1	1	0	28.2756569589081	0.425125302372607
3	0	0	1	1	1	0	12.2380702117546	0.512061333509037
4	0	1	1	1	1	0	3.92181845593615	0.56178272897213
5	1	1	1	1	1	0	5.22380260258146	0.565146966595023
6	1	1	1	1	1	1	7	0.566225631401972

Table 3 Best model for each number of parameters

```
    A     B     C     D     E     F
FALSE  TRUE  TRUE  TRUE  TRUE FALSE
```

Output 2: Sites included in the best model

```
Call:
lm(formula = sp2 ~ ., data = bestest2)

Residuals:
    Min      1Q  Median      3Q     Max
-28.708  -7.958  -1.922   5.880  45.711

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.3184     3.5032   3.802 0.000258 ***
B            -3.2190     0.9963  -3.231 0.001713 **
C             3.7386     0.8859   4.220 5.72e-05 ***
D            -5.8172     0.9259  -6.282 1.09e-08 ***
E             5.8472     0.8301   7.044 3.33e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.62 on 92 degrees of freedom
Multiple R-squared:  0.5618,     Adjusted R-squared:  0.5427
F-statistic: 29.49 on 4 and 92 DF,  p-value: 8.842e-16
```

Output 3: Summary of the best model

```
Analysis of Variance Table

Response: sp2
           Df  Sum Sq Mean Sq F value    Pr(>F)
B           1  1377.5  1377.5  7.4267  0.007693 **
C           1  3614.3  3614.3 19.4862 2.755e-05 ***
D           1  7682.0  7682.0 41.4168 5.466e-09 ***
E           1  9202.0  9202.0 49.6118 3.327e-10 ***
Residuals  92 17064.2   185.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output 4: Best models anova Statistics

```
(Intercept)           B           C           D           E
 13.318357   -3.219040    3.738559   -5.817171    5.847189
```

Output 5: Model Coefficients for the best model

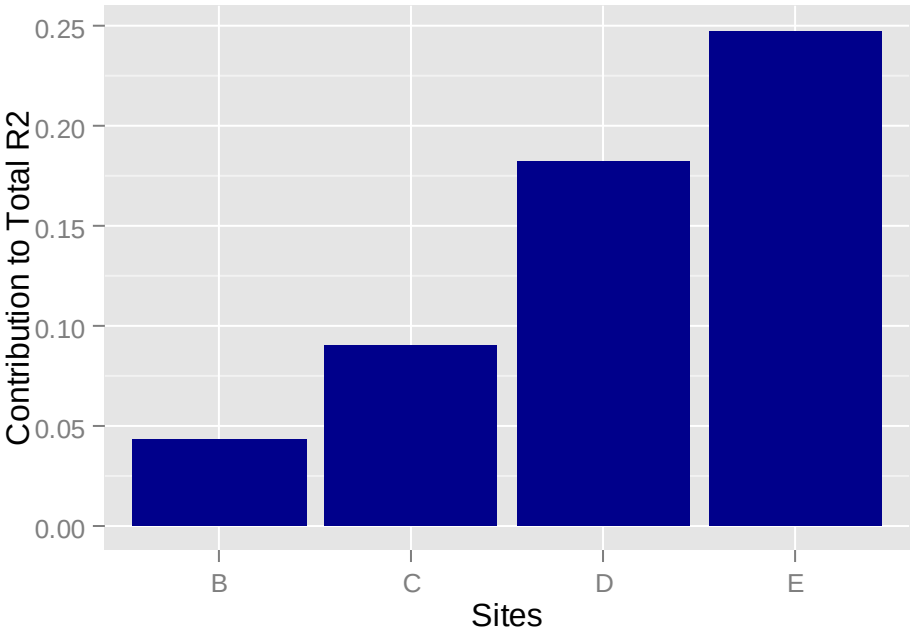


Figure 4: Relative contribution of each TFRE-block parameter to the 'best' model's r^2 statistic.

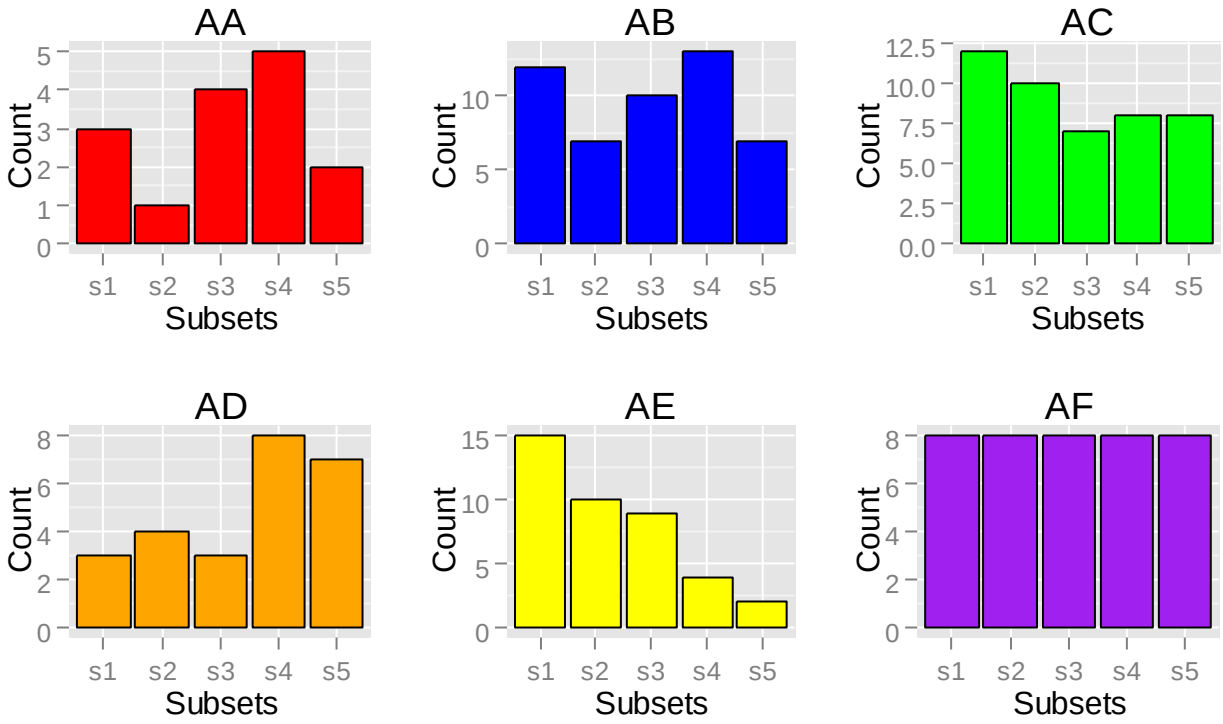
These data identify potential design solutions to specifically tailor next-generation library activities by indicating optimal TFRE block stoichiometries to increase promoter strengths (*i.e.* best model coefficients). Accordingly, at the conclusion of this analysis stage second generation library construction strategies can be implemented by employing design-led TFRE block ratios in ligation reactions.

Section 4: String analysis

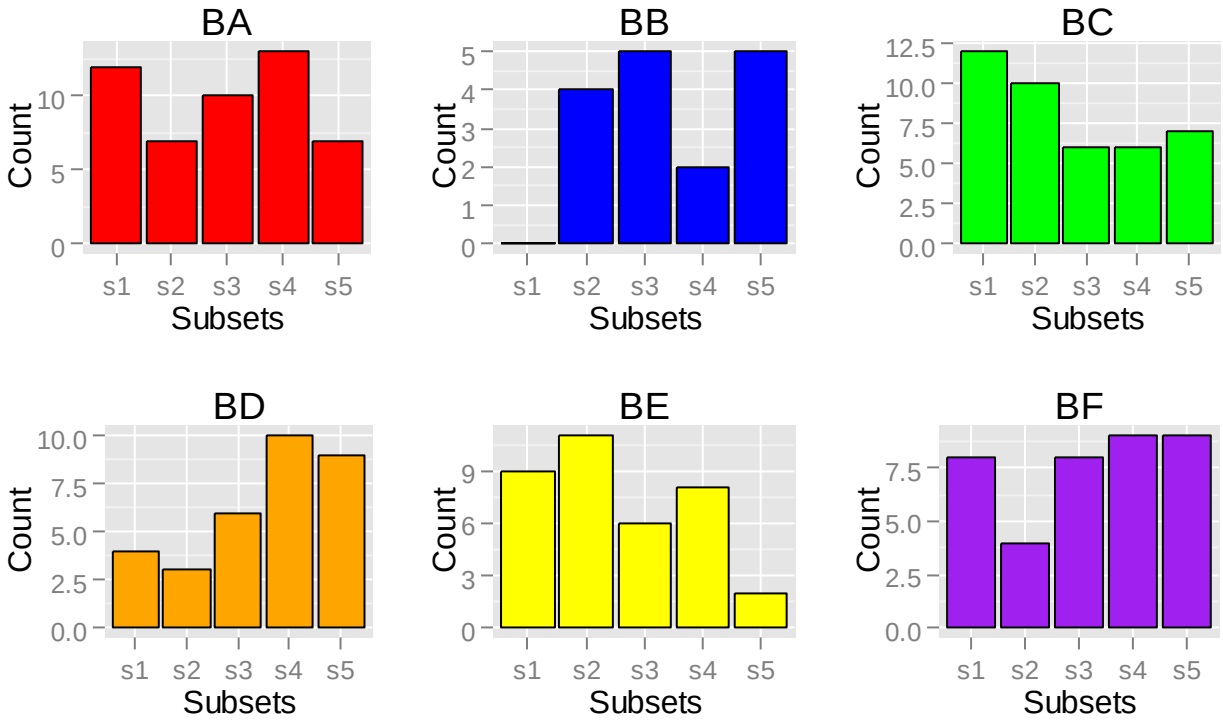
The fourth analysis section provides the data required to evaluate the effects of combinatorial interactions between neighbouring blocks within promoters. The functionally of every two and three block TFRE-string is analysed to both i) determine how neighbouring sites impact each blocks function and ii) allow characterisation of larger parts that could be utilised in future design spaces (*i.e.* blocks containing two or three discrete TFREs could be included in ligation reactions). Strings are evaluated both in the context of each discrete TFRE and as a collection of comparative block parts.

First, the relative abundance of 2-block TFRE strings in synthetic promoters of varying activities is analysed. The number of promoters within discrete library subsections (subsection 1 contains the top 20% of promoters, ranked according to strength) that contain each possible 2-block string is reported. Figures are grouped to show the relative differences between every possible 2-block string containing each discrete TFRE. Counts represent either orientation, e.g. AB could occur as either ‘A B’ or ‘B A’.

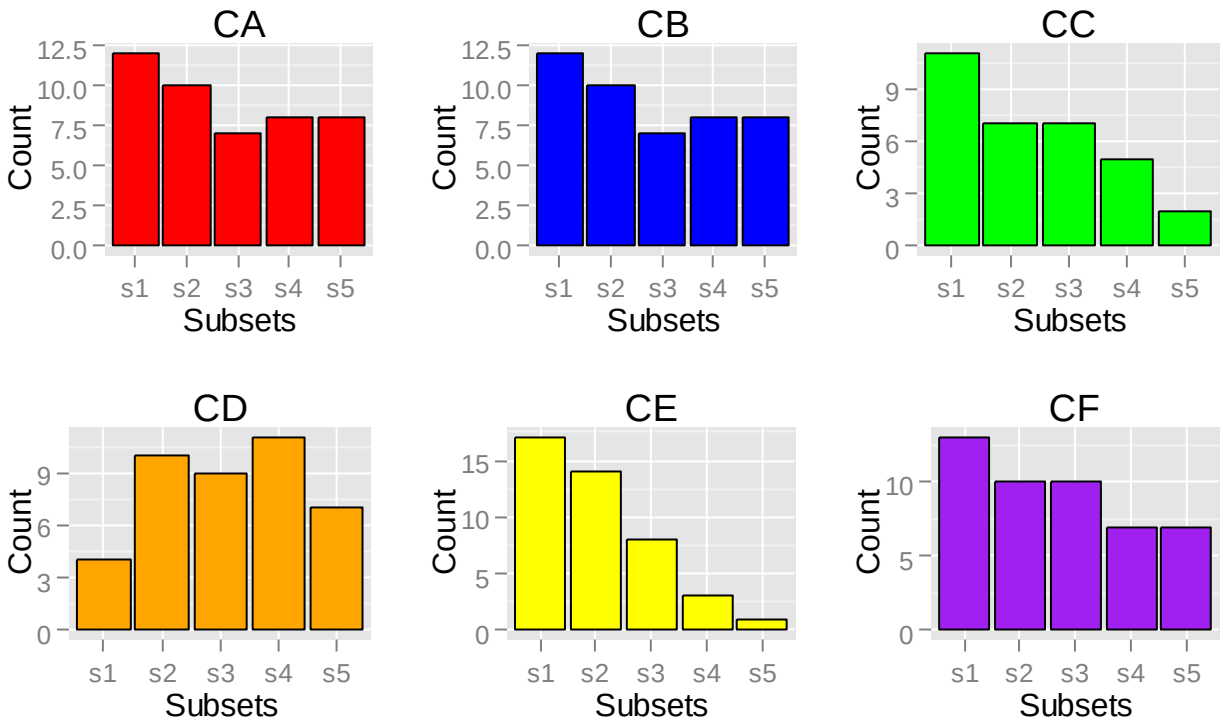
Statistics for site A



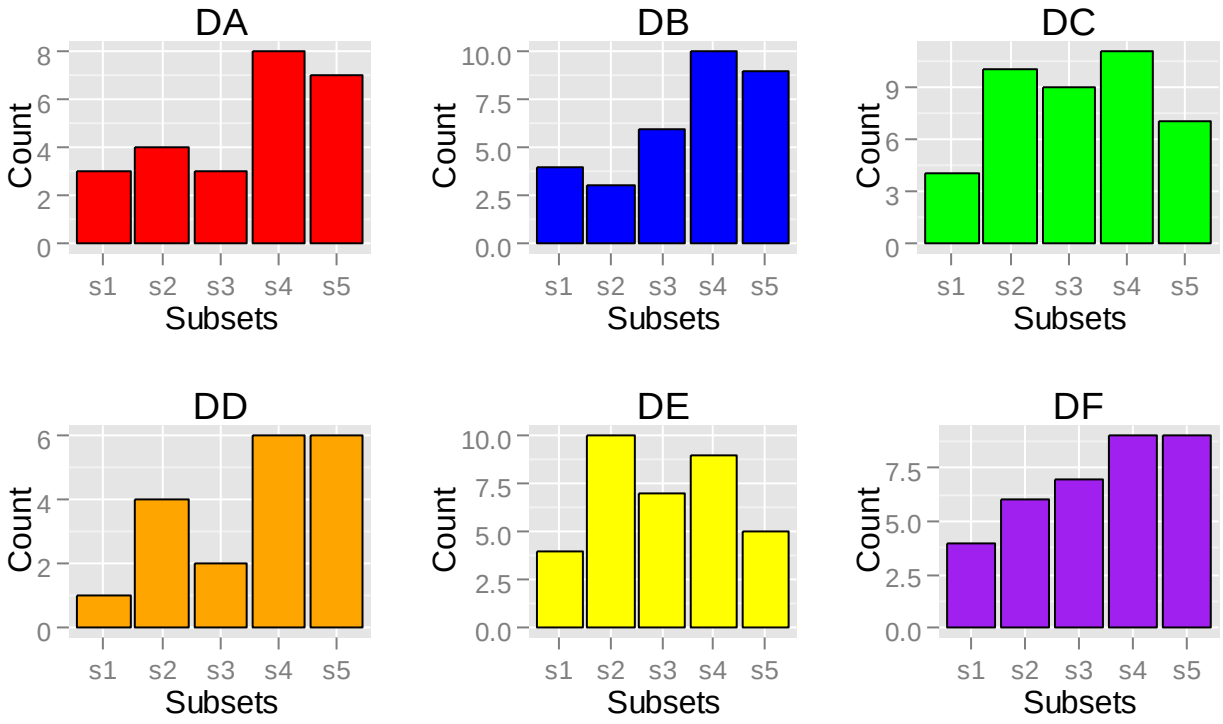
Statistics for site B



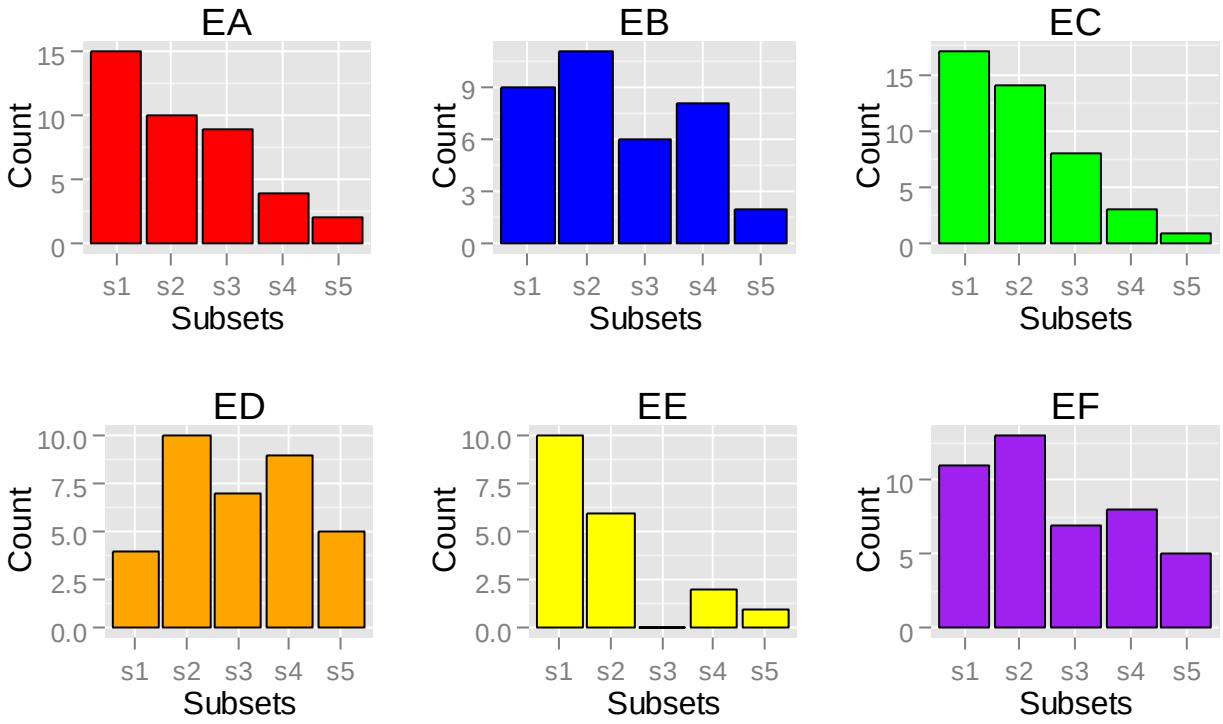
Statistics for site C



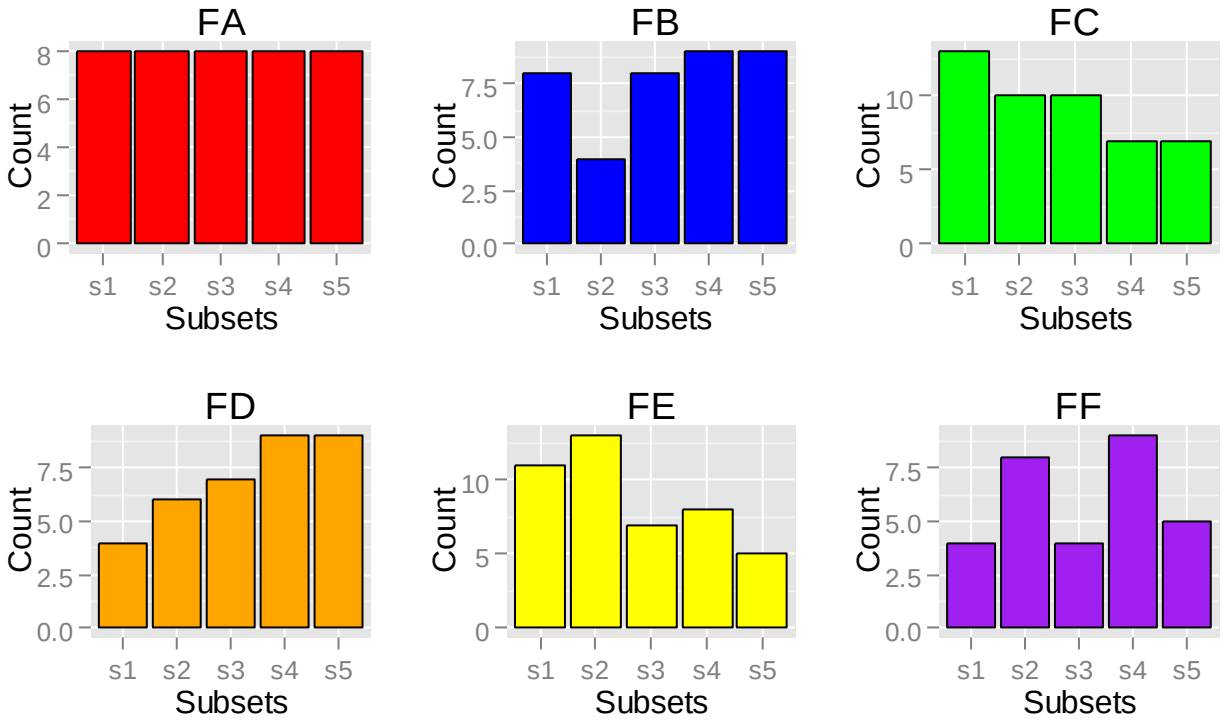
Statistics for site D



Statistics for site E



Statistics for site F



It is now useful to consider which of these two strings are most associated with high expression. Firstly, Figure 6 shows the percentage of the top 15 promoters which contain each 2string. Figure 7 shows the average library position of promoters containing each 2 string. These two figures enable identification of strings significantly associated with strong promoters.

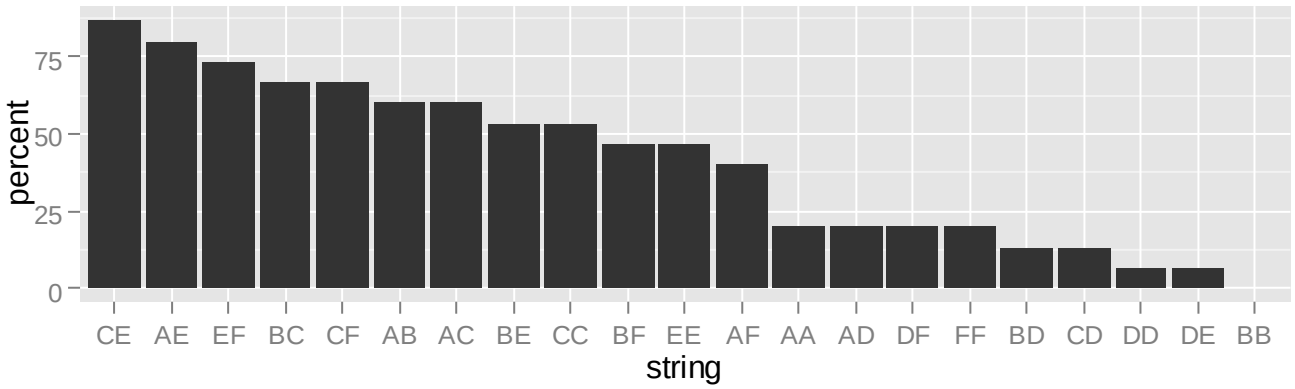


Figure 7: Percentage of the top 15 promoters in the library that contain each possible 2 string.

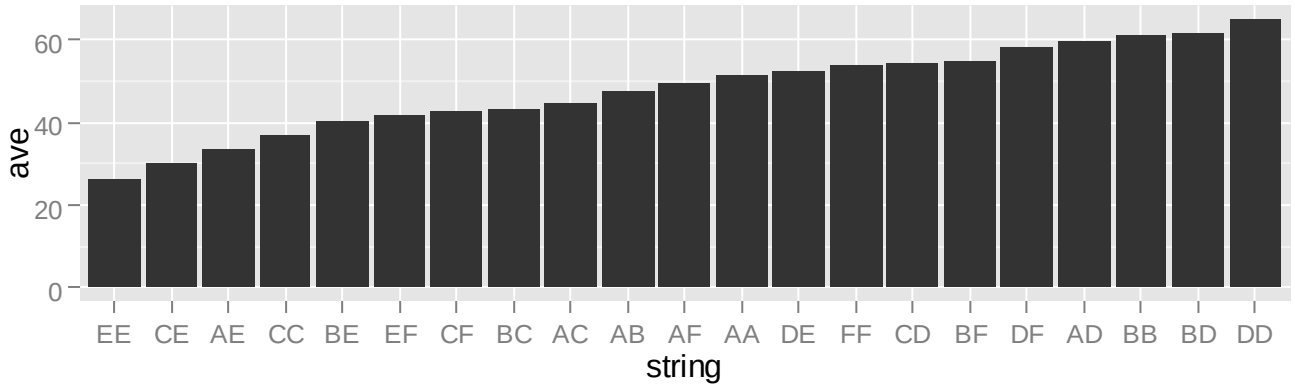
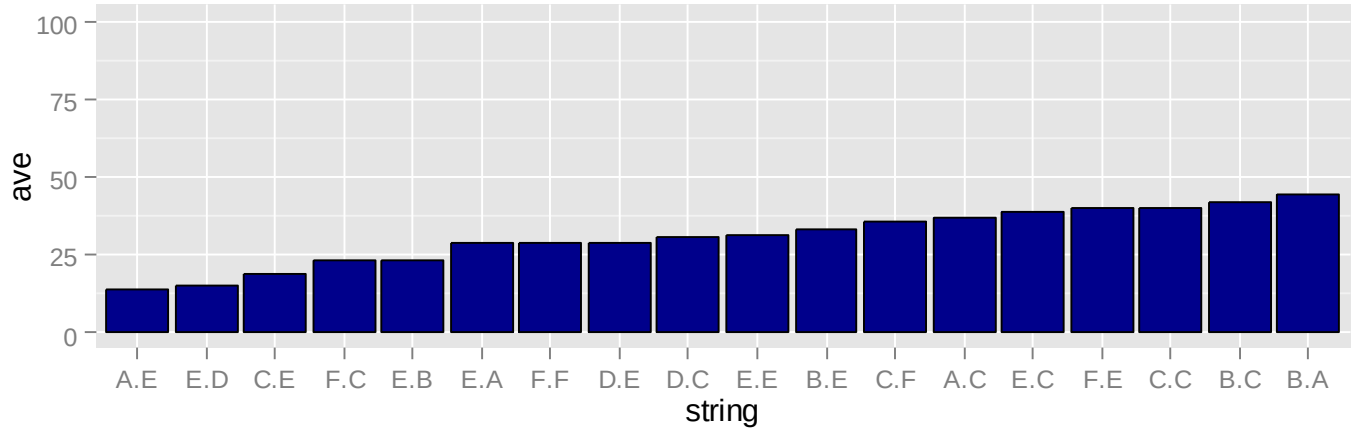
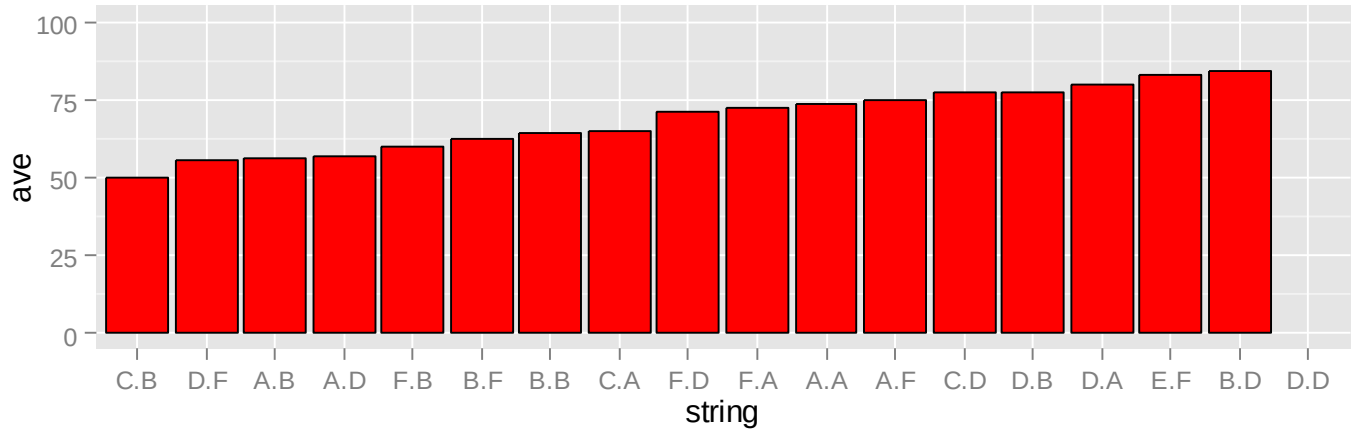


Figure 8: Average library position of promoters containing each possible two string.

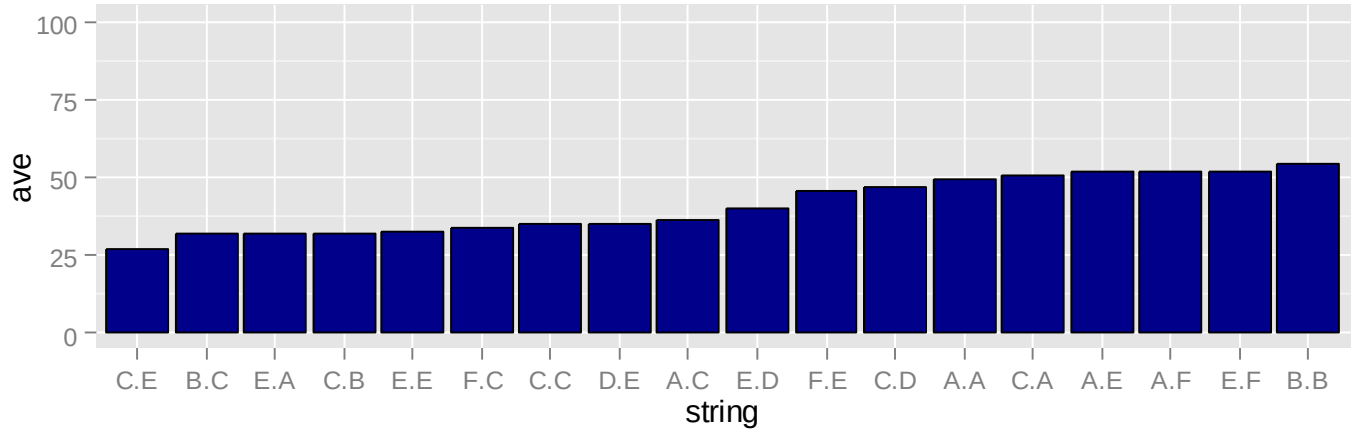
Next, the relative abundance of 3-block TFRE strings in synthetic promoters of varying activities is analysed. For each discrete TFRE a figure is generated detailing the average library position of promoters containing each possible 3-string where that TFRE is the central block (e.g. A.E and C.B refer to AEE and CEB respectively). Each TFRE figure is split into two sections to highlight higher (blue) and lower (red) performing 3-strings respectively.

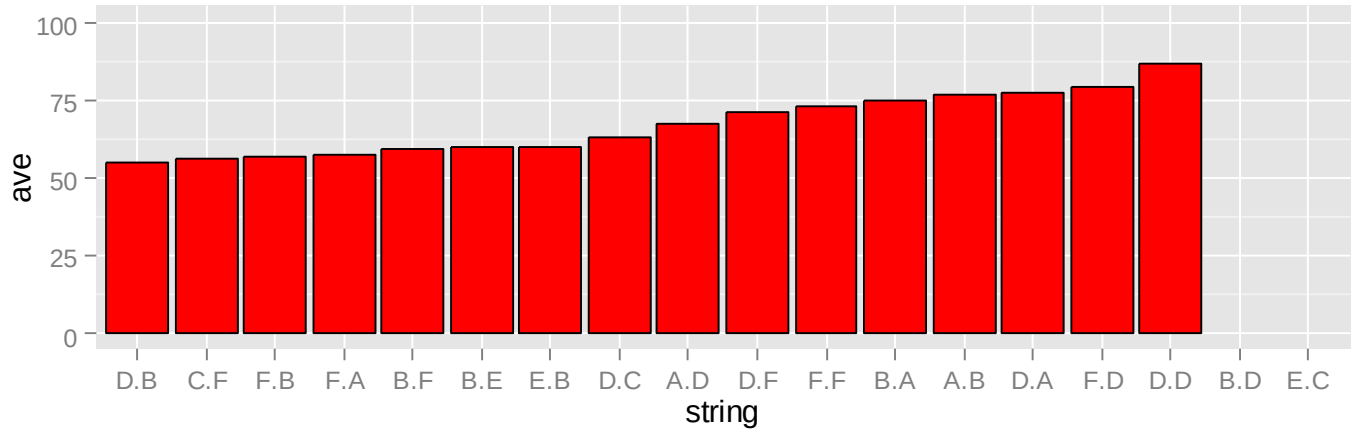
Statistics for site A



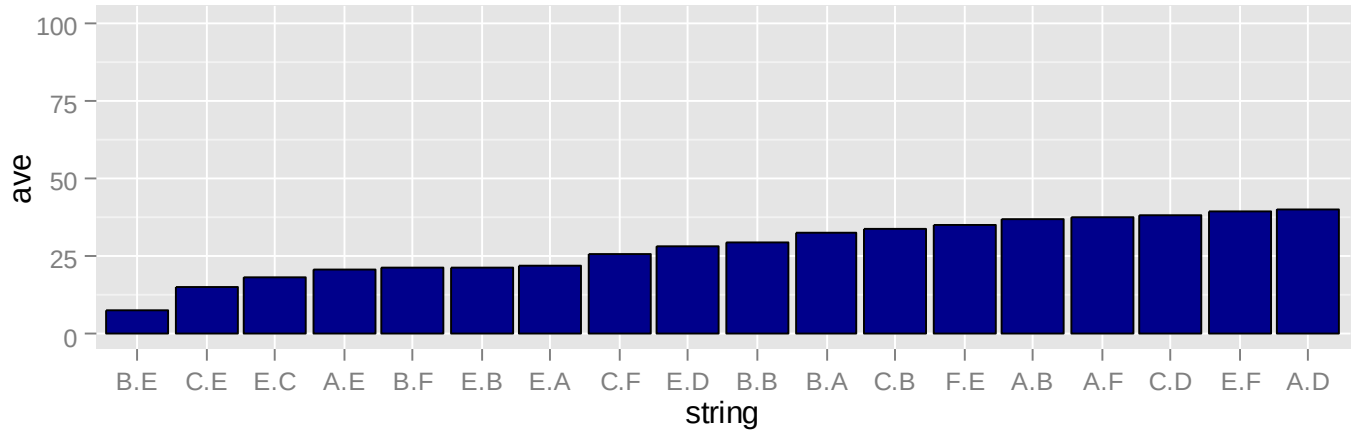


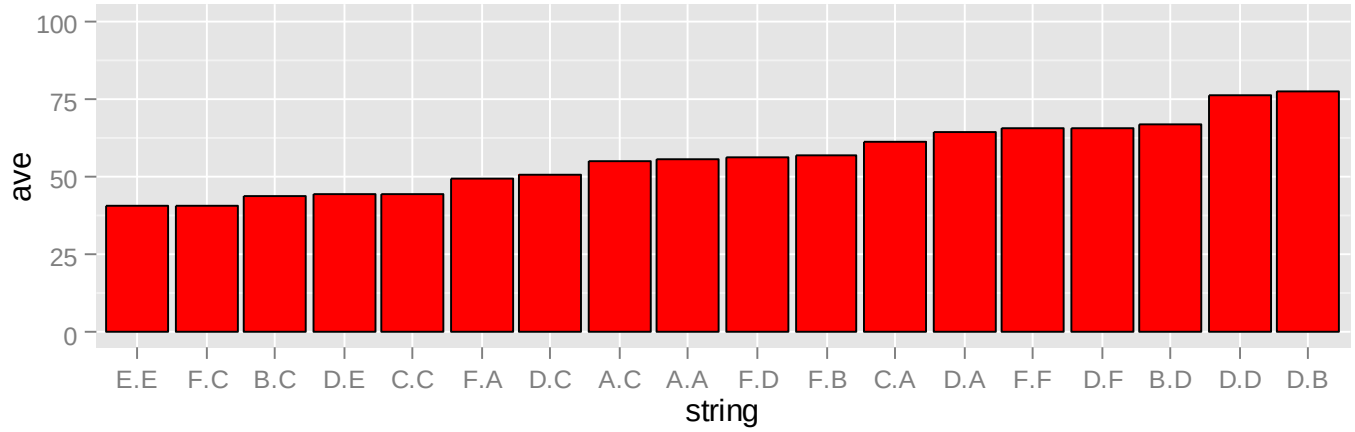
Statistics for site B



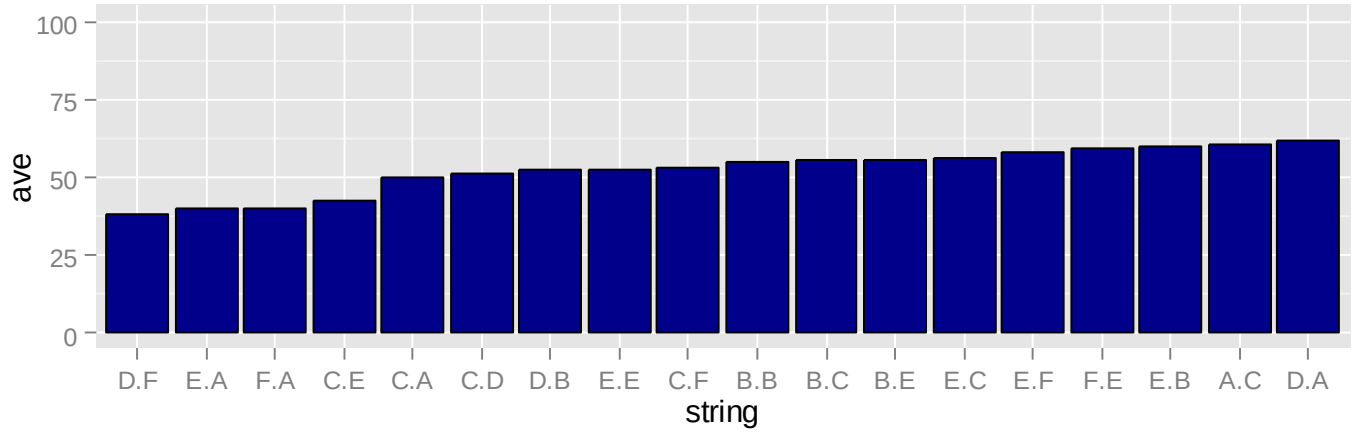


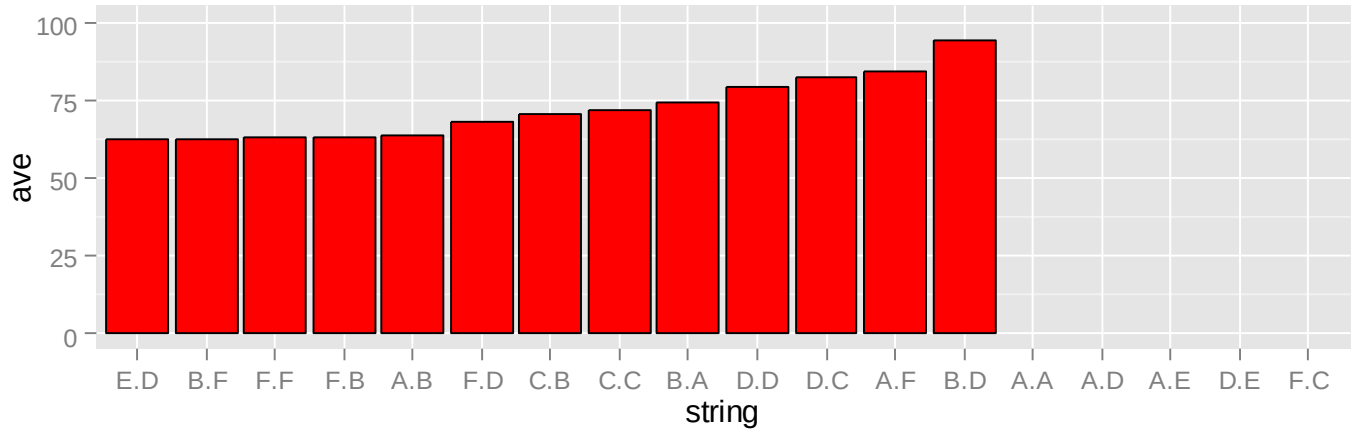
Statistics for site C



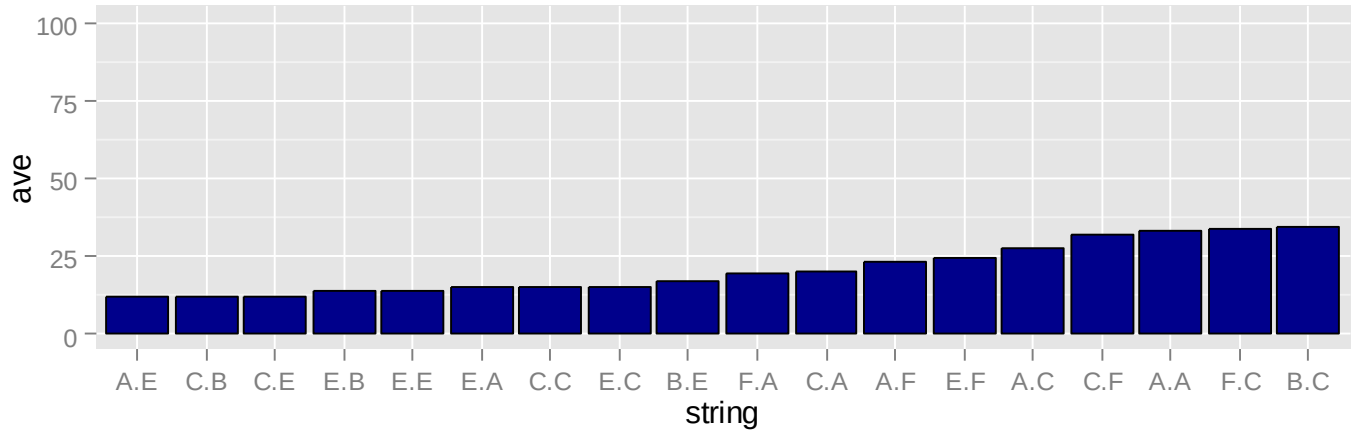


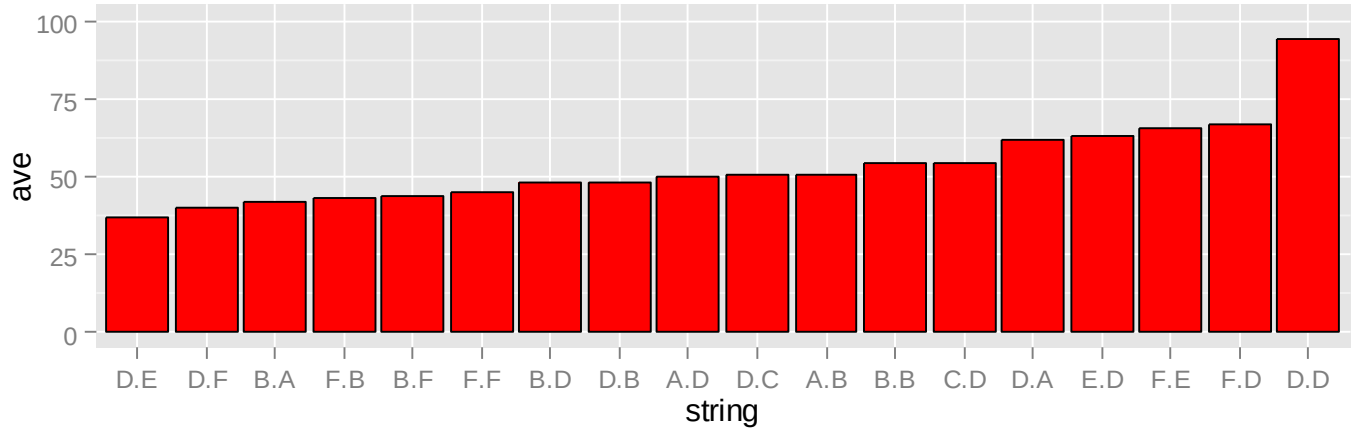
Statistics for site D



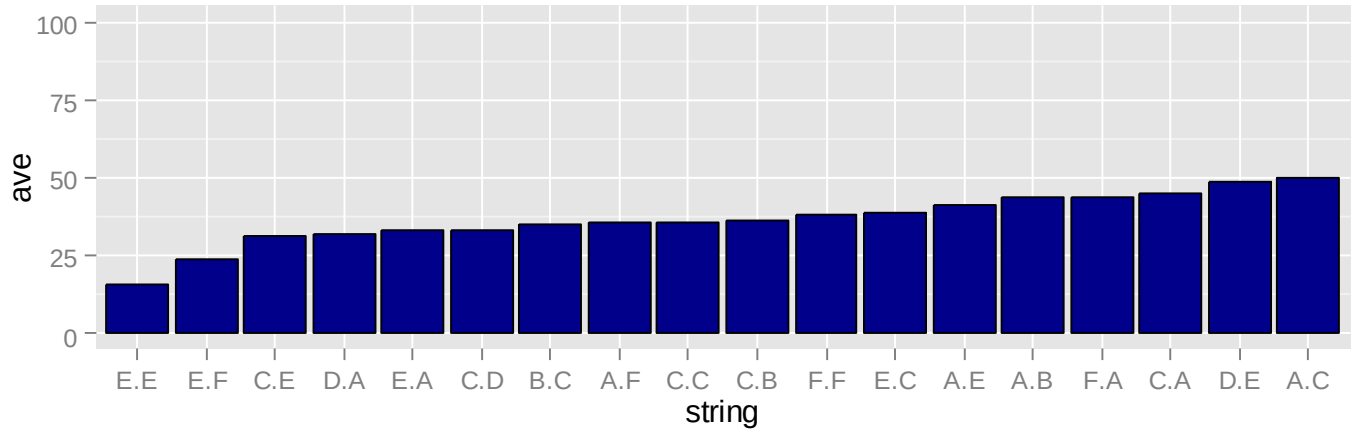


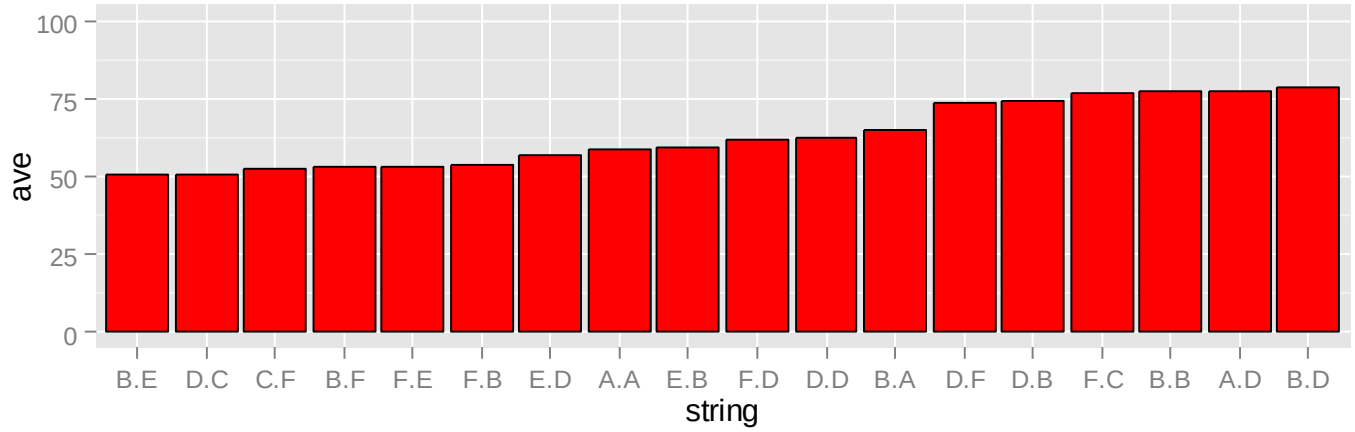
Statistics for site E





Statistics for site F





It is now useful to consider which of these three strings are most associated with high expression. Firstly, Figure 7 shows the top 20 3strings in terms of the average library position of promoters they appear in. Figure 8 is a check for any strings that are significantly associated with the very top promoters, showing the percentage of the top 10 promoters (ranked by activity) that contain each 3string (the top 10 strings are shown). Finally, table 4 shows the relative activities of all 3-block TFRE strings. The average library position of promoters containing each possible 3-string is shown.

The data generated in this section identifies whether discrete TFRE-strings need to be avoided (for example if neutral regulators negatively affect neighbouring positive regulators) or employed in next-generation library construction. Accordingly, second generation library design spaces are finalised at the conclusion of this final analysis stage.

Assumption 4: Neutral regulators do not negatively affect neighbouring positive regulators. If false these sites will be removed from library 2 = TRUE/FALSE

Assumption 5: Discrete combinations of positive (or neutral) regulators are not positively correlated with high promoter activity. If false these larger parts can be utilised in the next design space. = TRUE/FALSE

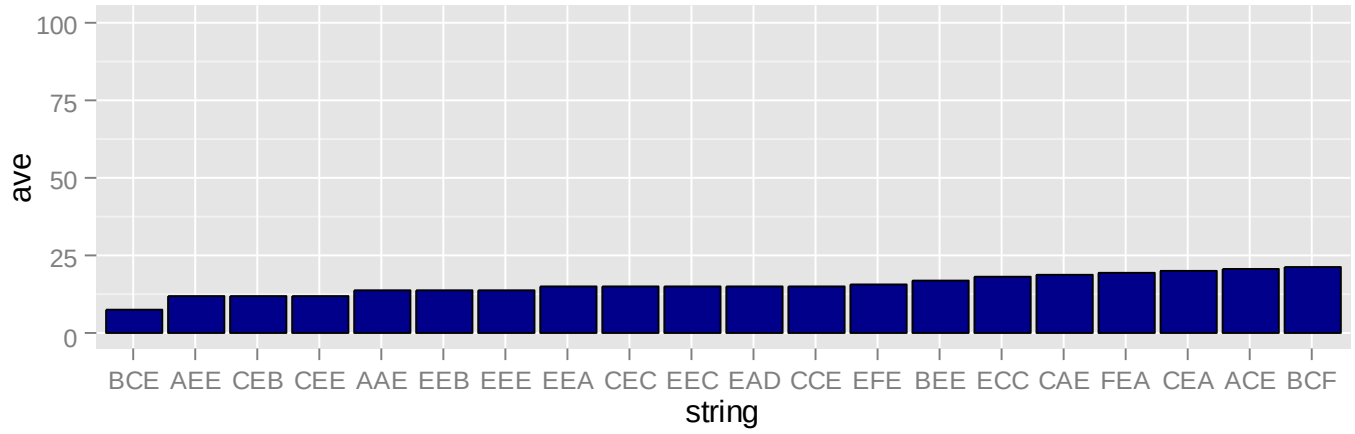


Figure 7: Top 20 3strings in terms of average library position of promoters that contain each string.

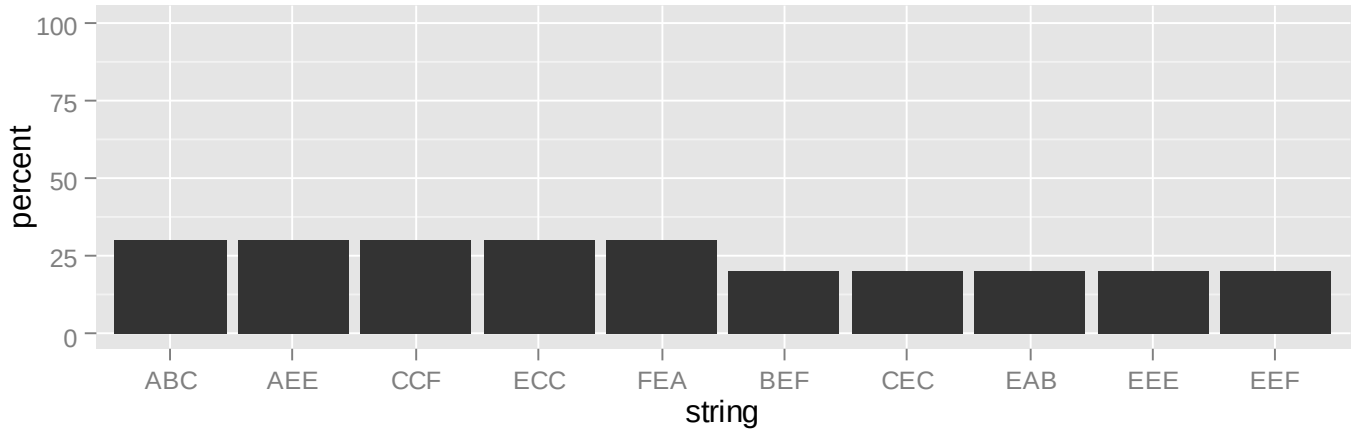


Figure 8: Top 10 3strings in terms of the percentage of the strongest 10 promoters which contain them.

	String	Ave	string	Ave	string	Ave	string	Ave	string	Ave	string	Ave
1	AAA	73.50000	ABA	49.50000	ACA	55.57143	ADA	NA	AEA	33.28571	AFA	58.80000
2	AAB	56.00000	ABB	76.50000	ACB	36.60000	ADB	63.75000	AEB	50.66667	AFB	43.50000
3	AAC	36.50000	ABC	36.30000	ACC	54.80000	ADC	60.25000	AEC	27.16667	AFC	50.00000
4	AAD	57.00000	ABD	67.00000	ACD	39.80000	ADD	NA	AED	49.66667	AFD	77.50000
5	AAE	14.00000	ABE	51.60000	ACE	20.87500	ADE	NA	AEE	11.83333	AFE	40.83333
6	AAF	74.75000	ABF	51.83333	ACF	37.44444	ADF	84.20000	AEF	23.00000	AFF	35.50000
7	BAA	44.00000	BBA	74.66667	BCA	32.60000	BDA	74.00000	BEA	41.71429	BFA	64.66667
8	BAB	64.14286	BBB	54.50000	BCB	29.00000	BDB	55.00000	BEB	54.00000	BFB	77.20000
9	BAC	41.60000	BBC	31.50000	BCC	43.50000	BDC	55.25000	BEC	34.00000	BFC	34.75000
10	BAD	84.00000	BBD	NA	BCD	66.37500	BDD	94.00000	BED	47.75000	BFD	78.50000
11	BAE	33.12500	BBE	59.75000	BCE	7.75000	BDE	55.66667	BEE	17.00000	BFE	50.20000
12	BAF	62.25000	BBF	59.50000	BCF	21.00000	BDF	62.50000	BEF	43.71429	BFF	52.71429
13	CAA	64.66667	CBA	50.33333	CCA	61.28571	CDA	49.75000	CEA	20.25000	CFA	44.80000
14	CAB	49.88889	CBB	32.00000	CCB	33.75000	CDB	70.33333	CEB	12.00000	CFB	36.14286
15	CAC	40.20000	CBC	34.75000	CCC	44.50000	CDC	71.66667	CEC	14.75000	CFC	35.66667
16	CAD	77.00000	CBD	47.00000	CCD	38.28571	CDD	51.00000	CED	54.00000	CFD	33.33333
17	CAE	18.66667	CBE	26.83333	CCE	15.25000	CDE	42.22222	CEE	12.00000	CFE	31.00000
18	CAF	35.60000	CBF	55.80000	CCF	25.85714	CDF	52.75000	CEF	31.50000	CFF	52.40000
19	DAA	80.00000	DBA	77.00000	DCA	64.40000	DDA	61.50000	DEA	61.60000	DFA	31.50000
20	DAB	77.50000	DBB	54.75000	DCB	77.00000	DDB	52.50000	DEB	48.00000	DFB	74.25000
21	DAC	30.66667	DBC	63.25000	DCC	50.33333	DDC	82.00000	DEC	50.33333	DFC	50.20000
22	DAD	NA	DBD	86.33333	DCD	75.83333	DDD	79.25000	DED	94.00000	DFD	62.62500
23	DAE	28.66667	DBE	34.75000	DCE	44.00000	DDE	NA	DEE	37.00000	DFE	48.50000
24	DAF	55.66667	DBF	71.20000	DCF	65.50000	DDF	38.00000	DEF	39.66667	DFF	73.42857
25	EAA	28.40000	EBA	31.71429	ECA	22.00000	EDA	40.00000	EEA	14.66667	EFA	33.00000
26	EAB	23.33333	EBB	60.00000	ECB	21.50000	EDB	59.66667	EEB	14.00000	EFB	59.00000
27	EAC	38.90909	EBC	NA	ECC	18.11111	EDC	55.83333	EEC	14.75000	EFC	38.42857
28	EAD	15.00000	EBD	40.00000	ECD	28.00000	EDD	62.33333	EED	62.66667	EFD	57.00000
29	EAE	30.85714	EBE	32.50000	ECE	40.25000	EDE	52.50000	EEE	14.00000	EFE	15.50000
30	EAF	83.00000	EBF	52.00000	ECF	39.20000	EDF	57.66667	EEF	24.33333	EFF	23.66667
31	FAA	72.00000	FBA	57.25000	FCA	49.33333	FDA	40.00000	FEA	19.60000	FFA	43.71429
32	FAB	59.66667	FBB	56.50000	FCB	56.75000	FDB	63.00000	FEB	43.25000	FFB	53.50000
33	FAC	23.16667	FBC	33.75000	FCC	40.66667	FDC	NA	FEC	33.42857	FFC	76.66667
34	FAD	71.00000	FBD	79.00000	FCD	56.00000	FDD	67.80000	FED	66.62500	FFD	61.75000
35	FAE	39.80000	FBE	45.60000	FCE	35.00000	FDE	59.50000	FEE	65.33333	FFE	52.80000
36	FAF	28.50000	FBF	73.00000	FCF	65.25000	FDF	62.77778	FEF	44.66667	FFF	37.80000

Table 4. Relative activities of each 3-block TFRE string.

Section 5: Next generation synthetic promoter library design criteria

Based on the analysis provided, the optimal design space for next-generation promoter library construction can now be confirmed.

TFRE	Element Name	Stochiometric Weighting
A	**X**	**X**
B	**X**	**X**
C	**X**	**X**
D	**X**	**X**
E	**X**	**X**
F	**X**	**X**

Table 5. Next-generation library construction criteria

Appendix C: Published article 1 – Brown *et al.* 2014

ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Synthetic Promoters for CHO Cell Engineering

Adam J. Brown,[1] Bernie Sweeney,[2] David O. Mainwaring,[2] David C. James[1]

[1]Department of Chemical and Biological Engineering, University of Sheffield, University of Sheffield, Mappin St., Sheffield S1 3JD, England; telephone: +44-0-114-222-7505; fax: +44-0-114-222-7501; e-mail: d.c.james@sheffield.ac.uk;

[2]Protein Expression and Purification Group, UCB, Berkshire, England

ABSTRACT: We describe for the first time the creation of a library of 140 synthetic promoters specifically designed to regulate the expression of recombinant genes in CHO cells. Initially, 10 common viral promoter sequences known to be active in CHO cells were analyzed using bioinformatic sequence analysis programs to determine the identity and relative abundance of transcription factor regulatory elements (TFREs; or transcription factor binding sites) they contained. Based on this, 28 synthetic reporters were constructed that each harbored seven repeats of a discrete TFRE sequence upstream of a minimal CMV core promoter element and secreted alkaline phosphatase (SEAP) reporter gene. After evaluation of the relative activity of TFREs by transient expression in CHO-S cells, we constructed a first generation library of 96 synthetic promoters derived from random ligation of six active TFREs inserted into the same reporter construct backbone. Comparison of the sequence and relative activity of first generation promoters revealed that individual TFRE blocks were either relatively abundant in active promoters (NFκB, E-box), equally distributed across promoters of varying activity (C/EBPα, GC-box) or relatively abundant in low activity promoters (E4F1, CRE). These data were utilized to create a second generation of 44 synthetic promoters based on random ligation of a fixed ratio of 4 TFREs (NFκB 5: E-box 3: C/EBPα 1: GC-box 1). Comparison of the sequence and relative activity of second generation promoters revealed that the most active promoters contained relatively high numbers of both NFκB and E-box TFREs in approximately equal proportion, with a correspondingly low number of GC-box and C/EBPα blocks. The most active second generation promoters achieved approximately twice the activity of a control construct harboring the human cytomegalovirus (CMV) promoter. Lastly, we evaluated the function of a subset of synthetic promoters exhibiting a broad range of activity in different CHO cell host cell lines (CHO-S, CHO-K1, and CHO-DG44) and across extended fed-batch transient expression in CHO-S cells. In general, the different synthetic promoters both maintained their relative activity and the most active promoters consistently and significantly exceeded the activity of the CMV control promoter. For advanced cell engineering strategies our synthetic promoter libraries offer precise control of recombinant transcriptional activity in CHO cells spanning over two orders of magnitude.

Biotechnol. Bioeng. 2014;9999: 1–10.

KEYWORDS: synthetic promoter; Chinese hamster ovary cells; cell engineering; synthetic biology

The authors have a patent application filed based on the work in this paper.
Correspondence to: D.C. James
Received 18 December 2013; Revision received 18 February 2014; Accepted 19 February 2014
Accepted manuscript online xx Month 2014;
Article first published online in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/bit.25227

Introduction

Recombinant gene transcription in Chinese hamster ovary (CHO) cells is still routinely controlled with strong viral promoters (such as the human cytomegalovirus immediate early 1 (hCMV-IE1) promoter) despite links to cellular stress induction, cell-cycle dependency, and epigenetic silencing (Dale, 2006; Kim et al., 2011). CMV is a relatively potent promoter, currently used to drive expression of many biopharmaceutical products. It is a highly complex element evolved by the virus to enable it to infect many mammalian host cells (Stinski and Isomura, 2008). However (surprisingly, given its use for >25 years to drive biopharmaceutical production) little is known about how it functions mechanistically in the CHO cell and therefore strategies to precisely control or improve its transcriptional activity are not generally available. Synthetic promoters therefore offer a potentially attractive solution, as they can replace functionally ill defined and uncontrollable genetic elements in expression vectors with sophisticated, bespoke controllers that can engineer host cell function predictably. They are fundamental components for the successful application of synthetic biology to mammalian cell factories, where synthetic biology aims to assemble characterized, modular parts into novel, predictable devices (Pasotti et al., 2012). Design of predictable genetic systems is a core objective underpinning our ability to create next-generation mammalian cell factories.

Synthetic promoters have been traditionally produced by screening either (i) randomised DNA sequences or synthetic oligonucleotide repeats or (ii) assemblies of known *cis*-regulatory elements [e.g., transcription factor binding sites,

or regulatory elements (TFREs)], both upstream of minimal core promoter motifs. These efforts have resulted in synthetic promoter libraries designed to function in a range of organisms, predominatly microbial, such as *Corynebacterium* (Yim et al., 2013), *Saccharomyces* (Blazeck et al., 2012), *Pichia* (Stadlmayr et al., 2010), and *Streptomyces* (Seghezzi et al., 2011), and also in mammalian cells (Ferreira et al., 2011; Ogawa et al., 2007; Schlabach et al., 2010). However, these studies have demonstrated that the exquisite gene expression control offered by synthetic promoters is highly context-dependent, necessitating promoters to be specifically constructed for each host cell type. For example, Schlabach et al. (2010) describe the synthesis of synthetic promoters for mammalian cells screened in HeLa cells (207% of CMV activity), but also report highly variable reporter gene expression in different mammalian cell types (21–113% of CMV).

Few studies have previously explored the utilization of synthetic promoters in CHO. Tornøe et al. (2002) created a small pool of promoters (<20) with a tenfold range in activity by randomizing the sequences separating TFREs within a chimeric promoter. Further, Grabherr et al. (2011) constructed five synthetic promoters with approximately equivalent activities by constructing contiguous sequences with nucleotide compositions mimicking those found in highly active promoters. Both of these studies targeted broad activity in mammalian cells (tested in multiple diverse cell lines) and neither sought to specifically design synthetic promoters to function in concert with the transactivational machinery of CHO cell factories. Accordingly, neither of these small libraries enable predictable, precise, robust control of CHO gene expression over broad dynamic ranges and synthetic promoter activities reported were significantly below that of CMV. Lastly, Le et al. (2013) recently utilized transcriptomics data to identify CHO endogenous promoters with desired expression dynamics. Whilst this is a promising avenue to identify promoters with discrete expression levels, it is limited to naturally occurring activities. Moreover, it can be a significant challenge to define the genomic regulatory sequences controlling expression of specific genes.

In this study we constructed the first libraries of synthetic promoters designed specifically to harness the pre-existing transcriptional activation machinery of CHO cell factories. We functionally screened TFRE function in CHO cells and utilized active elements to construct libraries of synthetic promoters that exhibited variable activity over two orders of magnitude, significantly exceeding that of hCMV-IE1 in transient production processes. Moreover, relative promoter activities across a broad dynamic range were maintained in both different CHO cell hosts and across a fed-batch transient production process. The precision control of recombinant gene transcription enabled by this synthetic promoter technology will facilitate the design of novel, predictable synthetic constructs for diverse CHO cell engineering applications.

Materials and Methods

In Silico Analysis of Transcription Factor Regulatory Elements

The following promoter sequences were retrieved from GenBank: hCMV-IE1 (accession number M60321.1), mouse CMV-IE1 (M11788), rat CMV-IE1 (U62396), guinea pig CMV-IE1 (CS419275), mouse CMV-IE2 (L06816.1), simian virus 40 early promoter and enhancer (NC_001669.1), adenovirus major late promoter (KF268310), myeloproliferative sarcoma virus long terminal repeat (LTR) (K01683.1), rous sarcoma virus LTR (J02025.1), and human immunodeficiency virus LTR (K03455.1). Promoters were analyzed using the Transcription Element Search System (TESS: http://www.cbil.upenn.edu/cgi-bin/tess/tess) and the Transcription Affinity Prediction tool (TRAP: http://trap.molgen.mpg.de/cgi-bin/trap_form.cgi) according to the methods previously described by Manke et al. (2008) and Schug (2008).

TFRE-Reporter Vector Construction

Previously described (Brown et al., 2013) promoterless reporter-vectors (subcloned from pSEAP2control (Clontech, Oxford, UK)) were utilized in this study. These plasmids contain a minimal hCMV-IE1 core promoter (5′-AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGAC-3′) upstream of either the secreted alkaline phosphatase (SEAP) or turbo green fluorescent protein (GFP) open reading frame (ORF). To create TFRE reporter plasmids synthetic oligonucleotides containing 7× repeat copies of the TFRE consensus sequences in Table I were synthesized (Sigma, Poole, UK), PCR amplified, and inserted into *Kpn*I and *Xho*I sites upstream of the CMV core promoter. Three clonally derived plasmids for each TFRE reporter were purified using a Qiagen plasmid mini kit (Qiagen, Crawley, UK). The sequence of all plasmid constructs was confirmed by DNA sequencing.

Synthetic Promoter Library Construction

Synthetic promoter building blocks were constructed from complementary single stranded 5′ phosphorylated oligonucleotides (Sigma), annealed in STE buffer (100 mM NaCl, 50 mM Tris–HCl, 1 mM EDTA, pH 7.8, Sigma) by heating at 95°C for 5 min, prior to ramp cooling to 25°C over 2 h. Oligonucleotides were designed such that the resulting double stranded blocks contained the specific TFRE (Table I) and a 4 bp TCGA single stranded overhang at each 5′ termini. For example, the sequences used for the NFκB-RE block were as follows (RE site underlined): 5′-TCGATGGGACTTTCCA-3′ and 5′-TCGATGGAAAGTCCCA-3′. Synthetic promoter libraries were constructed by ligating block molecules at appropriate stoichiometric molar ratios with high concentration T4 DNA ligase (Life Technologies, Paisley, UK). A "cloning-block" containing *Kpn*I and *Xho*I sites was included

Table I. DNA sequences of transcription factor regulatory elements identified by bioinformatic survey of viral promoters.

Transcription factor response/regulatory element (RE)	Sequence
Activator protein 1 (AP1)	TGACTCA
CC(A/T)$_6$GG element (CArG)	CCAAATTTGG
CCAAT displacement protein (CDP)	GGCCAATCT
CCAAT-enhancer binding protein alpha (C/EBPα)	TTGCGCAA
Cellular myeloblastosis (cMyb)	TAACGG
cAMP RE (CRE)	TGACGTCA
Elongation factor 2 (E2F)	TTTCGCGC
E4F1	GTGACGTAAC
Early growth response protein 1 (EGR1)	CGCCCCCGC
Estrogen-related receptor alpha RE (ERRE)	AGGTCATTTTGACCT
Enhancer box (E-box)	CACGTG
GATA-1 (GATA)	AGATAG
GC-box	GGGGCGGGG
Glucocorticoid RE (GRE)	AGAACATTTTGTTCT
Growth factor independence 1 (Gfi1)	AAAATCAAC
Helios RE (HRE)	AATAGGGACTT
Hepatocyte nuclear factor 1 (HNF)	GGGCCAAAGGTCT
Insulin promoter factor 1 (IPF1)	CCCATTAGGGAC
Interferon-stimulated RE (ISRE)	GAAAAGTGAAACC
Myocyte enhancer factor 2 (MEF2)	CTAAAAATAG
Msx homeobox (MSX)	CGGTAAATG
Nerve growth factor-induced gene-B RE (NBRE)	AAAGGTCA
Nuclear factor 1 (NF1)	TTGGCTATATGCCAA
Nuclear factor of activated T cells (NFAT)	AGGAAATC
Nuclear factor kappa B (NFκB)	GGGACTTTCC
Octamer motif (OCT)	ATTAGCAT
Retinoic acid RE (RARE)	AGGTCATCAAGAGGTCA
Yin yang 1 (YY1)	CGCCATTTT
Random 8mer (8mer)	TTTCTTTC

Ten viral promoters known to exhibit activity in CHO cells were surveyed for the presence of discrete transcription factor regulatory elements (transcription factor binding sites) using Transcription Element Search System (TESS) and Transcription Affinity Prediction (TRAP) algorithms using stringent search parameters to minimize false positives. DNA sequences of single TFREs that occur in more than one viral promoter are listed. Measurement of their relative ability to activate transcription of recombinant reporter genes in CHO-S cells is shown in Figure 1.

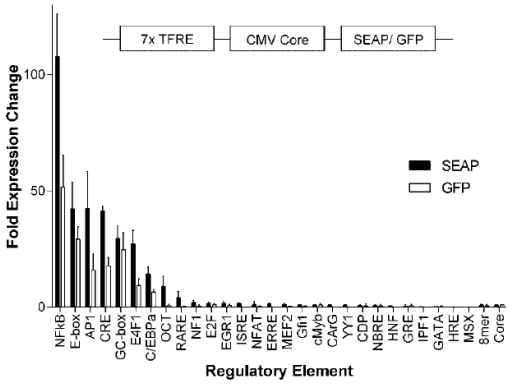


Figure 1. Identification of active transcription factor regulatory elements in CHO-S cells. Seven copies of each TFRE (as described in Table I) were cloned in series upstream of a minimal CMV core promoter in reporter vectors encoding either GFP or SEAP reporters. CHO-S cells (2×10^5) in 24-well plates were transfected with 1 μg of SEAP (black bars) or GFP (white bars) TFRE reporter-vector. SEAP activity in cell culture supernatant and intracellular GFP were measured 24 h post-transfection. Data are expressed as a fold-change with respect to the activity of a vector containing only a minimal CMV core promoter (Core). A random 8 bp sequence with no known homology to TFRE sequences (8mer) was also used as a control. Bars represent the mean + SD of three independent experiments each performed in triplicate, using three clonally derived plasmids for each TFRE-reporter construct.

in ligation mixes at a 1:20 molar ratio of the RE blocks. The ligated molecules were digested with *Kpn*I and *Xho*I (Promega, Southampton, UK), gel extracted (Qiaquick gel extraction kit, Qiagen), and inserted upstream of the minimal CMV core promoter in the promoterless SEAP reporter vector. Clonally derived plasmids were purified and sequenced.

Cell Culture and Transfection

CHO-S and CHO-K1 cells were cultured in CD-CHO medium (Life Technologies) supplemented with 8 and 6 mM L-glutamine (Sigma), respectively. CHO-DG44 cells were cultured in CD-DG44 medium (Life Technologies) supplemented with 8 mM L-glutamine and 18 mL/L pluronic F68 (Life Technologies). All cells were routinely cultured at 37°C in 5% (v/v) CO_2 in vented Erlenmeyer flasks (Corning, UK), shaking at 140 rpm and subcultured every 3–4 days at a seeding density of 2×10^5 cells/mL. Cell concentration and viability were determined by an automated Trypan Blue exclusion assay using a Vi-Cell cell viability analyzer (Beckman–Coulter, High Wycombe, UK). Two hours prior to transfection, 2×10^5 cells from a mid-exponential phase culture were seeded into individual wells of a 24-well plate (Nunc, Stafford, UK). Cells were transfected with DNA-lipid complexes comprising DNA and Lipofectamine (Life Technologies), prepared according to the manufacturer's instructions. Transfected cells were incubated for 24 h prior to protein expression analysis.

To eliminate potential promoter–promoter interference, individual transfections did not include a co-transfected reporter vector to compare transfection efficiencies. However, each 24-well plate included a set of three external reporters (CMV-SEAP, SV40-SEAP and 7× CRE-SEAP) to confirm reproducible transfection performance.

Fed-Batch Transient Transfection

Two hours prior to transfection 6×10^6 cells from a mid-exponential phase CHO-S culture were seeded into 50 mL CultiFlask bioreactors (Sartorius, Surrey, UK) at a working volume of 6 mL. Cells were transfected with DNA: lipofectamine complexes, prepared according to the manufacturer's instructions. Fed-batch cultures were maintained for 7 days by nutrient supplementation with 10% (v/v) CHO CD

Efficient Feed A (Life Technologies) on Day 2, 4, and 6. SEAP expression and cell growth were measured at 24 h intervals.

Quantification of Reporter Expression

SEAP protein expression was quantified using the Sensolyte pNPP SEAP colorimetric reporter gene assay kit (Cambridge Biosciences, Cambridge, UK) according to the manufacturer's instructions. GFP protein expression was quantified using a Flouroskan Ascent FL Flourometer (Excitation filter: 485 nm, Emission filter: 520 nm). Background fluorescence/absorbance was determined in cells transfected with a promoterless vector.

Results

Identification of Active Transcription Factor Regulatory Elements in CHO-S Cells

In order to identify discrete TFREs (transcription factor binding sites) capable of recombinant gene transactivation in CHO-S cells we surveyed, in silico, putative TFREs in 10 viral promoters generally known to be active in CHO cells. We rationalized that these promoters must contain one or more constituent TFREs able to bind cognate transcription factors present in CHO cells. Using online search tools that scan DNA sequences for transcription factor (TF) binding sites, specifically Transcription Element Search System (TESS) and Transcription Affinity Prediction tool (TRAP), we employed stringent search parameters (Manke et al., 2008; Schug, 2008) to minimize false positives. Across all viral promoter sequences, 67 discrete TFREs were identified as being present in one or more. To further minimize this pool (design space) we filtered out TFREs that did not occur in at least two promoters. Table I lists the final set of 28 TFREs incorporated into the functional screen. Viral promoter-specific TFRE compositions are listed in Supplementary Table SI.

To determine the relative transcriptional activity of each TFRE in CHO-S cells we created sets of both GFP and SEAP reporter constructs that each contained seven repeat copies of a specific TFRE in series, upstream of a minimal hCMV-IE1 core promoter (−34 to +48 relative to the transcriptional start site (TSS), containing a TATA box and an initiator element). We previously determined that this core promoter drives higher basal levels of expression than common alternatives (e.g., CHEF1α, SV40) in CHO cells (data not shown). There are relatively few available motifs to construct a synthetic core (Juven-Gershon and Kadonaga, 2010; Juven-Gershon et al., 2006). Further, as transcriptional activators have been shown to function specifically with different core promoters, we considered it optimal to comparatively evaluate different TFREs (and synthetic promoters) using a single fixed core (Juven-Gershon and Kadonaga, 2010).

Measurement of GFP and SEAP reporter production after transient transfection of CHO-S cells with each TFRE-reporter plasmid is shown in Figure 1. This analysis identified seven TFREs that significantly increased expression of both SEAP and GFP over basal expression from the minimal core promoter up to 100-fold (NFκB, E-box, AP1, CRE, GC-box, E4F1, C/EBPα). A further two elements (OCT and RARE) increased SEAP expression but showed no observable increase in GFP, likely due to differences in reporter measurement (e.g., intracellular turnover of GFP, extracellular accumulation of SEAP). No other TFREs mediated a significant increase in reporter expression above core control levels. These data identify a group of TFREs that can independently mediate activation of recombinant gene transcription in CHO-S cells using available transcription factor activity. The relative level of reporter expression is a function of TF relative abundance, affinity of the TF for its cognate TFRE and the discrete mechanism of transcriptional activation. Lastly, our data imply that many TFREs that exist in common viral promoters used to drive recombinant gene expression in CHO cells may be functionally redundant.

First Generation Synthetic Promoters Exhibit a Broad Activity Range Up To That of hCMV-IE1

In order to construct a first generation synthetic promoter library we utilized all TFREs identified as transcriptionally active in CHO-S cells. Oligonucleotide building blocks containing a single copy of each TFRE sequence (TFRE blocks) were chemically synthesized [NFκB, CRE, E-box, GC-box, E4F1, and C/EBPα; AP1 was omitted from the library due to previously observed functional redundancy between CRE and AP1 sites (Hai and Curran, 1991)], and ligated at an equal ratio to assemble random TFRE-combinations which were inserted upstream of the minimal CMV core promoter in SEAP reporter plasmids. A control CMV promoter reporter plasmid was constructed using the hCMV-IE1 promoter (−559 to +48 relative to the TSS, i.e., the complete hCMV-IE1 enhancer containing the distal, proximal and core promoter regions, hereafter referred to as CMV) upstream of the SEAP ORF. Purified plasmid DNA from 110 transformed *E. coli* colonies picked at random was utilized for measurement of SEAP reporter production.

Transient production was employed to determine the relative activity of synthetic promoters as it both maximizes throughput and provides a direct readout of synthetic promoter transactivation without potential interference from integration-specific effects or silencing. Whilst SEAP production is not a direct measurement of transcriptional activity, previous experiments in this laboratory have confirmed that SEAP activity in cell culture supernatant is linearly correlated with SEAP mRNA levels post-transfection. Moreover we optimized assay conditions such that control CMV-SEAP reporter activity was in the center of the linear assay range with respect to plasmid copy number (DNA load) and measured SEAP output (data not shown). SEAP production at 24 h post-transfection was measured for each synthetic promoter, and each promoter was sequenced to reveal its TFRE-block composition. A small proportion (14) of reporter plasmids were found to be lacking a promoter insert and these were excluded from further analysis. The

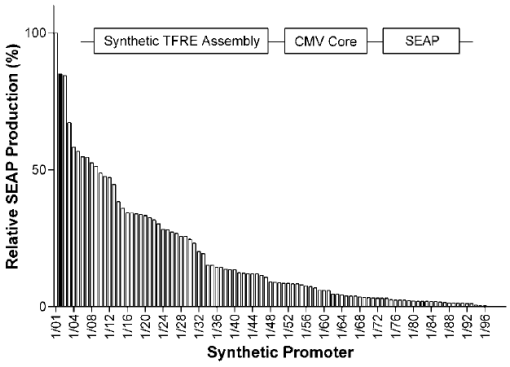


Figure 2. First generation synthetic promoters vary in activity up to that of CMV. First generation synthetic promoters were constructed by random ligation of NFκB, CRE, E-box, GC-box, E4F1, and C/EBPα TFREs in equal proportion. Synthetic promoters were inserted upstream of a minimal CMV core promoter in SEAP reporter plasmids and transfected into CHO-S cells. SEAP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by promoter 1/01 (Supplementary Table SII). SEAP production from the control CMV-SEAP reporter is shown as the black bar. Each bar represents the mean of two transfections, for each promoter less than 10% variation in SEAP production was observed.

relative transcriptional activity of the remaining 96 promoters is shown in Figure 2, and their TFRE-block compositions are listed in Supplementary Table SII. These data show that generation 1 synthetic promoter activities spanned two orders of magnitude, where the most active synthetic promoter exhibited a 1.2-fold increase in SEAP production over that deriving from the CMV control vector. Analysis of synthetic promoter composition revealed that (i) synthetic promoter length varied between 7 and 31 TFRE blocks (mean = 11.9 ± 4.2 blocks; 189 ± 66 bp), although relative transcriptional activity was unrelated to promoter length (ii) across the generation 1 library the relative abundance of the six TFRE building blocks was approximately equivalent and (iii) individual TFRE blocks could occur in either forward or reverse orientation (i.e., the consensus TF recognition sequence (see Table I) could occur on either DNA strand) but this was not apparently related to synthetic promoter activity, either with respect to the general frequency of occurrence or with respect to the relative orientation of specific TFRE blocks. Therefore, we inferred that variation in synthetic promoter activity was a consequence of the differing relative abundance of specific TFRE blocks within promoters and/or positional effects (i.e., that specific neighboring or distal combinations of TFRE blocks may affect promoter strength). Whilst the latter is computationally intractable given the size of the library, we addressed the former by determination of the relative frequency with which individual TFRE blocks occurred within synthetic promoters of varying activity. These data are shown in Figure 3. Whilst no single TFRE block exhibited an obviously

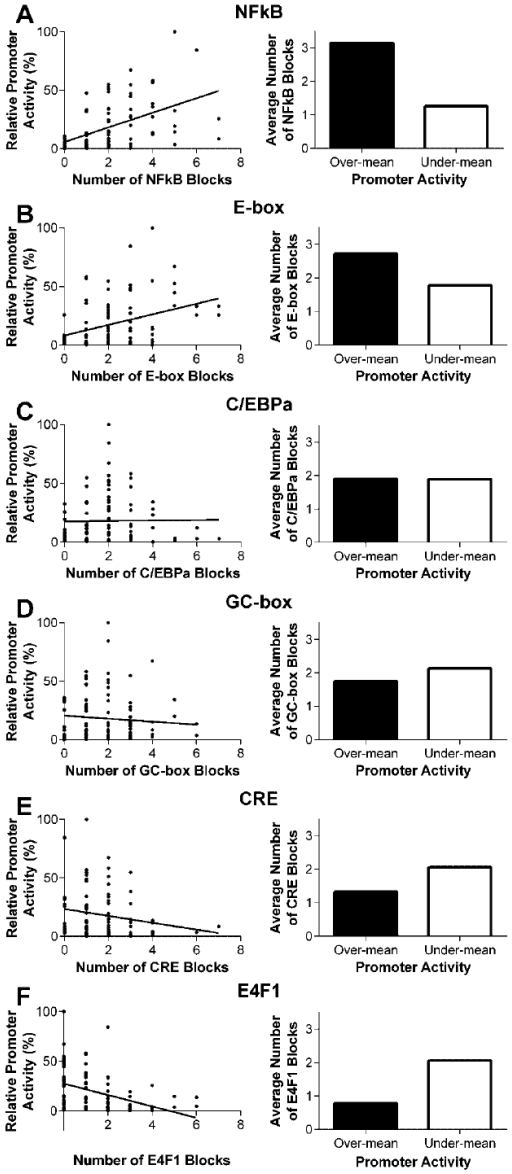


Figure 3. Relative abundance of transcription factor regulatory elements in first generation synthetic promoters. First generation synthetic promoters were sequenced to enable assignment of TFRE block composition (listed in Supplementary Table SII). The number of each TFRE block in each synthetic promoter is plotted against the relative activity of that promoter (**A**–**F**). In each case the linear regression line is shown, where the slope of the line indicates the extent to which each TFRE occurs in promoters of varying activity. The mean number of each TFRE block in higher or lower activity promoters (over or under mean first generation promoter activity) is indicated in each case.

dominant influence over synthetic promoter strength, individual TFRE blocks were either relatively abundant in active promoters (NFκB, E-box), equally distributed across promoters (C/EBPα, GC-box) or relatively abundant in low activity promoters (E4F1, CRE). This bias was confirmed by multiple linear regression analysis, where either an all factor model (inclusion of all six TFREs, $r^2 = 0.57$, $P = 1.7 \times 10^{-14}$) or a parsimonious model excluding C/EBPα and GC-box TFREs (as these do not improve model fit; $r^2 = 0.56$, $P = 8.84 \times 10^{-16}$) predicted the optimal stoichiometry of TFRE blocks to be NFκB 1.58: E-box 1. The other TFRE blocks were either neutral (C/EBPα, GC-box) or negative effectors (E4F1, CRE). Analysis of specific promoter sequences throughout the library confirmed site designations as positive, neutral or negative. For example, the strongest promoter (1/01) contains the highest ratio of positive (NFκB, E-box): negative (E4F1, CRE) sites (9: 1) in the library. Moreover, the three most active promoters (1/01–1/03) are the only promoters in the library containing more than seven positive sites and less than three negative sites. There are also multiple examples where high numbers of positive sites are apparently counteracted by high numbers of negative sites to produce relatively weak promoters, for example, promoters 1/37, 1/52, and 1/68 which have positive: negative ratios of 8: 8, 8: 8, and 9: 11, respectively.

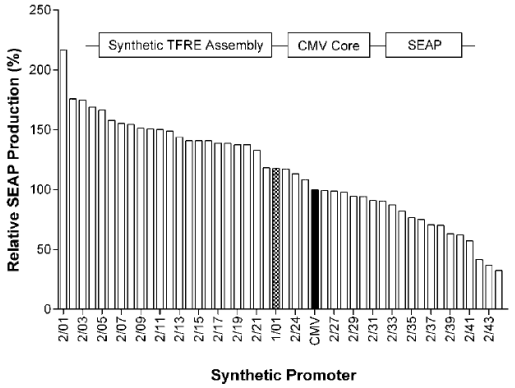


Figure 4. Generation two synthetic promoters achieve twice the activity of the CMV promoter. Second generation synthetic promoters were constructed by random ligation of NFκB, E-box, GC-box, C/EBPα TFREs in the ratio 5:3:1:1. Synthetic promoters were inserted upstream of a minimal CMV core promoter in SEAP reporter plasmids and transfected into CHO-S cells. SEAP expression was quantified 24 h post-transfection. Data are expressed as a percentage of the production exhibited by CMV control promoter (black bar). SEAP production from the most active promoter from the first generation library (1/01; Fig. 2) reporter is shown as a checked bar. Otherwise, each bar represents the mean of two transfections, for each promoter less than 10% variation in SEAP production was observed.

Second Generation Synthetic Promoters Achieve Twice the Activity of CMV

In order to further improve synthetic promoter activity we created a second generation library using random ligation of a mixture of TFRE blocks at an optimal ratio derived from analysis of the composition of first generation promoters. Specifically, negative TFREs were omitted (E4F1, CRE, Fig. 3), positive TFREs were included at the ratio NFκB 5: E-box 3, and neutral TFREs were included at the ratio C/EBPα 1: GC-box 1. We included the latter based on the hypothesis that increased complexity could be advantageous. For example, the three most active synthetic promoters in the first generation library all contained at least two copies of both neutral TFREs (Supplementary Table SII) and thus they could contribute to unknown positional effects. We expected that second generation promoters would contain the same average number of TFRE blocks (12) as first generation promoters.

A second generation library was created as described previously; 50 transformed *E. coli* colonies were picked at random, synthetic promoters in purified plasmid DNA were sequenced and 44 reporter plasmids containing promoter sequences were utilized for measurement of SEAP reporter production. The relative transcriptional activity of second generation promoters is shown in Figure 4, and their TFRE-block compositions are listed in Supplementary Table SIII. Second generation promoters exhibited significantly increased activity. The mean expression level (relative to CMV) shifted from 21.2% for first generation promoters to 116% for the second generation library. Twenty-five synthetic promoters (57% of the library) achieved a higher SEAP production than the CMV control, with the strongest promoter (2/01) exhibiting a 2.2-fold increase.

Analysis of the TFRE block composition of second generation promoters revealed that the relative stoichiometry of TFRE blocks across the library was approximately as designed (NFκB 5: E-box 2.81: GC-box 1.32: C/EBPα 1.15). In silico analysis utilizing TFRE-prediction tools confirmed that unexpected, additional TFRE sites had not formed at TFRE-block junctions during promoter assembly. In contrast to first generation promoters the slope of the fitted linear regression line relating total TFRE block number (synthetic promoter length) and promoter activity was slightly positive ($r^2 = 0.31$, $P = 8.9 \times 10^{-5}$), although we do not regard this as a significant factor. As shown in Figure 5, for second-generation promoters the influence of GC-box and C/EBPα is generally negative, whereas NFκB and E-box remain positive effectors. However, considering the composition of second generation promoters listed in Supplementary Table SIII, our data do not support the conclusion that either NFκB or E-box TFRE blocks could support high transcriptional activity alone—clearly a combination of both is necessary. The most powerful promoters (2/01–2/03) contain relatively high numbers of both TFREs in approximately equal proportion, with a correspondingly low number of negative GC-box and C/EBPα blocks. Some lower activity promoters do contain relatively large numbers of NFκB or E-box blocks (e.g., 2/11, 2/13, 2/17) but (i) contain a sub-

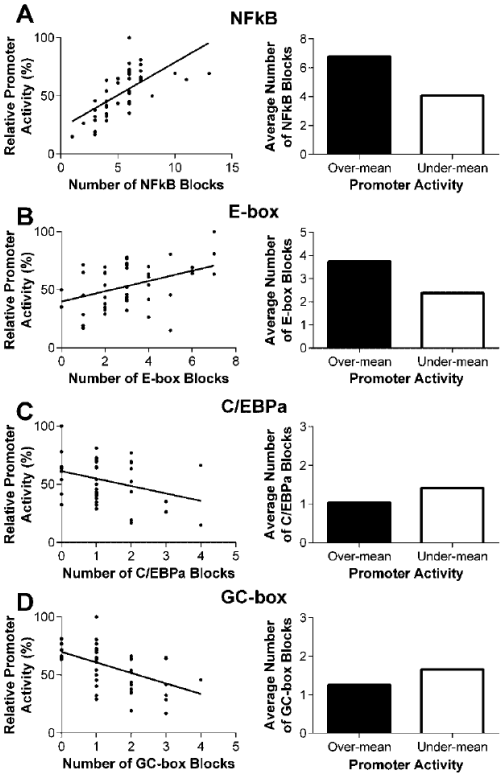


Figure 5. Relative abundance of transcription factor regulatory elements in second generation synthetic promoters. Second generation synthetic promoters were sequenced to enable assignment of TFRE block composition (listed in Supplementary Table SIII). The number of each TFRE block in each synthetic promoter are plotted against the relative activity of that promoter (**A–D**). In each case the linear regression line is shown, where the slope of the line indicates the extent to which each TFRE occurs in promoters of varying activity. The mean number of each TFRE block in higher or lower activity promoters (over or under mean second generation promoter activity) is indicated in each case.

optimal ratio of NFκB: E-box (2/11, 2/17) or (ii) also contain relatively large numbers of GC-box and C/EBPα blocks (2/13).

Synthetic Promoters Exhibit Conserved Relative Activity in Different CHO Host Cell Lines and Through a Fed-Batch Transient Production Process

In order to determine if synthetic promoters performed robustly and predictably we evaluated their relative functional capability in different CHO host lines and through a fed-batch transient production process. With respect to the former, we hypothesized that different CHO hosts may contain varying proportions of transcription factors that could markedly influence synthetic promoter function, especially as the complexity of synthetic promoters is significantly reduced compared to CMV. With respect to the latter, we hypothesized that transition of CHO cells through a production process, with the associated dynamic variation in cell physiology and function (e.g., growth rate, suspension culture) may change the relative proportion of endogenous transcription factors affecting synthetic promoter activity.

A panel of seven promoters from both first and second generation libraries were selected that cover a broad range of promoter activity (1/51 < 1/17 < 1/04 < 1/02 < 2/19 < 2/03 < 2/01). These were compared to the activity of CMV. Figure 6 shows transient SEAP production from all promoters in three commonly utilized host lines; CHO-S, CHO-DG44, and CHO-K1. The relative rank order of promoter activity is maintained in all three cell lines, with the exception that 2/03 outperforms 2/01 in CHO-K1. In contrast to the original screen, promoters 2/03 and 2/01 have approximately equivalent expression in CHO-DG44 and CHO-S. In each cell line the top performing synthetic promoter drives significantly higher SEAP production than CMV; 3.1-fold 1.9-fold, and 1.7-fold in CHO-DG44, CHO-S, and CHO-K1 cells respectively. We note that in general CHO-DG44 cells exhibited significantly less reporter production than either CHO-S or CHO-K1 cells, presumably due to their reduced "transfectability" by lipofection.

Lastly, we hypothesized that transfection of cells with synthetic promoters specific for a small number of transcription factors may function as decoys, effectively competing

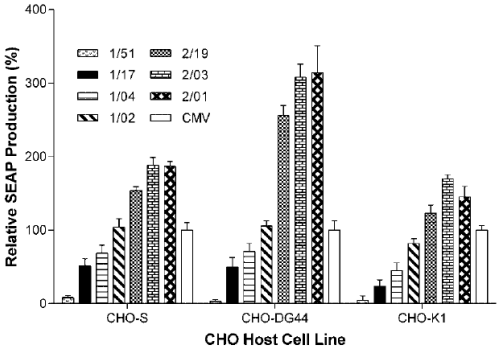


Figure 6. Synthetic promoters exhibit conserved relative activity in different CHO host cell lines. The relative activity of seven synthetic promoters with differing relative activity was determined in CHO-S cells (chequered bars), CHO-K1 cells (black bars), and CHO-DG44 cells (striped bars). Cells (2×10^5) were transfected with 250 ng SEAP-reporter vector, and SEAP production was quantified 24 h post-transfection. Data are expressed as a percentage of the activity of the CMV promoter in each cell line. Note that in general CHO-DG44 cells exhibited significantly less reporter production than either CHO-S or CHO-K1 cells (SEAP production from the CMV promoter in each cell line (100%) occurred at the ratio CHO-S 1: CHO-K1 0.81: CHO-DG44 0.32). Values represent the mean + SD of three independent experiments performed in triplicate.

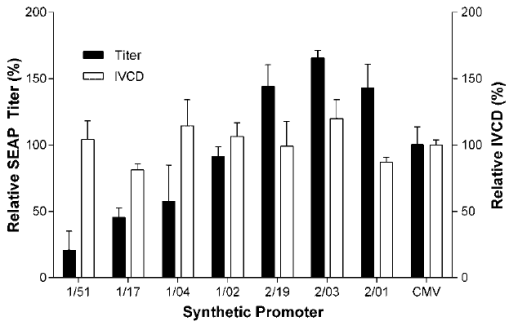


Figure 7. Synthetic promoters exhibit conserved relative activity in a fed-batch transient production process. CHO-S cells (6×10^6) were transfected with 7.5 μg of SEAP reporter-vectors, where SEAP expression was under the control of synthetic promoters with varying activity or the control CMV promoter. SEAP production and viable cell concentration were measured over the course of a 7-day fed-batch process in tube-spin bioreactors. The mean IVCD at Day 7 (white bars) and SEAP titer (black bars) are shown. SEAP data are expressed as a percentage of the control CMV promoter activity. Two independent transfections were performed in duplicate.

with the host cell genome for transactivation, thus reducing host cell proliferation. Therefore, we evaluated the same panel of promoters in a longer-term fed-batch SEAP production process over 7 days, utilizing CHO-S host cells in suspension (Fig. 7). A transient system was employed (rather than stable) to ensure production variability was directly linked to differences in promoter activity rather than cell line specific, site-specific integration or promoter silencing (e.g., methylation, deletion) artifacts. We observed that the differences in synthetic promoter activity observed in static microplates were maintained in the fed-batch transient system. The highest SEAP titer, driven by promoter 2/03, was over 1.65-fold that obtained by CMV-mediated expression. We observed that no synthetic promoter had a significant effect on the integral of viable cell density (IVCD) at the end of the transient production process, disproving our hypothesis.

Discussion

In this study we have constructed the first bespoke synthetic promoters designed specifically to function with the transcriptional machinery of CHO cell factories. Synthetic promoter activity was primarily a function of a promoter's relative composition of NFκB and E-box TFREs. TFs that bind these sites are amongst the most frequently occurring oncogenes, and their impaired regulation is a principal mechanism by which cancerous cells achieve uncontrolled proliferation and suppressed apoptosis (Dolcet et al., 2005; Lin et al., 2012). It is therefore likely that our synthetic promoters will be functional in a range of transformed mammalian cell types and may have use in cancer-targeted gene therapeutic applications. We anticipate that there are more complex design rules governing NFκB and E-box function (spatial effects, combinatorial interactions) and that many additional combinations of TFREs could be utilized to construct CHO cell synthetic promoters. Determination of the former would require construction of larger, specifically designed synthetic promoter libraries with the associated computational modeling, whilst the latter would require a more comprehensive screen of CHO cell TFRE activities. Moreover, we envisage that emerging transcriptomics data will enable identification of endogenous TFREs that are active in different CHO cell hosts, or under specific bioprocess conditions (Datta et al., 2013).

We indirectly identified TFREs that do not exhibit activity in CHO-S cells. Whilst inactivity may be explained by assay-design features (i.e., suboptimal TFRE sequences, TFs unable to drive transactivation independently) it could be evidence that their cognate TFs are not expressed in CHO cells. It would not be surprising if CHO cells expressed a particularly narrow range of transactivators given that CHO cells have lost specific functionalities over the course of their long-term (>50 years) culture in vitro (Lewis et al., 2013; Xu et al., 2011). Unrecognized TFREs could be utilized with their cognate TFs as inducible, controllable elements to construct regulated gene expression systems (switches, oscillators, synthetic regulatory cascades) for synthetic biology applications in CHO cells (Tigges and Fussenegger, 2009; Weber and Fussenegger, 2010). Many of these inactive TFREs are present in the CMV promoter, suggesting that a large proportion of the CMV sequence (and that of other viral promoters) may be functionally redundant in CHO cells. The CMV promoter has evolved to access a broad range of TFs to enable the virus to infect a wide range of host cells with varying physiology and TF activity (Stinski and Isomura, 2008). Specifically designed, rather than evolved, our synthetic promoters have far more efficient transcriptional activity per unit DNA sequence (synthetic promoter 2/01 exhibited a 2.2-fold increase in activity over CMV but is less than half the size).

Synthetic promoters that achieved higher titers than CMV in fed-batch transient SEAP production could be utilized to optimize the transient production of early stage products [i.e., development material for toxicology and clinical trials testing (Daramola et al., 2013)]. Although we have not tested our promoters in a stable expression format, and therefore cannot definitively claim that they will maintain either their relative activities or expression stability in stable transfectants, they may offer additional advantages beyond high level transcriptional activity. For example, the CMV promoter can contribute to production instability via promoter methylation and gene deletion [the latter likely via homologous recombination, i.e., "looping-out" caused by two identical CMV sequences surrounding a light chain gene copy (Kim et al., 2011)]. This homologous recombination also predisposes synthetic genetic circuits to failure (Sleight et al., 2010). Our most potent synthetic promoters do not contain CpG islands and two different synthetic promoter sequences could be utilized within constructs to minimize gene copy loss. Further, as CpG islands in CMV have been shown to interfere

with the functionality of matrix attachment regions (Girod et al., 2005), our synthetic promoters could be more compatible with existing transcription enhancing technologies such as ubiquitously acting chromatin opening elements, bacterial artificial chromosomes, and site-specific integration systems (Mader et al., 2012; Zhou et al., 2010). Moreover, they could be utilized with other expression control technologies in research applications requiring highly precise regulation. For example, they could be employed in conjunction with the TFRE-specific decoy technology we have recently invented (Brown et al., 2013), or recently described synthetic elements that control translation initiation rates (Ferreira et al., 2013). Lastly, we suspect that by utilizing specific TFRE combinations our synthetic promoters could be refined to exhibit desirable bioproduction functionalities such as increased activity at sub-physiological temperatures (Al-Fageeh et al., 2006) or during stationary phase cell growth (Prentice et al., 2007).

The synthetic promoters presented in this study enable precise control of recombinant gene expression in CHO host cells over a broad dynamic range. For easy to express proteins, where transcription rates have been shown to exert a high level of control over production (McLeod et al., 2011; O'Callaghan et al., 2010), they could be utilized to maximize recombinant gene transcription levels (e.g., by using synthetic promoter 2/01). For difficult to express proteins (e.g., bispecific antibodies, fusion proteins), where maximizing transcription is unlikely to be beneficial, they could be utilized to provide optimized protein-specific transcription activity kinetically coordinated with polypeptide-specific folding and assembly rates. An obvious potential application is in monoclonal antibody (mAb) expression where synthetic promoters of varying activity could be used to achieve mAb-specific light chain: heavy chain (LC: HC) expression ratios to optimize mAb production (Ho et al., 2013; Pybus et al., 2013). For example, the recent study in our laboratory by Pybus et al. demonstrated that three different difficult to express mAbs had three discrete optimal LC: HC expression ratios of 9: 1, 4: 1 and 2.3: 1. By utilizing synthetic promoter 2/01 (to maximize LC gene expression) in conjunction with synthetic promoters 2/25, 1/03 or 1/24, each of these ratios could be functionally achieved. Lastly, the provision of 140 discrete promoter activities, covering over two orders of magnitude, will enable CHO cell engineers to precisely engineer the cell factory, where systems level control of cell function may require the constitutive expression of several genes to be stoichiometrically balanced. By enabling sophisticated multigene engineering systems these synthetic promoters will therefore facilitate exploitation of the current revolution in CHOmics (Datta et al., 2013).

A.J.B. would like to thank the EPSRC for a DTA studentship.

References

Al-Fageeh MB, Marchant RJ, Carden MJ, Smales CM. 2006. The cold-shock response in cultured mammalian cells: Harnessing the response for the improvement of recombinant protein production. Biotechnol Bioeng 93(5):829–835.

Blazeck J, Garg R, Reed B, Alper HS. 2012. Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. Biotechnol Bioeng 109(11):2884–2895.

Brown AJ, Mainwaring DO, Sweeney B, James DC. 2013. Block decoys: Transcription factor decoys designed for in vitro gene regulation studies. Anal Biochem 443(2):205–210.

Dale L. 2006. Mammalian expression cassette engineering for high-level protein production. BioProcess Int 4:14–22.

Daramola O, Stevenson J, Dean G, Hatton D, Pettman G, Holmes W, Field R. 2013. A high-yielding CHO trasnient system: Coexpression of genes encoding EBNA-1 and GS enhances transient protein expression. Biotechnol Prog 30:132–141.

Datta P, Linhardt RJ, Sharfstein ST. 2013. An'omics approach towards CHO cell engineering. Biotechnol Bioeng 110(5):1255–1271.

Dolcet X, Llobet D, Pallares J, Matias-Guiu X. 2005. NF-kB in development and progression of human cancer. Virchows Arch 446(5):475–482.

Ferreira JP, Overton KW, Wang CL. 2013. Tuning gene expression with synthetic upstream open reading frames. Proc Natl Acad Sci USA 110(28):11284–11289.

Ferreira JP, Peacock RW, Lawhorn IE, Wang CL. 2011. Modulating ectopic gene expression levels by using retroviral vectors equipped with synthetic promoters. Syst Biol Synth Biol 5(3–4):131–138.

Girod PA, Zahn-Zabal M, Mermod N. 2005. Use of the chicken lysozyme 5′ matrix attachment region to generate high producer CHO cell lines. Biotechnol Bioeng 91(1):1–11.

Grabherr MG, Pontiller J, Mauceli E, Ernst W, Baumann M, Biagi T, Swofford R, Russell P, Zody MC, Di Palma F. 2011. Exploiting nucleotide composition to engineer promoters. PLoS ONE 6(5):e20136.

Hai T, Curran T. 1991. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc Natl Acad Sci USA 88(9):3720–3724.

Ho SC, Koh EY, van Beers M, Mueller M, Wan C, Teo G, Song Z, Tong Y, Bardor M, Yang Y. 2013. Control of IgG LC: HC ratio in stably transfected CHO cells and study of the impact on expression, aggregation, glycosylation and conformational stability. J Biotechnol 165(3–4):157–166.

Juven-Gershon T, Cheng S, Kadonaga J. 2006. Rational design of a super core promoter that enhances gene expression. Nat Methods 3(11):917–922.

Juven-Gershon T, Kadonaga JT. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 339(2):225–229.

Kim M, O'Callaghan PM, Droms KA, James DC. 2011. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol Bioeng 108(10):2434–2446.

Le H, Vishwanathan N, Kantardjieff A, Doo I, Srienc M, Zheng X, Somia N, Hu W-S. 2013. Dynamic gene expression for metabolic engineering of mammalian cells in culture. Metab Eng 20:212–220.

Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C. 2013. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nat Biotechnol 31(8):759–765.

Lin CY, Lovén J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA. 2012. Transcriptional amplification in tumor cells with elevated c-Myc. Cell 151(1):56–67.

Mader A, Prewein B, Zboray K, Casanova E, Kunert R. 2012. Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. Appl Microbiol Biotechnol 97(9):4049–4054.

Manke T, Roider HG, Vingron M. 2008. Statistical modeling of transcription factor binding affinities predicts regulatory interactions. PLoS Comput Biol 4(3):e1000039.

McLeod J, O'Callaghan PM, Pybus LP, Wilkinson SJ, Root T, Racher AJ, James DC. 2011. An empirical modeling platform to evaluate the relative control discrete CHO cell synthetic processes exert over recombinant monoclonal antibody production process titer. Biotechnol Bioeng 108(9):2193–2204.

O'Callaghan PM, McLeod J, Pybus LP, Lovelady CS, Wilkinson SJ, Racher AJ, Porter A, James DC. 2010. Cell line-specific control of recombinant

monoclonal antibody production by CHO cells. Biotechnol Bioeng 106(6):938–951.
Ogawa R, Kagiya G, Kodaki T, Fukuda S, Yamamoto K. 2007. Construction of strong mammalian promoters by random cis-acting element elongation. BioTechniques 42(5):628.
Pasotti L, Politi N, Zucca S, De Angelis MGC, Magni P. 2012. Bottom-up engineering of biological systems through standard bricks: A modularity study on basic parts and devices. PLoS ONE 7(7):e39407.
Prentice H, Tonkin C, Caamano L, Sisk W. 2007. High level expression of proteins using sequences from the ferritin heavy chain gene locus. J Biotechnol 128(1):50–60.
Pybus LP, Dean G, West NR, Smith A, Daramola O, Field R, Wilkinson SJ, James DC. 2013. Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells. Biotechnol Bioeng 111:372–385.
Schlabach MR, Hu JK, Li M, Elledge SJ. 2010. Synthetic design of strong promoters. Proc Natl Acad Sci USA 107(6):2538.
Schug J. 2008. Using TESS to predict transcription factor binding sites in DNA sequence. Curr Protoc Bioinform 21:2.6.1–2.6.15.
Seghezzi N, Amar P, Koebmann B, Jensen PR, Virolle M-J. 2011. The construction of a library of synthetic promoters revealed some specific features of strong Streptomyces promoters. Appl Microbiol Biotechnol 90(2):615–623.
Sleight SC, Bartley BA, Lieviant JA, Sauro HM. 2010. Designing and engineering evolutionary robist genetic circuits. J Biol Eng 4:12.
Stadlmayr G, Mecklenbräuker A, Rothmüller M, Maurer M, Sauer M, Mattanovich D, Gasser B. 2010. Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production. J Biotechnol 150(4):519–529.
Stinski MF, Isomura H. 2008. Role of the cytomegalovirus major immediate early enhancer in acute infection and reactivation from latency. Med Microbiol Immunol 197(2):223–231.
Tigges M, Fussenegger M. 2009. Recent advances in mammalian synthetic biology—design of synthetic transgene control networks. Curr Opin Biotechnol 20(4):449–460.
Tornøe J, Kusk P, Johansen TE, Jensen PR. 2002. Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. Gene 297(1):21–32.
Weber W, Fussenegger M. 2010. Synthetic gene networks in mammalian cells. Curr Opin Biotechnol 21(5):690–696.
Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S. 2011. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29(8):735–741.
Yim SS, An SJ, Kang M, Lee J, Jeong KJ. 2013. Isolation of fully synthetic promoters for high-level gene expression in Corynebacterium glutamicum. Biotechnol Bioeng 110(11):2959–2969.
Zhou H, Liu Z-g, Sun Z-w, Huang Y, Yu W-y. 2010. Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. J Biotechnol 147(2):122–129.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix D: Published article 2 – Brown *et al.* 2013

Analytical Biochemistry 443 (2013) 205–210

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

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

Adam J. Brown [a], David O. Mainwaring [b], Bernie Sweeney [b], David C. James [a,*]

[a] Department of Chemical and Biological Engineering, University of Sheffield, Sheffield S1 3JD, England, United Kingdom
[b] Protein Expression and Purification Group, UCB, Slough, Berkshire SL1 4EN, England, United Kingdom

ARTICLE INFO

Article history:
Received 23 June 2013
Received in revised form 21 August 2013
Accepted 3 September 2013
Available online 10 September 2013

Keywords:
Transcription factor
Decoy
Regulatory element
Binding site
Chimera
Promoter regulation

ABSTRACT

Transcription-factor decoys are short synthetic oligodeoxynucleotides that sequester cognate transcription factors and prevent their binding at target promoters. Current methods of decoy formation have primarily been optimized for potential therapeutic applications. However, they are not ideally suited to in vitro investigations into multi-transcription factor-mediated processes that may require multiple regulatory elements to be inhibited in varying combinations. In this study we describe a novel method for chimeric decoy formation in which blocks containing discrete transcription factor binding sites are combined into circular molecules. Unlike currently available methods, block decoys allow rapid construction of chimeric decoys targeting multiple regulatory elements. Further, they enable fine-tuning of binding-site copy ratios within chimeras, allowing sophisticated control of the cellular transcriptional landscape. We show that block decoys are exonuclease-resistant and specifically inhibit expression from target binding sites. The potential of block decoys to inhibit multiple elements simultaneously was demonstrated using a chimeric decoy containing molar optimized ratios of three regulatory elements, NF-κB-RE, CRE, and E-box. The chimeric decoy inhibited expression from all three elements simultaneously at equivalent levels. The primary intended use of block decoys is in vitro gene regulation studies in which bespoke chimeras can be rapidly constructed and utilized to determine a promoter's functional regulation.

Transcription factors (TFs) bind to specific DNA motifs (regulatory elements (REs)) within gene promoters and enhancers to regulate the levels of gene transcription [1]. Disruption of TF binding to target sites is accordingly a well-established method to investigate promoter regulation. An effective method to achieve this is the use of transcription-factor decoys [2–4]: short synthetic oligodeoxynucleotides (ODNs) that contain a specific TF-binding motif. When introduced into a cell the decoys compete for available TFs, preventing their association at target promoters [5]. This site-specific sequestration of TFs makes decoys an attractive method to determine the functional contribution of individual REs to a promoter's activity.

The key determinants of decoy effectiveness are stability, specificity, and uptake [6]. Multiple methods of decoy formation have been developed to improve these factors, primarily focusing on their stability against intracellular nucleases. These include chemical modifications such as the use of phosphorothioate groups [5] and circular dumbbell ODNs that have enzymatically ligated termini [7], conferring resistance to exonucleases (the primary cause of intracellular degradation [8]). Although such advancements have greatly improved decoy functionality, particularly in potential therapeutic applications [9], currently available methods are not ideally suited to in vitro gene regulation studies.

As most promoters contain binding sites for multiple TFs, gene regulation studies utilizing decoys are likely to require multiple decoys, targeting varying combinations of different binding motifs. Ideally, where multiple REs are targeted at once they would be included on a single decoy molecule to avoid the uneven distribution of various decoys across the transfected cell population. Phosphorothioate and dumbbell decoys targeting two [10,11] and three [12] REs have been described (and shown to be far superior to using individual decoys against each site) but these formation methods do not allow for the rapid creation of bespoke chimeric decoys. Further, they do not provide the capability to fine-tune the molar ratio of different sites within one molecule.

Here, we describe a method of decoy formation in which chimeric, ratio-optimized, molecules can be synthesized from a set of input RE blocks. We show that block decoys are resistant to exonuclease degradation and can specifically inhibit multiple REs simultaneously. This novel method offers significant advantages for multitarget decoy studies. Using existing methods each

* Corresponding author.
E-mail address: d.c.james@sheffield.ac.uk (D.C. James).

0003-2697/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ab.2013.09.003

chimera must be designed and synthesized individually. Block decoys allow potentially hundreds of different chimeric combinations to be quickly constructed from a pool of input RE blocks, saving time and substantial costs. The ability to tailor RE ratios enables more efficient inhibition, reducing the decoy concentration required and potentially facilitating more REs to be targeted simultaneously. Furthermore, the relative level of inhibition of each target can be altered, allowing sophisticated control of the cellular transcriptional landscape. Block decoys are therefore a new tool for investigating multiple TF-mediated phenotypes and are particularly suited to in vitro gene regulation studies.

Materials and methods

Construction of block decoys

The method of block decoy construction is shown schematically in Fig. 1. RE block molecules were developed by annealing two complementary, single-stranded 5′-phosphorylated DNA ODNs (Sigma, Poole, UK) in STE buffer (100 mM NaCl, 50 mM Tris–HCl, 1 mM EDTA, pH 7.8; Sigma). ODNs were heated at 95 °C for 5 min and then ramp cooled to 25 °C over 2 h to create RE blocks that contain a transcription factor binding site and a 4-bp TCGA single-stranded overhang at each 5′ terminus. RE blocks (12 μg) were then ligated with 5 units of high-concentration T4 DNA ligase (Life Technologies, Paisley, UK) at room temperature for 3 h to form block decoys. The cohesive ends enable RE blocks to be ligated together into extending concatamers. At sizes greater than 100 bp DNA molecules are likely to bend [13,14], allowing ligation of cohesive termini [15,16]. Therefore, ligation of input blocks theoretically results in covalently closed circular block decoys. Chimeric decoys were constructed by ligating varying molar concentrations of different RE blocks. The sequences of RE blocks employed were as follows (consensus site in italic): nuclear factor κB response element (NF-κB-RE), 5′-TCGAT*GGGACTTTCCA*-3′ and 5′-TCGAT*GGAAAGTCC*-*CA*-3′; cyclic AMP-responsive element (CRE), 5′-TCGATT*TGAC* *GTCA*TT-3′ and 5′-TCGAAA*TGACGTCA*AA-3′; enhancer box (E-box), 5′-TCGAAA*CACGTG*AGA-3′ and 5′-TCGATCT*CACGTG*TT-3′. Scrambled decoys contained the following scrambled consensus sites: NF-κB-RE, AATCGCAAGT; CRE, GACTAGAG; E-box, GCTCAG. All block decoys were extracted and stored at 350 ng/μl.

Analysis of block decoy structure

Block decoy population size distribution was analyzed by ethidium bromide agarose gel electrophoresis utilizing molecular weight markers (Hyperladder II; Bioline, London, UK). To confirm block decoy circularization, 1.5 μg of purified block decoy was added to 5 units of high-concentration T4 DNA ligase before gel analysis. To test the stability of block decoys against exonuclease, 4 μg of block decoy was incubated with 300 units of exonuclease III (Promega, Southampton, UK) and the mixture was incubated at 37 °C. A mixture of linear ODNs spanning the molecular weight range of the block decoys was used as a positive control.

Construction of RE-specific reporter vectors

A promoterless vector was subcloned from pSEAP2control (Clontech, Oxford, UK) by PCR amplification of appropriate vector regions with Phusion high-fidelity polymerase (New England Biolabs, Hitchin, UK). A minimal core promoter from the human cytomegalovirus (CMV) was synthesized (Sigma) and cloned into the XhoI and EcoRI sites directly upstream of the secreted alkaline phosphatase (SEAP) open reading frame (ORF). The core promoter sequence used was as follows: 5′-AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCCACGCT GTTTTGACCTCCATAGAAGAC-3′. A second reporter plasmid was created by replacing the SEAP ORF with the Turbo green fluorescent protein (GFP) ORF. To create binding-site reporter plasmids, synthetic oligonucleotides containing 7× repeat copies of NF-κB-RE, CRE, and E-box were synthesized (Sigma), PCR amplified, and inserted into KpnI and XhoI sites upstream of the CMV core pro-

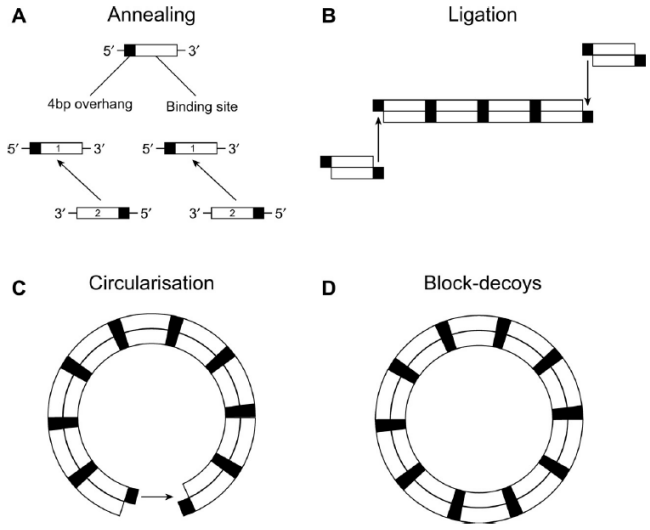


Fig.1. Schematic of block decoy formation. (A) Single-stranded oligonucleotides are annealed to form regulatory element blocks containing a transcription factor binding site and a 4-bp single-stranded overhang at the 5′ terminus. (B) Regulatory element blocks are ligated together into extending concatamers, (C) which circularize, allowing intramolecular ligation of cohesive termini, yielding (D) covalently closed circular block decoys containing multiple copies of the target binding site.

moter. The sequence of all plasmid constructs was confirmed by DNA sequencing.

Cell culture and transfection

CHO-S cells, a suspension-adapted variant of CHO-K1, were obtained from Life Technologies. CHO-S cells were routinely cultured in CD-CHO medium (Life Technologies) supplemented with 8 mM L-glutamine (Sigma) at 37 °C in 5% (v/v) CO_2 in vented Erlenmeyer flasks (Corning Glassware, Corning, NY, USA), shaking at 140 rpm. Cells were subcultured every 3 or 4 days at a seeding density of 2×10^5 cells/ml. Cell concentration and viability were determined by an automated trypan blue exclusion assay using a Vi-Cell cell viability analyzer (Beckman-Coulter, High Wycombe, UK). Two hours prior to transfection, 2×10^5 cells from a mid-exponential-phase culture were seeded into individual wells of a 24-well plate (Nunc, UK). Cells were transfected with DNA–lipid complexes comprising 1 μg DNA per 3 μl Lipofectamine (Life Technologies), prepared according to the manufacturer's instructions. Transfected cells were incubated for 24 h prior to protein expression analysis. All transfections were carried out in triplicate and experiments repeated three times.

Quantification of reporter expression

SEAP protein expression was quantified using the Sensolyte pNPP SEAP colorimetric reporter gene assay kit (Cambridge Biosciences, Cambridge, UK) according to the manufacturer's instructions. GFP protein expression was quantified using a Fluoroskan Ascent FL fluorimeter (excitation filter 485 nm, emission filter 520 nm). Background fluorescence/absorbance was determined in cells transfected with a promoterless vector.

Results and discussion

Block decoy formation and stability

Block decoy formation (Fig. 1) was confirmed by gel electrophoresis. This analysis showed that different RE-specific decoys constructed using the appropriate RE blocks exhibited near-identical size distributions, with the vast majority of molecules between 100 and 300 bp in size (Fig. 2). To test the hypothesis that circularization of decoys prevented further ligation (thus limiting their size) we (i) utilized purified block decoys as the substrate in further ligation reactions and (ii) evaluated block decoy stability against digestion by exonuclease III active against linear DNA. In both cases, no variation in block decoy size distribution was observed (Fig. 2). We conclude that this method of block decoy construction yields circular ODNs containing approximately 7–20 copies of a target TF binding site.

Block decoy function and specificity

The use of block decoys to inhibit the activity of specific REs was evaluated using GFP and SEAP reporter plasmids containing five to seven copies of a discrete RE motif upstream of a core human CMV promoter. Preliminary experiments showed that minimal reporter expression was observed with the core promoter alone (1–3% of reporter activity of RE-containing plasmids; data not shown).

Three reporter plasmids, each utilizing a specific RE (NF-κB-RE, CRE, or E-box) to drive reporter expression, were used to validate the specific inhibitory effects of block decoys in vitro. In each case, RE-specific reporter expression was inhibited only by the corresponding block decoy. For example, as shown in Fig. 3A, the NF-κB-RE block decoy inhibited expression from NF-κB-RE–GFP reporter plasmid in a dose-dependent manner. Moreover, the concentration of NF-κB-RE block decoy exhibiting maximal inhibition of NF-κB-RE–GFP reporter expression (2 μg/ml) had no significant effect on GFP expression from either CRE or E-box RE reporter plasmids (Fig. 3B). Last, block decoys constructed from E-box and CRE RE blocks did not significantly affect expression from the NF-κB-RE–GFP reporter plasmid (Fig. 3C). All block decoys exhibited similar RE-specific inhibition (data not shown), and we conclude that each block decoy functions to specifically sequester its cognate RE-binding transcription factors, inhibiting expression from promoters dependent on their activity.

Chimeric block decoys

A major advantage of the block decoy strategy is that it can be utilized to construct stoichiometrically optimized chimeric decoys targeting multiple REs. To test the hypothesis that chimeric block decoys can be constructed with a controlled ratio of RE blocks we (i) ligated six distinct RE blocks in an equimolar ratio and (ii) ligated four RE blocks in a 1.0:0.6:0.2:0.2 ratio. A "cloning block" containing XhoI and KpnI sites was added to each ligation mix to enable subsequent linearization. Linearized molecules were cloned into an acceptor plasmid and 50 clonally derived plasmids from each library were sequenced. The actual stoichiometric ratios of RE blocks across the block decoy libraries was approximately equal to the input ratios (1.18:1.09:1.03:1.00:0.96:0.86 and 1.00:0.57:0.25:0.22). We conclude that chimeric block decoys containing stoichiometrically tailored quantities of RE blocks can be constructed by controlling the molar ratio of RE blocks in the ligation reaction and therefore that cells transfected in vitro with chimeric molecules will contain RE blocks at the specified

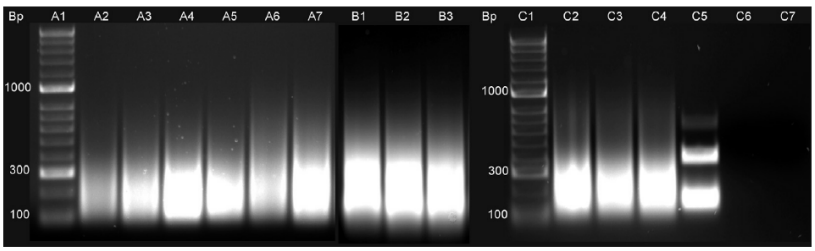


Fig.2. Circular block decoys contain numerous regulatory element binding sites. Agarose gel analysis of block decoys constructed from NF-κB-RE (lanes A2 and A3), E-box (lanes A4 and A5), and CRE (lanes A6 and A7) regulatory element binding site blocks. Circularization of a purified block decoy (lane B1) was demonstrated by (i) two further sequential ligation reactions (lanes B2 and B3), which showed no additional increase in decoy size distribution, and (ii) stability on digestion with exonuclease III for 0, 1, and 6 h at 37 °C (lanes C2–C4, respectively). Lanes C5, C6, and C7 show digestion of linear DNA sampled at the same time points.

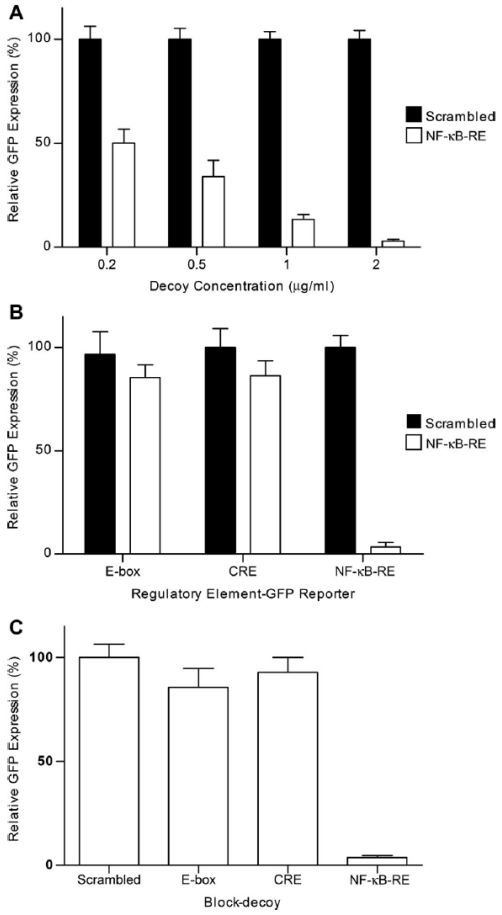


Fig.3. Circular block decoys specifically inhibit expression mediated by discrete regulatory elements. (A) CHO-S cells (2×10^5) were cotransfected with a NF-κB-RE-dependent GFP reporter plasmid with either a NF-κB-RE block decoy or a scrambled NF-κB-RE block decoy at concentrations of 0.2–2 μg/ml DNA per transfection. Data were normalized with respect to GFP expression in the presence of the scrambled decoy in each case. (B) Cotransfection of NF-κB-RE block decoy or scrambled NF-κB-RE block decoy and GFP reporter plasmids varying in transcription factor specificity (CRE, E-box, or NF-κB-RE), illustrating specific inhibition of NF-κB-RE-mediated reporter expression. (C) Cotransfection of NF-κB-RE-dependent GFP reporter with scrambled NF-κB-RE block decoy (control) or various regulatory element block decoys illustrating specific inhibition with NF-κB-RE block decoy. In (A), (B), and (C) each bar represents the mean + SEM of three independent experiments each performed in triplicate.

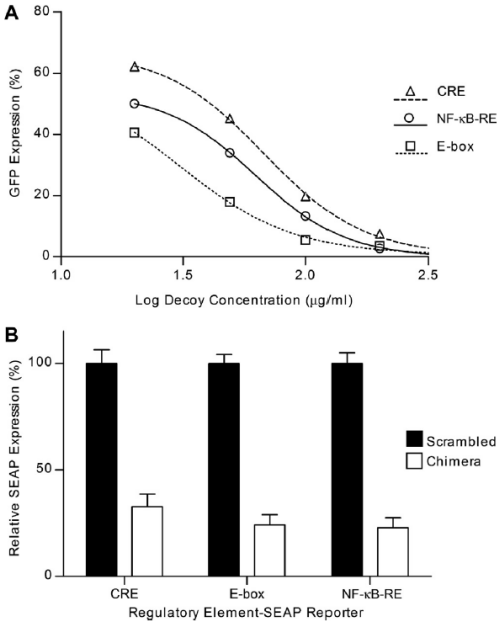


Fig.4. Stoichiometric optimization of chimeric block decoys targeting three regulatory elements. To determine the correct stoichiometry of various RE blocks in chimeric decoys to achieve equivalent inhibition of each regulatory element, the relative ability of separate RE blocks to inhibit RE-specific reporter expression was first quantified. (A) Separate cotransfection of CHO-S cells with three different regulatory element-specific block decoys (CRE, NF-κB-RE, and E-box) or the corresponding scrambled block decoy controls at varying block decoy concentrations with the corresponding RE-specific GFP-reporter plasmids (at a ratio of decoy:reporter plasmid maintained at 1:1). GFP expression in block decoy-transfected cells is shown as a percentage of reporter expression in cells transfected with the same concentration of scrambled decoy control. Best fit curves obtained by nonlinear regression analysis were utilized to determine the relative ratio of RE-specific blocks employed to construct chimeric decoys. (B) Chimeric block decoy function was evaluated by cotransfection of CHO-S cells with 3.5 μg/ml chimeric block decoys and 2 μg/ml of CRE, E-box, or NF-κB-RE SEAP-reporter plasmid. Chimeric decoys were constructed using a stoichiometric ratio of RE blocks in the ratio CRE 1.0:NF-κB-RE 0.8:E-box 0.5 (control scrambled chimeric decoy contained the same ratio of scrambled RE blocks). Each bar shows SEAP expression in chimeric decoy-treated cells relative to expression with the same concentration of scrambled decoy. Values represent the means + SEM of three independent experiments performed in triplicate.

concentrations across the cell population. To demonstrate this novel capability we constructed a chimeric block decoy targeting all three REs: NF-κB-RE, CRE, and E-box.

To determine the optimal ratio of RE blocks to construct a chimeric block decoy exhibiting maximal, equivalent inhibition of all RE–reporter plasmids we adjusted the stoichiometry of RE blocks in the ligation reaction according to the extent individual RE-specific block decoys inhibited expression of the cognate RE reporter (i.e., the relative "potency" of each RE block). We assume that the relative extent to which each RE-specific block decoy inhibits reporter expression from its corresponding RE–reporter plasmid is a function of block decoy-specific differences in (i) the relative intracellular abundance of TFs and (ii) TF-RE block binding kinetics. As shown in Fig. 4A each RE-specific block decoy exhibited a characteristic inhibitory dose–response relationship, in which, at the highest concentrations, expression from each corresponding RE-specific reporter was inhibited over 90%. Log transformation of block decoy concentration data and nonlinear regression analysis enabled determination of the relative potency of each block decoy and revealed that their inhibitory potency occurred in the order E-box > NF-κB-RE > CRE, with a stoichiometry of E-box 0.5:NF-κB-RE 0.8:CRE 1.0 (calculated by interpolation to determine relative inhibitory concentrations). Thus, to achieve concurrent inhibition of all REs to a similar extent using the block decoy approach we ligated RE blocks in this stoichiometric molar ratio.

Anticipating that chimeric decoys would require a greater concentration of decoy to be transfected to achieve a specific reduction in RE–reporter expression (as the number of copies of each RE

block is effectively reduced with an increase in the number of different RE blocks utilized to construct a chimeric decoy), we (i) increased total RE-decoy DNA load per transfection and (ii) utilized alternative RE–SEAP reporter constructs to enable more sensitive detection of RE-driven reporter expression. Preliminary experiments showed that a decoy concentration of 3.5 μg/ml decoy was the maximal decoy load that could be cotransfected with RE–reporter plasmid while still maintaining quantitation in the linear range from each RE-specific SEAP reporter plasmid (transfected at 2 μg/ml) (data not shown). Chimeric decoys were therefore transfected at this concentration, equating to RE block concentrations of 0.76 (E-box), 1.22 (NF-κB-RE), and 1.52 μg/ml (CRE). Through interpolation of the single-decoy data summarized in Fig. 4A we predicted that under these conditions chimeric decoys would inhibit expression from E-box, NF-κB, and CRE–SEAP reporter plasmids by 88%.

Fig. 4B shows that the chimeric decoy significantly inhibited expression from all three REs at approximately equivalent levels. E-box-, NF-κB-RE-, and CRE-dependent SEAP expression was reduced to 77, 76, and 68%, respectively, showing that the chimeric decoy simultaneously sequestered a substantial proportion of the intracellular cognate TFs that bind to each of the three REs. The slight reduction in decoy potency compared to predicted values may be explained by (i) the reduced transfection efficiency associated with transfecting a higher concentration of DNA (resulting in fewer copies of each RE block per cell) or (ii) TF-binding dynamics being affected by the presence of multiple RE blocks. Nonetheless, the results show that three transcription factor-binding pathways can be inhibited simultaneously using a chimeric block decoy containing stoichiometrically tailored quantities of each RE block. This is the first time that a transcription factor decoy has been shown to target multiple elements by using an optimized number of copies of each binding site.

It is likely that greater concentrations of chimeric decoy would have increased inhibition of each target element. In these experiments the concentration of block decoy was limited by cotransfection with binding-site reporters. Promoters investigated in vitro are commonly either endogenous or significantly stronger than single RE promoters. In these cases higher decoy concentrations could be employed. Nonetheless, when DNA load is a restricting factor in vitro, block decoys offer a significant advantage for concurrent inhibition of multiple REs. Fine-tuning of binding-site copy ratios reduces the final decoy concentration required. This potentially enables a greater number of elements to be targeted simultaneously compared to existing methods. Adjustable control of RE block ratios enables the optimal inhibition of each target element at any decoy concentration.

The results show that block decoys are a powerful tool for inhibiting multiple REs in vitro. The method's primary advantages are the ability to rapidly construct chimeric molecules and to control their binding-site ratios. However, block decoys have other potential benefits. Circular DNA has been associated with improved transfection efficiencies, compared to linear ODNs [17,18]. Further, multiple copies of the same binding site within a single decoy molecule may enhance TF sequestration [19,20]. It was previously shown that decoys with three site copies achieved stronger inhibition than those containing a single site [12]. Therefore, the 7–20+ binding sites per block decoy may enhance decoy function and efficiency.

If a promoter of interest contains 8 unique binding sites, there are over 200 possible unique chimeric site combinations. Using the block decoy methodology any of these could be constructed rapidly following the initial synthesis of eight RE blocks. This compares to existing methods, by which all chimeras would have to be synthesized independently. By utilizing block decoys the binding-site ratios within each chimera could be adjusted to precisely control the relative extent to which each RE is inhibited in each transfected cell. This is a major advantage over the use of mixtures of single decoys, whose relative distribution within transfected cells is unpredictable. Current methods do not allow this sophisticated control of TF activity. Block decoys therefore offer significant potential benefits for in vitro gene regulation studies.

Conclusion

Block decoys are a novel method of transcription factor decoy formation. The method described enables construction of chimeric decoys containing stoichiometrically optimized ratios of input RE blocks. We demonstrated that block decoys are able to inhibit expression from multiple target elements simultaneously using a bespoke chimeric ODN. Block decoys have significant advantages over existing decoy methods for studies requiring the simultaneous inhibition of multiple elements in defined combinations. Block decoys could be applied to the investigation of any multi-transcription factor-mediated cell function or phenotype.

Acknowledgment

A.J.B. thanks the EPSRC for a DTA studentship.

References

[1] N.J. Fuda, M.B. Ardehali, J.T. Lis, Defining mechanisms that regulate RNA polymerase II transcription in vivo, Nature 461 (2009) 186–192.
[2] S. Tomita, N. Tomita, T. Yamada, L. Zhang, Y. Kaneda, R. Morishita, T. Ogihara, V.J. Dzau, M. Horiuchi, Transcription factor decoy to study the molecular mechanism of negative regulation of renin gene expression in the liver in vivo, Circ. Res. 84 (1999) 1059–1066.
[3] V. Bezzerri, M. Borgatti, A. Finotti, A. Tamanini, R. Gambari, G. Cabrini, Mapping the transcriptional machinery of the IL-8 gene in human bronchial epithelial cells, J. Immunol. 187 (2011) 6069–6081.
[4] E. Renard, B. Porée, C. Chadjichristos, M. Kypriotou, L. Maneix, N. Bigot, F. Legendre, D. Ollitrault, B. De Crombrugghe, F. Malléin-Gérin, Sox9/Sox6 and Sp1 are involved in the insulin-like growth factor-I-mediated upregulation of human type II collagen gene expression in articular chondrocytes, J. Mol. Med. 90 (2012) 649–666.
[5] A. Bielinska, R.A. Shivdasani, L. Zhang, G.J. Nabel, Regulation of gene expression with double-stranded phosphorothioate oligonucleotides, Science 250 (1990) 997–1000.
[6] M.K. Osako, H. Nakagami, R. Morishita, Modification of decoy oligodeoxynucleotides to achieve stability and therapeutic efficacy, Curr. Top. Med. Chem. 12 (2012) 1603–1607.
[7] M.K. Osako, N. Tomita, H. Nakagami, Y. Kunugiza, M. Yoshino, K. Yuyama, T. Tomita, H. Yoshikawa, T. Ogihara, R. Morishita, Increase in nuclease resistance and incorporation of NF-κB decoy oligodeoxynucleotides by modification of the 3′-terminus, J. Gene Med. 9 (2007) 812–819.
[8] H.B. Gamper, M.W. Reed, T. Cox, J.S. Virosco, A.D. Adams, A.A. Gall, J.K. Scholler, R.B. Meyer, Facile preparation of nuclease resistant 3′ modified oligodeoxynucleotides, Nucleic Acids Res. 21 (1993) 145–150.
[9] R. Gambari, Recent patents on therapeutic applications of the transcription factor decoy approach, Expert Opin. Ther. Pat. 21 (2011) 1755–1771.
[10] T. Miyake, M. Aoki, H. Nakashima, T. Kawasaki, M. Oishi, K. Kataoka, K. Tanemoto, T. Ogihara, Y. Kaneda, R. Morishita, Prevention of abdominal aortic aneurysms by simultaneous inhibition of NFκB and ets using chimeric decoy oligonucleotides in a rabbit model, Gene Ther. 13 (2006) 695–704.
[11] W.R. Lee, K.H. Kim, H.J. An, Y.Y. Park, K.S. Kim, C.K. Lee, B.K. Min, K.K. Park, Effects of chimeric decoy oligodeoxynucleotide in the regulation of transcription factors NF-κB and Sp1 in an animal model of atherosclerosis, Basic Clin. Pharmacol. Toxicol. 112 (2012) 236–243.
[12] H. Gao, J. Xiao, Q. Sun, H. Lin, Y. Bai, L. Yang, B. Yang, H. Wang, Z. Wang, A single decoy oligodeoxynucleotide targeting multiple oncoproteins produces strong anticancer effects, Mol. Pharmacol. 70 (2006) 1621–1629.
[13] Y. Zhang, D.M. Crothers, Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization, Biophys. J. 84 (2003) 136–153.
[14] L. Ulanovsky, M. Bodner, E.N. Trifonov, M. Choder, Curved DNA: design, synthesis, and circularization, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 862–866.
[15] D. Shore, J. Langowski, R.L. Baldwin, DNA flexibility studied by covalent closure of short fragments into circles, Proc. Natl. Acad. Sci. U.S.A. 78 (1981) 4833–4837.
[16] B. Révet, A. Fourcade, Short unligated sticky ends enable the observation of circularised DNA by atomic force and electron microscopies, Nucleic Acids Res. 26 (1998) 2092–2097.

[17] A. Dhanoya, B.M. Chain, E. Keshavarz-Moore, The impact of DNA topology on polyplex uptake and transfection efficiency in mammalian cells, J. Biotechnol. 155 (2011) 377–386.
[18] P. Chancham, J.A. Hughes, Relationship between plasmid DNA topological forms and in vitro transfection, J. Liposome Res. 11 (2001) 139–152.
[19] V. Gotea, A. Visel, J.M. Westlund, M.A. Nobrega, L.A. Pennacchio, I. Ovcharenko, Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers, Genome Res. 20 (2010) 565.
[20] T.H. Lee, N. Maheshri, A regulatory role for repeated decoy transcription factor binding sites in target gene expression, Mol. Syst. Biol. 8 (2012) 576.

References

Aggarwal SR. 2012. What's fueling the biotech engine - 2011 to 2012. Nature biotechnology 30(12):1191-1197.

Agrawal V, Bal M. 2012. Strategies for rapid production of therapeutic proteins in mammalian cells. BioProcess international 10(4):32-48.

Al-Fageeh MB, Marchant RJ, Carden MJ, Smales CM. 2006. The cold-shock response in cultured mammalian cells: Harnessing the response for the improvement of recombinant protein production. Biotechnology and bioengineering 93(5):829-835.

Baldi L, Hacker DL, Adam M, Wurm FM. 2007. Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. Biotechnology letters 29(5):677-684.

Bandaranayake AD, Almo SC. 2013. Recent advances in mammalian protein production. FEBS letters 588(2):253-260

Bauer AP, Leikam D, Krinner S, Notka F, Ludwig C, Längst G, Wagner R. 2010. The impact of intragenic CpG content on gene expression. Nucleic acids research 38(12):3891-3908.

Baycin-Hizal D, Tabb DL, Chaerkady R, Chen L, Lewis NE, Nagarajan H, Sarkaria V, Kumar A, Wolozny D, Colao J. 2012. Proteomic analysis of Chinese hamster ovary cells. Journal of proteome research 11(11):5265-5276.

Berlec A, Štrukelj B. 2013. Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells. Journal of industrial microbiology & biotechnology 40(3-4):257-274.

Bezzerri V, Borgatti M, Finotti A, Tamanini A, Gambari R, Cabrini G. 2011. Mapping the transcriptional machinery of the IL-8 gene in human bronchial epithelial cells. The Journal of Immunology 187(11):6069-6081.

Bicknell AA, Cenik C, Chua HN, Roth FP, Moore MJ. 2012. Introns in UTRs: why we should stop ignoring them. Bioessays 34(12):1025-1034.

Bielinska A, Shivdasani RA, Zhang L, Nabel GJ. 1990. Regulation of gene expression with double-stranded phosphorothioate oligonucleotides. Science 250(4983):997-1000.

Birch J, Racher A. 2006. Antibody production. Advanced drug delivery reviews 58(5-6):671-685.

Black CB, Duensing TD, Trinkle LS, Dunlay RT. 2011. Cell-based screening using high-throughput flow cytometry. Assay and drug development technologies 9(1):13-20.

Blazeck J, Garg R, Reed B, Alper HS. 2012. Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. Biotechnology and bioengineering 109(11):2884-2895.

Bosch D, Castilho A, Loos A, Schots A, Steinkellner H. 2013. N-Glycosylation of plant-produced recombinant proteins. Current pharmaceutical design 19(31):5503-5512.

Brown AJ, Mainwaring DO, Sweeney B, James DC. 2013. Block decoys: Transcription-factor decoys designed for in vitro gene regulation studies. Analytical biochemistry 443(2):205-210.

Brown AJ, Sweeney B, Mainwaring DO, James DC. 2014. Synthetic promoters for CHO cell engineering. Biotechnology and bioengineering doi: 10.1002/bit.25227.

Butler M. 2005. Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. Applied microbiology and biotechnology 68(3):283-291.

Cain K, Peters S, Hailu H, Sweeney B, Stephens P, Heads J, Sarkar K, Ventom A, Page C, Dickson A. 2013. A CHO cell line engineered to express XBP1 and ERO1-Lα has increased levels of transient protein expression. Biotechnology progress 29(3):697-706.

Calo-Fernández B, Martínez-Hurtado JL. 2012. Biosimilars: company strategies to capture value from the biologics market. Pharmaceuticals 5(12):1393-1408.

Caposio P, Luganini A, Bronzini M, Landolfo S, Gribaudo G. 2010. The Elk-1 and serum response factor binding sites in the major immediate-early promoter of human cytomegalovirus are required for efficient viral replication in quiescent cells and compensate for inactivation of the NF-κB sites in proliferating cells. Journal of virology 84(9):4481-4493.

Chancham P, Hughes JA. 2001. Relationship between plasmid DNA topological forms and in vitro transfection. Journal of Liposome Research 11(2-3):139-152.

Chao SH, Harada JN, Hyndman F, Gao X, Nelson CG, Chanda SK, Caldwell JS. 2004. PDX1, a cellular homeoprotein, binds to and regulates the activity of human cytomegalovirus immediate early promoter. Journal of biological chemistry 279(16):16111-16120.

Chen C-H, Liu Y-K, Lin Y-L, Chuang H-Y, Hsu W-T, Chiu Y-H, Cheng T-L, Liao K-W. 2012. A rapid and convenient method to enhance transgenic expression in target cells. Preparative Biochemistry and biotechnology 42(5):448-461.

Chen J, Haverty J, Deng L, Li G, Qiu P, Liu Z, Shi S. 2013. Identification of a novel endogenous regulatory element in Chinese hamster ovary cells by promoter trap. Journal of biotechnology 167(3):255-261

Chia YL, Ng CH, Lashmit P, Chu KL, Lew QJ, Ho JP, Lim HL, Nissom PM, Stinski MF, Chao SH. 2014. Inhibition of human cytomegalovirus replication by overexpression of CREB1. Antiviral research 102:11-22.

Choi KH, Basma H, Singh J, Cheng PW. 2005. Activation of CMV promoter-controlled glycosyltransferase and β-galactosidase glycogenes by butyrate, tricostatin A, and 5-Aza-2′-deoxycytidine. Glycoconjugate journal 22(1-2):63-69.

Chusainow J, Yang Y, Yeo J, Toh P, Asvadi P, Wong N, Yap M. 2009. A study of monoclonal antibody producing CHO cell lines: What makes a stable high producer? Biotechnology and bioengineering 102(4):1182-1196.

Cockett MI, Bebbington CR, Yarranton GT. 1991. The use of engineered E1A genes to transactivate the hCMV-MIE promoter in permanent CHO cell lines. Nucleic acids research 19(2):319-325.

Coulon A, Chow CC, Singer RH, Larson DR. 2013. Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nature reviews genetics 14(8):572-584.

Cuevas-Bennett C, Shenk T. 2008. Dynamic histone H3 acetylation and methylation at human cytomegalovirus promoters during replication in fibroblasts. Journal of virology 82(19):9525-9536.

Czajkowsky DM, Hu J, Shao Z, Pleass RJ. 2012. Fc-fusion proteins: new developments and future perspectives. EMBO molecular medicine 4(10):1015-1028.

Dale L. 2006. Mammalian expression cassette engineering for high-level protein production. BioProcess International 4:14-22.

Daramola O, Stevenson J, Dean G, Hatton D, Pettman G, Holmes W, Field R. 2013. A high-yielding CHO transient system: Coexpression of genes encoding EBNA-1 and GS enhances transient protein expression. Biotechnology progress 30(1):132-141.

Datta P, Linhardt RJ, Sharfstein ST. 2013. An 'omics approach towards CHO cell engineering. Biotechnology and bioengineering 110(5):1255-1271.

Davis J. 2007. Systems for cell culture scale-up. In: Stacey G, editor. Medicines from animal cell culture. New York: John Wiley and Sons p143-171

de Boer L, Gray PP, Sunstrom NA. 2004. Enhanced productivity of G1 phase Chinese hamster ovary cells using the GADD153 promoter. Biotechnology letters 26(1):61-65.

De Jesus MJ, Wurm FM. 2013. Scale-up and predictability in process development with suspension cultures of mammalian cells for recombinant protein manufacture: comments on a trend reversal. Pharmaceutical bioprocessing 1(4):333-335.

De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ. 2007. Construction and model-based analysis of a promoter library for E. coli: an indispensable tool for metabolic engineering. BMC biotechnology 7(1):34.

Dhanoya A, Chain BM, Keshavarz-Moore E. 2011. The impact of DNA topology on polyplex uptake and transfection efficiency in mammalian cells. Journal of biotechnology 155(4):377-386.

Dharshanan S, Chong H, Hung CS, Zamrod Z, Kamal N. 2011. Rapid automated selection of mammalian cell line secreting high level of humanized monoclonal antibody using Clone Pix FL system and the correlation between exterior median intensity and antibody productivity. Electronic journal of biotechnology 14(2):8-8.

DiMasi JA. 2014. Pharmaceutical R&D performance by firm size: Approval success rates and economic returns. American journal of therapeutics 21(1):26-34.

Dinnis D, James D. 2005. Engineering mammalian cell factories for improved recombinant monoclonal antibody production: lessons from nature? Biotechnology and bioengineering 91(2):180-189.

Dolcet X, Llobet D, Pallares J, Matias-Guiu X. 2005. NF-kB in development and progression of human cancer. Virchows Archiv 446(5):475-482.

Edelman GM, Meech R, Owens GC, Jones FS. 2000. Synthetic promoter elements obtained by nucleotide sequence variation and selection for activity. Proceedings of the national academy of sciences 97(7):3038-3043.

Elliott P, Billingham S, Bi J, Zhang H. 2013. Quality by design for biopharmaceuticals: a historical review and guide for implementation. Pharmaceutical bioprocessing 1(1):105-122.

Esvelt KM, Wang HH. 2013. Genome-scale engineering for systems and synthetic biology. Molecular Systems biology 9(1):10.1038/msb.2012.66.

Fan L, Kadura I, Krebs LE, Hatfield CC, Shaw MM, Frye CC. 2012. Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. Biotechnology and bioengineering 109(4):1007-1015.

Fath S, Bauer AP, Liss M, Spriestersbach A, Maertens B, Hahn P, Ludwig C, Schäfer F, Graf M, Wagner R. 2011. Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. PloS one 6(3):e17596.

Ferreira JP, Overton KW, Wang CL. 2013. Tuning gene expression with synthetic upstream open reading frames. Proceedings of the national academy of sciences 110(28):11284-11289.

Ferreira JP, Peacock RW, Lawhorn IE, Wang CL. 2011. Modulating ectopic gene expression levels by using retroviral vectors equipped with synthetic promoters. Systems and synthetic biology 5(3-4):131-138.

Fuda NJ, Ardehali MB, Lis JT. 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461(7261):186-192.

Fux C, Fussenegger M. 2003. Bidirectional expression units enable streptogramin-adjustable gene expression in mammalian cells. Biotechnology and bioengineering 83(5):618-625.

Galle CS. 2013. Characterization of cis-acting partners within the cytomegalovirus major immediate-early enhancer that strengthen MIE gene expression and viral fitness. Unpublished PhD thesis, University of Iowa.

Gambari R. 2011. Recent patents on therapeutic applications of the transcription factor decoy approach. Expert opinion on therapeutic patents 21(11):1755-1771.

Gamper HB, Reed MW, Cox T, Virosco JS, Adams AD, Gall AA, Scholler JK, Meyer RB. 1993. Facile preparation of nuclease resistant 3′ modified oligodeoxynucleotides. Nucleic acids research 21(1):145-150.

Gao H, Xiao J, Sun Q, Lin H, Bai Y, Yang L, Yang B, Wang H, Wang Z. 2006. A single decoy oligodeoxynucleotides targeting multiple oncoproteins produces strong anticancer effects. Molecular pharmacology 70(5):1621-1629.

Ghazal P, DeMattei C, Giulietti E, Kliewer SA, Umesono K, Evans RM. 1992. Retinoic acid receptors initiate induction of the cytomegalovirus enhancer in embryonal cells. Proceedings of the national academy of sciences 89(16):7630-7634.

Girod PA, Le Fourn V, Calabrese D, Regamey A, Ley D, Mermod N. 2013. High performance CHO cell line development platform for enhanced production of recombinant proteins including difficult-to-express proteins. BMC proceedings 7(6):75

Girod PA, Nguyen DQ, Calabrese D, Puttini S, Grandjean M, Martinet D, Regamey A, Saugy D, Beckmann JS, Bucher P. 2007. Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. Nature methods 4(9):747-753.

Girod PA, Zahn-Zabal M, Mermod N. 2005. Use of the chicken lysozyme 5′ matrix attachment region to generate high producer CHO cell lines. Biotechnology and bioengineering 91(1):1-11.

Gitzinger M, Kemmer C, Fluri DA, El-Baba MD, Weber W, Fussenegger M. 2012. The food additive vanillic acid controls transgene expression in mammalian cells and mice. Nucleic acids research 40(5):e37.

Golebiowski FM, Górecki A, Bonarek P, Rapala-Kozik M, Kozik A, Dziedzicka-Wasylewska M. 2012. An investigation of the affinities, specificity and kinetics involved in the interaction between the Yin Yang 1 transcription factor and DNA. FEBS Journal 279(17):3147-3158.

Gordon S, Akopyan G, Garban H, Bonavida B. 2006. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. Oncogene 25(8):1125-1142.

Gotea V, Visel A, Westlund JM, Nobrega MA, Pennacchio LA, Ovcharenko I. 2010. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. Genome research 20(5):565-577.

Gottschalk U, Brorson K, Shukla AA. 2013. Innovation in biomanufacturing: the only way forward. Pharmaceutical bioprocessing 1(2):141-157.

Grabherr MG, Pontiller J, Mauceli E, Ernst W, Baumann M, Biagi T, Swofford R, Russell P, Zody MC, Di Palma F. 2011. Exploiting nucleotide composition to engineer promoters. PloS One 6(5):e20136.

Grainger RK, James DC. 2013. CHO cell line specific prediction and control of recombinant monoclonal antibody N-glycosylation. Biotechnology and bioengineering 110(11):2970-2983.

Grassi G, Maccaroni P, Meyer R, Kaiser H, D'ambrosio E, Pascale E, Grassi M, Kuhn A, Di Nardo P, Kandolf R. 2003. Inhibitors of DNA methylation and histone deacetylation activate cytomegalovirus promoter-controlled reporter gene expression in human glioblastoma cell line U87. Carcinogenesis 24(10):1625-1635.

Guye P, Li Y, Wroblewska L, Duportet X, Weiss R. 2013. Rapid, modular and reliable construction of complex mammalian gene circuits. Nucleic acids research 41(16):e156.

Hacker D, De Jesus M, Wurm F. 2009. 25 years of recombinant proteins from reactor-grown cells - where do we go from here? Biotechnology advances 27(6):1023-1027.

Hai T, Curran T. 1991. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proceedings of the national academy of sciences 88(9):3720-3724.

Han J, McLane B, Kim EH, Yoon JW, Jun HS. 2010. Remission of diabetes by insulin gene therapy using a hepatocyte-specific and glucose-responsive synthetic promoter. Molecular therapy 19(3):470-478.

Hanania EG, Fieck A, Stevens J, Bodzin LJ, Palsson BØ, Koller MR. 2005. Automated in situ measurement of cell-specific antibody secretion and laser-mediated purification for rapid cloning of highly-secreting producers. Biotechnology and bioengineering 91(7):872-876.

Hargrove JL, Schmidt FH. 1989. The role of mRNA and protein stability in gene expression. The FASEB journal 3(12):2360-2370.

Hartenbach S, Daoud-El Baba M, Weber W, Fussenegger M. 2007. An engineered L-arginine sensor of Chlamydia pneumoniae enables arginine-adjustable

transcription control in mammalian cells and mice. Nucleic acids research 35(20):e136.

Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A. 2008. Promoter library designed for fine-tuned gene expression in Pichia pastoris. Nucleic acids research 36(12):e76.

Haupt S, Grützner J, Rath BH, Möhlig H, Brüstle O. 2009. Automated selection and collection of pluripotent stem cell colonies using the CellCelector. Nature methods 6(6):iii-iv.

He B, Weber GF. 2004. Synergistic activation of the CMV promoter by NF-κB P50 and PKG. Biochemical and biophysical research communications 321(1):13-20.

He W, Qiang M, Ma W, Valente AJ, Quinones MP, Wang W, Reddick RL, Xiao Q, Ahuja SS, Clark RA. 2006. Development of a synthetic promoter for macrophage gene therapy. Human gene therapy 17(9):949-959.

Hitoshi N, Ken-ichi Y, Jun-ichi M. 1991. Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene 108(2):193-199.

Ho SC, Koh EY, van Beers M, Mueller M, Wan C, Teo G, Song Z, Tong Y, Bardor M, Yang Y. 2013. Control of IgG LC: HC ratio in stably transfected CHO cells and study of the impact on expression, aggregation, glycosylation and conformational stability. Journal of biotechnology 165(3-4):157-166.

Hogwood CE, Tait AS, Koloteva-Levine N, Bracewell DG, Smales CM. 2013. The dynamics of the CHO host cell protein profile during clarification and protein A capture in a platform antibody purification process. Biotechnology and bioengineering 110(1):240-251.

Huang CJ, Lin H, Yang X. 2012. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. Journal of industrial microbiology & biotechnology 39(3):383-399.

Huang T, Oka T, Asai T, Okada T, Merrills B, Gertson P, Whitson R, Itakura K. 1996. Repression by a differentiation-specific factor of the human cytomegalovirus enhancer. Nucleic acids research 24(9):1695-1701.

Huisinga KL, Brower-Toland B, Elgin SC. 2006. The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma 115(2):110-122.

Hung F, Deng L, Ravnikar P, Condon R, Li B, Do L, Saha D, Tsao YS, Merchant A, Liu Z. 2010. mRNA stability and antibody production in CHO cells: improvement through gene optimization. Biotechnology journal 5(4):393-401.

Hunninghake G, Monick M, Liu B, Stinski M. 1989. The promoter-regulatory region of the major immediate-early gene of human cytomegalovirus responds to T-lymphocyte stimulation and contains functional cyclic AMP-response elements. Journal of virology 63(7):3026-3033.

Hwang DY, Carlezon Jr WA, Isacson O, Kim KS. 2001. A high-efficiency synthetic promoter that drives transgene expression selectively in noradrenergic neurons. Human gene therapy 12(14):1731-1740.

Isomura H, Stinski MF, Kudoh A, Daikoku T, Shirata N, Tsurumi T. 2005. Two Sp1/Sp3 binding sites in the major immediate-early proximal enhancer of human cytomegalovirus have a significant role in viral replication. Journal of virology 79(15):9597-9607.

Isomura H, Tsurumi T, Stinski MF. 2004. Role of the proximal enhancer of the major immediate-early promoter in human cytomegalovirus replication. Journal of virology 78(23):12788-12799.

Jackson RJ, Hellen CU, Pestova TV. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. Nature reviews molecular cell biology 11(2):113-127.

James RI, Elton JP, Todd P, Kompala DS. 2000. Engineering CHO cells to overexpress a secreted reporter protein upon induction from mouse mammary tumor virus promoter. Biotechnology and bioengineering 67(2):134-140.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. 2011. Global cancer statistics. CA: a cancer journal for clinicians 61(2):69-90.

Jiang Z, Sharfstein ST. 2008. Sodium butyrate stimulates monoclonal antibody over-expression in CHO cells by improving gene accessibility. Biotechnology and bioengineering 100(1):189-194.

Jianwei D, Qianqian Z, Songcai L, Mingjun Z, Xiaohui R, Linlin H, Qingyan J, Yongliang Z. 2012. The combination of a synthetic promoter and a CMV promoter improves foreign gene expression efficiency in myocytes. Journal of biotechnology 158(3):91-96.

Juven-Gershon T, Cheng S, Kadonaga J. 2006. Rational design of a super core promoter that enhances gene expression. Nature methods 3(11):917-922.

Juven-Gershon T, Hsu JY, Theisen JWM, Kadonaga JT. 2008. The RNA polymerase II core promoter - the gateway to transcription. Current opinion in cell biology 20(3):253-259.

Juven-Gershon T, Kadonaga JT. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental biology 339(2):225-229.

Kalwy S, Rance J, Young R. 2006. Toward more efficient protein expression. Molecular biotechnology 34(2):151-156.

Kelen KVD, Beyaert R, Inzé D, Veylder LD. 2009. Translational control of eukaryotic gene expression. Critical reviews in biochemistry and molecular biology 44(4):143-168.

Keller MJ, Wu AW, Andrews JI, McGonagill PW, Tibesar EE, Meier JL. 2007. Reversal of human cytomegalovirus major immediate-early enhancer/promoter silencing in quiescently infected cells via the cyclic AMP signaling pathway. Journal of virology 81(12):6669-6681.

Kelly RJ, Smith TJ. 2014. Delivering maximum clinical benefit at an affordable price: engaging stakeholders in cancer care. The lancet oncology 15(3):e112-e118.

Khattak SF, Xing Z, Kenty B, Koyrakh I, Li ZJ. 2010. Feed development for fed-batch CHO production process by semisteady state analysis. Biotechnology progress 26(3):797-804.

Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. Genome research 23(5):800-811.

Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ. 2013. The emerging CHO systems biology era: harnessing the ‘omics revolution for biotechnology. Current opinion in biotechnology 24(6):1102-1107.

Kim JY, Kim Y-G, Lee GM. 2012a. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Applied microbiology and biotechnology 93(3):917-930.

Kim M, O'Callaghan PM, Droms KA, James DC. 2011. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnology and bioengineering 108(10):2434-2446.

Kim YG, Park B, Ahn JO, Jung J-K, Lee HW, Lee EG. 2012b. New cell line development for antibody-producing Chinese hamster ovary cells using split green fluorescent protein. BMC biotechnology 12(1):24.

Kiomy Osako M, Nakagami H, Morishita R. 2012. Modification of decoy oligodeoxynucleotides to achieve the stability and therapeutic efficacy. Current topics in medicinal chemistry 12(15):1603-1607.

Koh EY, Ho SC, Song Z, Bi X, Bardor M, Yang Y. 2013. An internal ribosome entry site (IRES) mutant library for tuning expression level of multiple genes in mammalian cells. PloS One 8(12):e82100.

Koschmann J, Machens F, Becker M, Niemeyer J, Schulze J, Bülow L, Stahl DJ, Hehl R. 2012. Integration of bioinformatics and synthetic promoters leads to the discovery of novel elicitor-responsive cis-regulatory sequences in Arabidopsis. Plant physiology 160(1):178-191.

Kreutzer S, Schmidt C, Fuchs MC, Dietze M, Fischer M, Fuchs M. 2012.Introducing an R package for luminescence dating analysis. Ancient TL 30(1):1-8.

Kwaks TH, Barnett P, Hemrika W, Siersma T, Sewalt RG, Satijn DP, Brons JF, van Blokland R, Kwakman P, Kruckeberg AL. 2003. Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. Nature biotechnology 21(5):553-558.

Lai T, Yang Y, Ng SK. 2013. Advances in mammalian cell line development technologies for recombinant protein production. Pharmaceuticals 6(5):579-603.

Langer ES. 2012. Biomanufacturing outsourcing outlook. BioPharm international 25(2):15-16.

Lashmit P, Wang S, Li H, Isomura H, Stinski MF. 2009. The CREB site in the proximal enhancer is critical for cooperative interaction with the other transcription factor binding sites to enhance transcription of the major intermediate-early genes in human cytomegalovirus-infected cells. Journal of virology 83(17):8893-8904.

Lashmit PE, Lundquist CA, Meier JL, Stinski MF. 2004. Cellular repressor inhibits human cytomegalovirus transcription from the UL127 promoter. Journal of virology 78(10):5113-5123.

Lattenmayer C, Trummer E, Schriebl K, Vorauer-Uhl K, Mueller D, Katinger H, Kunert R. 2007. Characterisation of recombinant CHO cell lines by investigation of protein productivities and genetic parameters. Journal of biotechnology 128(4):716-725.

Le Fourn V, Girod PA, Buceta M, Regamey A, Mermod N. 2014. CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. Metabolic engineering 21:91-102.

Le H, Vishwanathan N, Kantardjieff A, Doo I, Srienc M, Zheng X, Somia N, Hu WS. 2013. Dynamic gene expression for metabolic engineering of mammalian cells in culture. Metabolic engineering 20:212-220.

Lee J, Lau J, Chong G, Chao SH. 2007. Cell-specific effects of human cytomegalovirus unique region on recombinant protein expression. Biotechnology letters 29(12):1797-1802.

Lee JC, Wu TY, Huang CF, Yang FM, Shih SR, Hsu JT. 2005. High-efficiency protein expression mediated by enterovirus 71 internal ribosome entry site. Biotechnology and bioengineering 90(5):656-662.

Lee TH, Maheshri N. 2012. A regulatory role for repeated decoy transcription factor binding sites in target gene expression. Molecular systems biology 8(1):576.

Lee WR, Kim KH, An HJ, Park YY, Kim KS, Lee CK, Min BK, Park Kk. 2012. Effects of Chimeric Decoy Oligodeoxynucleotide in the regulation of transcription factors NF-κB and Sp1 in an animal model of atherosclerosis. Basic & clinical pharmacology & toxicology 112(4):236-243.

Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C. 2013. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nature biotechnology 31(8):759-765.

Li F, Shen A, Amanullah A. 2010. Cell culture processes in monoclonal antibody production. mAbs 2(5):466-477.

Li X, Eastman EM, Schwartz RJ, Draghia-Akli R. 1999. Synthetic muscle promoters: activities exceeding naturally occurring regulatory sequences. Nature biotechnology 17(3):241-245.

Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes & development 18(13):1606.

Lin CY, Lovén J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA. 2012. Transcriptional amplification in tumor cells with elevated c-Myc. Cell 151(1):56-67.

Liu R, Baillie J, Sissons JP, Sinclair JH. 1994. The transcription factor YY1 binds to negative regulatory elements in the human cytomegalovirus major immediate early enhancer/promoter and mediates repression in nonpermissive cells. Nucleic acids research 22(13):2453-2459.

Liu X, Wang X, Yan S, Zhang Z, Abecassis M, Hummel M. 2013. Epigenetic control of cytomegalovirus latency and reactivation. Viruses 5(5):1325-1345.

Liu X, Yan S, Abecassis M, Hummel M. 2010b. Biphasic recruitment of transcriptional repressors to the murine cytomegalovirus major immediate-early promoter during the course of infection in vivo. Journal of virology 84(7):3631-3643.

Liu X, Yuan J, Wu AW, McGonagill PW, Galle CS, Meier JL. 2010a. Phorbol ester-induced human cytomegalovirus major immediate-early (MIE) enhancer activation through PKC-Delta, CREB, and NF-κB desilences MIE gene expression in quiescently infected human pluripotent NTera2 cells. Journal of virology 84(17):8495-8508.

Ma S, Tang N, Tian J. 2012. DNA synthesis, assembly and applications in synthetic biology. Current opinion in chemical biology 16(3):260-267.

Mader A, Prewein B, Zboray K, Casanova E, Kunert R. 2012. Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. Applied microbiology and biotechnology 97(9):4049-4054.

Makrides SC. 1999. Components of vectors for gene transfer and expression in mammalian cells. Protein expression and purification 17(2):183-202.

Manke T, Roider HG, Vingron M. 2008. Statistical modeling of transcription factor binding affinities predicts regulatory interactions. PLoS computational biology 4(3):e1000039.

Mann CJ. 2007. Rapid isolation of antigen-specific clones from hybridoma fusions. Nature methods 4(4):i-ii.

Manna PR, Stocco DM. 2007. Crosstalk of CREB and Fos/Jun on a single cis-element: transcriptional repression of the steroidogenic acute regulatory protein gene. Journal of molecular endocrinology 39(4):261-277.

Mariati NY, Chao SH, Yap MG, Yang Y. 2010. Evaluating regulatory elements of human cytomegalovirus major immediate early gene for enhancing transgene expression levels in CHO K1 and HEK293 cells. Journal of biotechnology 147(3-4):160-3.

Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. 2011. Projections of the cost of cancer care in the United States: 2010–2020. Journal of the national cancer institute 103(2):117-128.

Masterton RJ, Smales CM. 2014. The impact of process temperature on mammalian cell lines and the implications for the production of recombinant proteins in CHO cells. Pharmaceutical bioprocessing 2(1):49-61.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen C-y, Chou A, Ienasescu H. 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic acids research 42(D1):D142-D147.

McGrew MJ, Sherman A, Ellard FM, Lillico SG, Gilhooley HJ, Kingsman AJ, Mitrophanous KA, Sang H. 2004. Efficient production of germline transgenic chickens using lentiviral vectors. EMBO reports 5(7):728-733.

McLeod J, O'Callaghan PM, Pybus LP, Wilkinson SJ, Root T, Racher AJ, James DC. 2011. An empirical modeling platform to evaluate the relative control discrete CHO cell synthetic processes exert over recombinant monoclonal antibody production process titer. Biotechnology and bioengineering 108(9):2193-2204.

Mead EJ, Chiverton LM, Smales CM, von der Haar T. 2009. Identification of the limitations on recombinant gene expression in CHO cell lines with varying luciferase production rates. Biotechnology and bioengineering 102(6):1593-1602.

Mella-Alvarado V, Gautier A, Le Gac F, Lareyre JJ. 2013. Tissue and cell-specific transcriptional activity of the human cytomegalovirus immediate early gene promoter (UL123) in zebrafish. Gene expression patterns 13(3):91-103.

Mellitzer A, Weis R, Glieder A, Flicker K. 2012. Expression of lignocellulolytic enzymes in Pichia pastoris. Microbial cell factories 11(1):61.

Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Mikkelsen TS. 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nature biotechnology 30(3):271-277.

Miller DW. 2012. Value-based pricing. Pharmaceutical medicine 26(4):217-222.

Miyake T, Aoki M, Nakashima H, Kawasaki T, Oishi M, Kataoka K, Tanemoto K, Ogihara T, Kaneda Y, Morishita R. 2006. Prevention of abdominal aortic aneurysms by simultaneous inhibition of NFκB and ets using chimeric decoy oligonucleotides in a rabbit model. Gene therapy 13(8):695-704.

Morton FS, Kyle M. 2012. Markets for pharmaceutical products. Handbook of health economics 2:763-823.

Murphy JC, Fischle W, Verdin E, Sinclair JH. 2002. Control of cytomegalovirus lytic gene expression by histone acetylation. The EMBO journal 21(5):1112-1120.

Nehlsen K, Schucht R, da Gama-Norton L, Krömer W, Baer A, Cayli A, Hauser H, Wirth D. 2009. Recombinant protein expression by targeting pre-selected chromosomal loci. BMC biotechnology 9(1):100.

Noh SM, Sathyamurthy M, Lee GM. 2013. Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production. Current opinion in chemical engineering 2(4):391-397.

Nott A, Meislin SH, Moore MJ. 2003. A quantitative analysis of intron effects on mammalian gene expression. RNA 9(5):607-617.

O'Callaghan P, McLeod J, Pybus L, Lovelady C, Wilkinson S, Racher A, Porter A, James D. 2010. Cell line-specific control of recombinant monoclonal antibody production by CHO cells. Biotechnology and bioengineering 106(6):938-951.

Ogawa R, Kagiya G, Kodaki T, Fukuda S, Yamamoto K. 2007. Construction of strong mammalian promoters by random cis-acting element elongation. BioTechniques 42(5):628.

Osako MK, Tomita N, Nakagami H, Kunugiza Y, Yoshino M, Yuyama K, Tomita T, Yoshikawa H, Ogihara T, Morishita R. 2007. Increase in nuclease resistance and incorporation of NF-κB decoy oligodeoxynucleotides by modification of the 3′-terminus. The journal of gene medicine 9(9):812-819.

Palermo D, DeGraaf M, Marotti K, Rehberg E, Post L. 1991. Production of analytical quantities of recombinant proteins in Chinese hamster ovary cells using sodium butyrate to elevate gene expression. Journal of biotechnology 19(1):35-47.

Pandhal J, Desai P, Walpole C, Doroudi L, Malyshev D, Wright PC. 2012. Systematic metabolic engineering for improvement of glycosylation efficiency in Escherichia coli. Biochemical and biophysical research communications 419(3):472-476.

Pasotti L, Politi N, Zucca S, De Angelis MGC, Magni P. 2012. Bottom-up engineering of biological systems through standard bricks: A modularity study on basic parts and devices. PloS one 7(7):e39407.

Pilbrough W, Munro TP, Gray P. 2009. Intraclonal protein expression heterogeneity in recombinant CHO cells. PloS one 4(12):e8432.

Pizzorno MC. 2001. Nuclear cathepsin B-like protease cleaves transcription factor YY1 in differentiated cells. Biochimica et biophysica acta - molecular basis of disease 1536(1):31-42.

Pollock J, Ho SV, Farid SS. 2013. Fed-batch and perfusion culture processes: economic, environmental, and operational feasibility under uncertainty. Biotechnology and bioengineering 110(1):206-219.

Pontiller J, Gross S, Thaisuchat H, Hesse F, Ernst W. 2008. Identification of CHO endogenous promoter elements based on a genomic library approach. Molecular biotechnology 39(2):135-139.

Pontiller J, Maccani A, Baumann M, Klancnik I, Ernst W. 2010. Identification of CHO endogenous gene regulatory elements. Molecular biotechnology 45(3):235-240.

Prentice H, Tonkin C, Caamano L, Sisk W. 2007. High level expression of proteins using sequences from the ferritin heavy chain gene locus. Journal of biotechnology 128(1):50-60.

Prösch S, Heine A-K, Volk H-D, Krüger DH. 2001. CCAAT/enhancer-binding proteins α and β negatively influence the capacity of tumor necrosis factor α to up-regulate the human cytomegalovirus IE1/2 enhancer/promoter by nuclear factor κB during monocyte differentiation. Journal of biological chemistry 276(44):40712-40720.

Pshenichkin S, Surin A, Surina E, Klauzińska M, Grajkowska E, Luchenko V, Dolińska M, Wroblewska B, Wroblewski JT. 2011. Heat shock enhances CMV-IE promoter-driven metabotropic glutamate receptor expression and toxicity in transfected cells. Neuropharmacology 60(7):1292-1300.

Puck TT. 1985. Development of the Chinese hamster ovary (CHO) cell for use in somatic cell genetics. In: Gottesman MM, editor. Molecular cell genetics. New York: John Wiley and Sons p37-64

Pybus LP, Dean G, West NR, Smith A, Daramola O, Field R, Wilkinson SJ, James DC. 2013. Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells. Biotechnology and bioengineering 111:372-385.

Qin JY, Zhang L, Clift KL, Hulur I, Xiang AP, Ren BZ, Lahn BT. 2010. Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PloS one 5(5):e10611.

Radhakrishnan P, Basma H, Klinkebiel D, Christman J, Cheng PW. 2008. Cell type-specific activation of the cytomegalovirus promoter by dimethylsulfoxide and 5-Aza-2'-deoxycytidine. The international journal of biochemistry and cell biology 40(9):1944-1955.

Rajkumar AS, Maerkl SJ. 2012. Rapid synthesis of defined eukaryotic promoter libraries. ACS synthetic biology 1(10):483-490.

Renard E, Porée B, Chadjichristos C, Kypriotou M, Maneix L, Bigot N, Legendre F, Ollitrault D, De Crombrugghe B, Malléin-Gérin F. 2012. Sox9/Sox6 and Sp1 are involved in the insulin-like growth factor-I-mediated upregulation of human type II collagen gene expression in articular chondrocytes. Journal of molecular medicine 90(6):649-666.

Révet B, Fourcade A. 1998. Short unligated sticky ends enable the observation of circularised DNA by atomic force and electron microscopies. Nucleic acids research 26(9):2092-2097.

Rita Costa A, Elisa Rodrigues M, Henriques M, Azeredo J, Oliveira R. 2010. Guidelines to cell engineering for monoclonal antibody production. European Journal of Pharmaceutics and Biopharmaceutics 74(2):127-138.

Running Deer J, Allison D. 2004. High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1 gene. Biotechnology progress 20(3):880-889.

Ruth C, Zuellig T, Mellitzer A, Weis R, Looser V, Kovar K, Glieder A. 2010. Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in Pichia pastoris. Systems and synthetic biology 4(3):181-191.

Schlabach MR, Hu JK, Li M, Elledge SJ. 2010. Synthetic design of strong promoters. Proceedings of the national academy of sciences 107(6):2538.

Schug J. 2008. Using TESS to predict transcription factor binding sites in DNA sequence. Current protocols in bioinformatics 21:2.6.1-2.6.15.

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. Nature 473(7347):337-342.

Seghezzi N, Amar P, Koebmann B, Jensen PR, Virolle MJ. 2011. The construction of a library of synthetic promoters revealed some specific features of strong Streptomyces promoters. Applied microbiology and biotechnology 90(2):615-623.

Sellick CA, Croxford AS, Maqsood AR, Stephens G, Westerhoff HV, Goodacre R, Dickson AJ. 2011. Metabolite profiling of recombinant CHO cells: designing tailored feeding regimes that enhance recombinant antibody production. Biotechnology and bioengineering 108(12):3025-3031.

Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, Keren L, Yakhini Z, Weinberger A, Segal E. 2012. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nature biotechnology 30(6):521-530.

Shitara K. 2009. Potelligent antibodies as next generation therapeutic antibodies. Yakugaku zasshi: journal of the pharmaceutical society of japan 129(1):3-9.

Shore D, Langowski J, Baldwin RL. 1981. DNA flexibility studied by covalent closure of short fragments into circles. Proceedings of the national academy of sciences 78(8):4833-4837.

Shukla AA, Gottschalk U. 2013. Single-use disposable technologies for biopharmaceutical manufacturing. Trends in biotechnology 31(3):147-154.

Shukla AA, Thömmes J. 2010. Recent advances in large-scale production of monoclonal antibodies and related proteins. Trends in biotechnology 28(5):253-261.

Siegel R, Naishadham D, Jemal A. 2013. Cancer statistics, 2013. CA: a cancer journal for clinicians 63(1):11-30.

Skoko N, Baralle M, Tisminetzky S, Buratti E. 2011. InTRONs in Biotech. Molecular biotechnology 48(3):290-297.

Smith RP, Riesenfeld SJ, Holloway AK, Li Q, Murphy KK, Feliciano NM, Orecchia L, Oksenberg N, Pollard KS, Ahituv N. 2013. A compact, in vivo screen of all 6-mers reveals drivers of tissue-specific expression and guides synthetic regulatory element design. Genome biology 14(7):R72.

Spadiut O, Capone S, Krainer F, Glieder A, Herwig C. 2014. Microbials for the production of monoclonal antibodies and antibody fragments. Trends in biotechnology 32(1):54-60.

Stadlmayr G, Mecklenbräuker A, Rothmüller M, Maurer M, Sauer M, Mattanovich D, Gasser B. 2010. Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production. Journal of biotechnology 150(4):519-529.

Stinski MF, Isomura H. 2008. Role of the cytomegalovirus major immediate early enhancer in acute infection and reactivation from latency. Medical microbiology and immunology 197(2):223-231.

R Core Development Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Thaisuchat H, Baumann M, Pontiller J, Hesse F, Ernst W. 2011. Identification of a novel temperature sensitive promoter in cho cells. BMC biotechnology 11(1):51.

Theisen JWM, Lim CY, Kadonaga JT. 2010. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. Molecular and cellular biology 30(14):3471.

Tigges M, Fussenegger M. 2009. Recent advances in mammalian synthetic biology - design of synthetic transgene control networks. Current opinion in biotechnology 20(4):449-460.

Tomita S, Tomita N, Yamada T, Zhang L, Kaneda Y, Morishita R, Ogihara T, Dzau VJ, Horiuchi M. 1999. Transcription factor decoy to study the molecular mechanism of negative regulation of renin gene expression in the liver in vivo. Circulation research 84(9):1059-1066.

Torella JP, Boehm CR, Lienert F, Chen JH, Way JC, Silver PA. 2014. Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. Nucleic acids research 42(1):681-689.

Tornøe J, Kusk P, Johansen TE, Jensen PR. 2002. Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. Gene 297(1):21-32.

Ulanovsky L, Bodner M, Trifonov EN, Choder M. 1986. Curved DNA: design, synthesis, and circularization. Proceedings of the national academy of sciences 83(4):862-866.

Umana P, Jean-Mairet J, Moudry R, Amstutz H, Bailey J. 1999. Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. Nature biotechnology 17:176-180.

Van Dam H, Castellazzi M. 2001. Distinct roles of Jun: Fos and Jun: ATF dimers in oncogenesis. Oncogene 20(19):2453-2464.

Van Helden J, André B, Collado-Vides J. 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. Journal of molecular biology 281(5):827-842.

Vasey D, Lillico S, Sang H, King T, Whitelaw C. 2009. CMV enhancer–promoter is preferentially active in exocrine cells in vivo. Transgenic research 18(2):309-314.

Vogl T, Hartner FS, Glieder A. 2013. New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris. Current opinion in biotechnology 24(6):1094-1101.

Walter CA, Humphrey RM, Adair GM, Nairn RS. 1991. Characterization of Chinese hamster ovary cells stably transformed by a plasmid with an inducible APRT gene. Plasmid 25(3):208-216.

Wang F, Wang TY, Tang YY, Zhang JH, Yang XJ. 2012. Different matrix attachment regions flanking a transgene effectively enhance gene expression in stably transfected Chinese hamster ovary cells. Gene 500(1):59-62.

Wang TY, Yang R, Qin C, Wang L, Yang XJ. 2008. Enhanced expression of transgene in CHO cells using matrix attachment region. Cell biology international 32(10):1279-1283.

Wang W, Yi X, Zhang Y. 2007. Gene transcription acceleration: main cause of hepatitis B surface antigen production improvement by dimethyl sulfoxide in

the culture of Chinese hamster ovary cells. Biotechnology and bioengineering 97(3):526-535.

Weber W, Fussenegger M. 2010. Synthetic gene networks in mammalian cells. Current opinion in biotechnology 21(5):690-696.

Weber W, Fux C, Daoud-El Baba M, Keller B, Weber CC, Kramer BP, Heinzen C, Aubel D, Bailey JE, Fussenegger M. 2002. Macrolide-based transgene control in mammalian cells and mice. Nature biotechnology 20(9):901-907.

Weber W, Lienhart C, Daoud-El Baba M, Fussenegger M. 2009. A biotin-triggered genetic switch in mammalian cells and mice. Metabolic engineering 11(2):117-124.

Wulhfard S, Tissot S, Bouchet S, Cevey J, De Jesus M, Hacker DL, Wurm FM. 2008. Mild hypothermia improves transient gene expression yields several fold in Chinese hamster ovary cells. Biotechnology progress 24(2):458-465.

Wurm FM. 2013. CHO quasispecies - implications for manufacturing processes. Processes 1(3):296-311.

Würtele H, Little K, Chartrand P. 2003. Illegitimate DNA integration in mammalian cells. Gene therapy 10(21):1791-1799.

Xia W, Bringmann P, McClary J, Jones P, Manzana W, Zhu Y, Wang S, Liu Y, Harvey S, Madlansacay M. 2006. High levels of protein expression using different mammalian CMV promoters in several cell lines. Protein expression and purification 45(1):115-124.

Xiong J, Sun W, Wang W, Liao Z, Zhou F, Kong H, Xu Y, Xie C, Zhou Y. 2012. Novel, chimeric, cancer-specific, and radiation-inducible gene promoters for suicide gene therapy of cancer. Cancer 118(2):536-548.

Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S. 2011. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature biotechnology 29(8):735-741.

Yamane-Ohnuki N, Kinoshita S, Inoue-Urakubo M, Kusunoki M, Iida S, Nakano R, Wakitani M, Niwa R, Sakurada M, Uchida K. 2004. Establishment of FUT8 knockout Chinese hamster ovary cells: An ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. Biotechnology and bioengineering 87(5):614-622.

Yang Y. 2010. DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. Journal of biotechnology 147(3-4):180-5.

Yim SS, An SJ, Kang M, Lee J, Jeong KJ. 2013. Isolation of fully synthetic promoters for high-level gene expression in corynebacterium glutamicum. Biotechnology and bioengineering 110(11):2959-2969.

Zago P, Baralle M, Ayala YM, Skoko N, Zacchigna S, Buratti E, Tisminetzky S. 2009. Improving human interferon-β production in mammalian cell lines by insertion of an intronic sequence within its naturally uninterrupted gene. Biotechnology and applied biochemistry 52(3):191-198.

Zhang Q, Stovall DB, Inoue K, Sui G. 2011. The oncogenic role of Yin Yang 1. Critical reviews in oncogenesis 16(3-4):163–*19*7

Zhang Y, Crothers DM. 2003. Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. Biophysical journal 84(1):136-153.

Zhou H, Liu Z, Sun Z, Huang Y, Yu W. 2010. Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. Journal of biotechnology 147(2):122-129.

Zhu J. 2012. Mammalian cell protein expression for biopharmaceutical production. Biotechnology advances 30(5):1158-1170.

Zweidler-McKay PA, Grimes HL, Flubacher MM, Tsichlis PN. 1996. Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor. Molecular and cellular biology 16(8):4024-4034.