Manipulation of RNA and RNA-Binding Proteins for Control of Biopharmaceutical Titre

James Baker

Submitted for the degree of Doctor of Philosophy November 2019

Department of Molecular Biology and Biotechnology

University of Sheffield

Supervisors

Prof. David James

Department of Chemical and Biological Engineering

University of Sheffield

Prof. Stuart Wilson

Department of Molecular Biology and Biotechnology

University of Sheffield

Dr. Suzanne Gibson

AstraZeneca

Granta Park, Cambridge

The 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.

Acknowledgements

This research would not have been possible without funding from AstraZeneca.

I firstly owe thanks to Professor David James, for the support, direction, and confidence I needed to see this project through. My thanks also go out to Dr Suzanne Gibson, who has been a constant supply of friendly and expert advice, going above and beyond despite her busy role at AstraZeneca, and Professor Stuart Wilson for taking me on in this project. I am deeply grateful to Professor Egbert Hoiczyk, Professor Jeffrey Green, and Linda Harris, for their kind personal advice and help.

My colleagues in the DCJ laboratory group have been ceaselessly welcoming, helpful and friendly, and it was a delight to share a space and to work with them in my research project. In particular, Dr. Yash Patel was an indispensable advisor, always there with an expert suggestion when things went wrong, and a good friend, always bringing a smile alongside it.

My family and friends, inside and outside of Sheffield, have supported me in ways I never anticipated I would need. On not one occasion was a call for help unanswered, and for this I am deeply grateful. Special thanks should go to Pete, for managing to live with me through each of our stressful last three years, for always making me laugh, and for his laudable efforts to not become a King Dedede main. My deepest gratitude goes out to everybody who loved, supported, and laughed with me along the way.

Finally, I'd like to thank my partner Chloé, for her unending smiles, support, and belief in me.

Abstract

The biopharmaceutical industry is at a crossroads, where changing industry and ethical demands means it must invest in research and development (R&D) to maintain a diverse pipeline of novel drugs, and simultaneously reduce the substantial risks and costs associated with this R&D. In the future, a complete toolkit of synthetic expression elements could be used to create a streamlined, rationalised next generation cell line development (CLD) process, maximising the benefits offered by synthetic biology. However, synthetic, predictable, titratable molecular biology tools have not yet been developed for all the desired steps of biopharmaceutical expression. In this thesis, I will present work endeavouring to expand this toolbox, by investigating synthetic control of glycoprotein expression through messenger RNA (mRNA) engineering.

The tethering of various RNA-binding proteins to recombinant mRNA was tested, as a method of controlling mRNA processing, and stimulating intronless mRNA export. *C1orf35* and *HuR* were identified as target effector genes to increase transient protein production. However, inconsistency in their effect disqualified them as effective molecular biology tools.

Two families of 3′ untranslated region (UTR) RNA elements were screened for their ability to control productivity through enhancement of mRNA stability. Though triple helices failed to increase expression, a stability element was discovered to increase transient SEAP productivity by 1.28-fold, via a mechanism of extension of SEAP mRNA half-life from 0.68h to 4.04h, compared to an industry-standard vector.

5′ terminal oligo-pyrimidine (TOP) motifs were investigated for their ability to control productivity through translation initiation. Enhancement of transient titre was demonstrated with various recombinant proteins and culture conditions, chemical supplements screened for their ability to specifically affect 5′TOP activation, mechanism of titre control investigated, and 5′TOP motifs integrated with synthetic proximal and core promoters. Different 5′TOP motifs were shown to control titre of a biotherapeutic fusion protein in an industry transient production process, up to a 2.12-fold increase compared to an industry-standard vector.

These synthetic elements were screened together for their modularity, demonstrating titratable control of SEAP titre, from a 0.82-fold decrease to a 5.23-fold increase compared to an industry-standard vector. Further steps were then recommended, to render these tools truly predictable, and contribute maximally towards a rationalised, next-generation CLD process.

Table of Contents

Table of Figures

Table of Tables

Acronyms and Abbreviations

3'SS	3' Splice Site
4E-BP	Eukaryotic Translation Initiation Factor 4E-Binding Protein
5'TOP	5' Terminal Oligo-Pyrimidine
ABCE1	ATP-Binding Cassette Sub-Family E Member 1
ActD	Actinomycin D
Ago2	Argonaute 2
AKT	AKT Serine/Threonine Kinase 1
ALYREF	Aly/REF Export Factor
AmpR	Ampicillin Resistance
ARE	A/U Rich Element
ATF6	Activating Transcription Factor 6
Bak	Bcl-2 Homologous Antagonist/Killer
Bax	*BCL2*-associated X Protein
BCL	B-cell Lymphoma 2
BHK	Baby Hamster Kidney
BiP	Binding Immunoglobulin Protein
bp	Base Pair
BP	Binding Protein
BS	Binding Sites
BTZ	Barentsz
C_orf_	Chromosome _ Open Reading Frame _
CAI	Codon Adaptation Index
CAR	Cytoplasmic-Accumulation Regions
CAR-T	Chimeric Antigen Receptor Therapy
CASTOR1	Cytosolic Arginine Sensor for MTORC1 Subunit 1
CBC	Cab-Binding Complex
CCR4	C-C Chemokine Receptor Type 4
CD47	Cluster of Differentiation 47

CD-CHO	Chemically-Defined CHO
cDNA	Complementary DNA
CDS	Coding Sequence
cFIm	Cleavage Factor
CHO	Chinese Hamster Ovary
CHTOP	Chromatin Target of PRMT1
CIRP	Cold-Inducible RNA-binding Protein
CLD	Cell Line Development
CMV	Cytomegalovirus
CPSF	Polyadenylation Specificity Factor
CRE	cAMP Response Element
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
CstF	Cleavage Stimulation Factor
CTD	Carboxy-Terminal Domain
CypB	Peptidyl-Prolyl Cis-Trans Isomerase B
DCP1/2	De-Capping Protein 1/2
DDX39B	Spliceosome RNA helicase *DDX39B*
DDX6	DEAD-Box Helicase 6
DHFR	Dihydrofolate Reductase
DNA	Deoxyribonucleotide
DTE	Difficult to Express
DTT	Dithiothreitol
EBNA	Epstein-Barr Nuclear Antigen
EBV	Epstein-Barr Virus
ECL	Enhanced Chemiluminescent
Edc1/2	Enhancer of mRNA-Decapping Protein 1
eEF	Eukaryotic Elongation Factor
Ef1α	Elongation Factor 1 Alpha
eIF	Eukaryotic Initiation Factor

EJC	Exon Junction Complex
ELISA	Enzyme-Linked Immunosorbent Assay
EMA	European Medical Association
EMCV	Encephalomyocarditis Virus
ENE	Elements for Nuclear Expression
EPO	Erythropoietin
ER	Endoplasmic Reticulum
ERAD	ER-Associated Degradation
eRF	Eukaryotic Recycling Factor
ERK	Extracellular Regulated Kinase 1
ERSE	ER Stress Response Element
ETE	Easy to Express
FACS	Fluorescence-Activated Cell Sorting
FDA	Food and Drug Administration
Flp/FRT	Flippase/Flippase Recognition Target
FPKM	Fragments Per Kilobase Million
FUT8	Fucosyltransferase 8
GDP	Guanosine Diphosphate
GFP	Green Fluorescent Protein
gRNA	Guide RNA
GS	Glutamine Synthetase
GTP	Guanosine Triphosphate
HBV	Hepatitis B Virus
HC	Heavy Chain
HDR	Homology-Directed Repair
HEK	Human Embryonic Kidney
HSP27	Heat-Shock Protein 27
HuR	Human Antigen R
IgM	Immunoglobulin M
Inr	Initiator

IRES	Internal Ribosome Entry Site
IVCD	Integral of Viable Cell Density
kb	Kilobase
KSHV	Kaposi's Sarcoma-Associated Herpesvirus
LARP1	LA-Related Protein 1
LB	Lysogeny Broth
LC	Light Chain
lncRNA	Long noncoding RNA
mAb	Monoclonal Antibody
MAGOH	Protein Mago Nashi Homolog
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MAR	Matrix Attachment Region
MENB	Multiple Endocrine Neoplasia-β
Met-tRNAi	Initiator Methionyl-tRNA
MGVT	Multi-Gene Vector (Transient)
miRNA	Micro Ribonucleotide
mLST8	*MTOR* Associated Protein, LST8 Homolog
MPRA	Massively Parallel Reporter Assay
mRNA	Messenger Ribonucleotide
mRNP	Messenger Ribonucleoprotein
MSX	L-Methionine Sulfoximine
mTOR	Mammalian Target of Rapamycin
MTORC1	*mTOR* Complex 1
MTX	Methotrexate
NF-KB	Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells
NGD	No-Go Decay
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localisation Sequence
NMD	Nonsense Mediated Decay

NOT	Negative Regulator of Transcription Factor 1
NPC	Nuclear Pore Complex
NSD	Non-Stop Decay
NXF1	Nuclear RNA Export Factor 1
NXT1	NTF2-Related Export Protein 1
pA	Poly A
PABP	PolyA Binding Protein
PAN	Polyadenylated Nuclear
PAN2/3	Deadenylation Complex Catalytic Subunit 2/3
PBA	4-Phenylbutyrate
PBS	Phosphate-Buffered Saline
pcd	Per Cell Day
PCR	Polymerase Chain Reaction
PDI	Protein Disulphide Isomerase
pDNA	Plasmid DNA
PEI	Polyethyleneimine
PERK	Proline-Rich Receptor-Like Protein 13
PIC	Pre-Initiation Complex
PP7d	PP7 Dimer
pri-miRNA	Primary Transcript miRNA
Prp19	Pre-mRNA Processing Factor 19
PTC	Premature Termination Codon
qP	Specific Productivity
qPCR	Quantitative PCR
RBM3	RNA Binding Motif Protein 3
RISC	RNA-Inducible Silencing Complex
RMCE	Recombinase-Mediated Cassette Exchange
RNA	Ribonucleotide
RNAPII	RNA Polymerase II

RNP	Ribonucleoprotein
RPU	Relative Promoter Units
rRNA	Ribosomal RNA
Rrp44	Exosome Complex Exonuclease RRP44
RSK	Ribosomal S6 Kinase
RT-qPCR	Reverse-Transcription Quantitative PCR
S6K	Ribosomal Protein S6 Kinase
scFP	Single-Chain Fusion Protein
scFv	Single Chain Variable Fragment
SDM	Site-Directed Mutagenesis
SE	Stability Element
SEAP	Secreted Alkaline Phosphatase
siRNA	Small Interfering RNA
SNAP23	Synaptosome Associated Protein 23
snRNA	Small Nuclear RNA
SP	Signal Peptide
SRP	Signal Recognition Particle
SV40	Simian Virus 40
TALEN	Transcription Activator-like Effector Nucleases
TBE	Tris-Borate EDTA
TBP	TATA Box Binding Protein
TBS	Tris-Buffered Saline
TFRE	Transcription Factor Response Element
TGE	Transient Gene Expression
THO	THO Complex Subunit
TNFα	Tumour Necrosis Factor Alpha
t-PA	Tissue Plasminogen Activator
tracrRNA	Trans-activating CRISPR RNA
TREX	Transcription/Export Complex
TRF2	TBP-Related Factor 2

tRNA	Transfer Ribonucleotide
TSC	Tuberous Sclerosis 1
TSS	Transcription Start Site
TU	Transcription Unit
UCOE	Ubiquitous Chromatin Opening Elements
UIF	UAP56-Interacting Factor
Upf1	Regulator of Nonsense Transcripts 1 Homolog
UPR	Unfolded Protein Response
UTR	Untranslated Region
VAMP8	Vesicle Associated Membrane Protein 8
VCD	Viable Cell Density
WT	Wild-Type
XBP1	X-box Binding Protein 1
Xrn1	5'-3' Exoribonuclease 1
YY1	Ying Yang 1

1. Introduction and Literature Review

1.1. Introduction

1.1.1. What are Biopharmaceuticals?

Tracing its intellectual lineage back to therapeutics as early as the invention of the cowpox vaccine in the 18th century by Edward Jenner (Riedel, 2005), the use of biologically inspired pharmaceuticals is a powerful tool in the fight for human health. The nomenclature and definitions attached to this area of medicine are wide, with products sometimes interchangeably referred to as biologics, biopharmaceuticals, and biological medicines. Regulatory bodies tend to define biopharmaceuticals by their production, for example the European Medicines Agency (EMA), defining them simply as 'A medicine whose active substance is made by a living organism' (EMA, 2019). Others argue for a more encompassing definition, pointing out that products such as oligonucleotides and peptides, while synthetically made, are clearly biologically derived, and should be classified as such (Rader, 2008). Even laying this broader definition aside, drugs regulated as biopharmaceuticals are a very large field, accounting for 25% of products and 49% of sales in pharmaceuticals in 2017, projected to increase to 31% and 52% respectively by 2024 (Evaluate Pharma, 2018). In the United States, the biopharmaceutical sector accounts for 854,000 jobs, and 3.8% of total economic output (Thakor *et al.*, 2017).

The debate in nomenclature stems partially from the number of diverse biopharmaceutical products and fields. Most famous, for their stunning success in disease prevention, and in some cases extinction, are vaccines, which can take credit for the elimination of diseases such as polio, smallpox, and rubella as critical public health concerns (Stern and Markel, 2005). Another area, gene therapy is finally beginning to deliver on its initial promise (Naldini, 2015), with promising new technologies such as genome editing via CRISPR/*Cas9* (Long *et al.*, 2014; Maeder and Gersbach, 2016), and delivery by novel mechanisms, such as nebulisation and inhalation of mRNA (messenger ribonucleotide) (Patel *et al.*, 2019). Elsewhere, lyophilized gut microbiota are administered as 'live therapeutics', specifically supplementing microbe populations and biosynthetic pathways affected by disease (Delday *et al.*, 2018; O'Toole *et al.*, 2017; Raftis *et al.*, 2018; Yuille *et al.*, 2018). The field of cell therapy has varied applications,

such as treatment of Parkinson's disease (Fu *et al.*, 2015), and at the fulcrum between gene and cell therapy, CAR-T cell therapy, involving the removal, genetic engineering, and re-introduction of immune cells is a field generating significant excitement (Yip and Webster, 2018). However, in a prominent theme in biopharmaceuticals, CAR-T cell therapy is hampered by very high costs, which complicate the excitement based on its' efficacy (Leech and Dusetzina, 2018; Xie, 2018). Whilst acknowledging the scope and breadth of the broader biopharmaceutical field, this review and thesis will be focusing on therapeutic glycoproteins, and specifically their upstream production in Chinese Hamster Ovary (CHO) cells.

1.1.2. Therapeutic Glycoproteins

As vaccines can trace their lineage back to Edward Jenner, so too can therapeutic glycoproteins trace theirs back to Banting and Macleod, and the first successful purification and administration of insulin, to a diabetic patient in 1922 (Joshi *et al.*, 2007). Starting in 1982 with the release of Humulin, the first recombinant insulin drug (Kinch, 2015), the original paradigm of production (extraction and purification from animals) has been supplanted by the expression of recombinant proteins. This paradigm, in which cultured cells containing recombinant DNA are used to synthesize an extracted product, remains by far the dominant industry methodology, despite the emergence of alternative expression platforms, such as cell-free expression systems, transgenic plants, and 'pharming' from genetically modified animals (Kesik-Brodacka, 2018; Lalonde and Durocher, 2017).

As desire to produce more complex products arose, the correct folding and post-translational modification, such as glycosylation, of these proteins became paramount, in order that they be bioactive, long-lasting, and non-immunogenic in humans (Lalonde and Durocher, 2017). Thus, established first by the release of tissue plasminogen activator (t-PA), produced in CHO cells in 1987 (Jayapal *et al.*, 2007), mammalian cell cultures became the leading technology for the production of these glycoproteins.

1.1.3. CHO Cells

Multiple mammalian cell types are available for biopharmaceutical production, such as CHO, Human Embryonic Kidney (HEK-293), Myeloma, and Baby Hamster Kidney (BHK) (Jostock and Knopf, 2012; Sarantos and Cleo, 2013). Despite this breadth of available cell types, CHO cells are the dominant type used in this field, for example accounting for 61.5% of

biopharmaceutical production in the European Union (Sarantos and Cleo, 2013), and 75% of published research papers concerning biopharmaceutical production (Kunert and Reinhart, 2016).

Since they were first cultured in 1957 (Tjio and Puck, 1958), CHO cells have been a mainstay for mammalian genetic studies, sometimes referred to as the mammalian equivalent of the model bacteria *E.coli* (Jayapal *et al.,* 2007). Most modern CHO cell lines for industrial glycoprotein production can trace their ancestry back to the CHO-K1 proline auxotroph created in 1967 (Puck and Kao, 1967), and subsequent addition of dihydrofolate-deficient mutations, creating the DXB11 and DG44 cell lines (Urlaub and Chasin, 1980; Urlaub *et al.,* 1983), allowing for widespread use of selective media in CHO cell culture, which will be discussed later in this review.

CHO cells have a number of favourable characteristics, which have led to their dominance. CHO cells are able to grow in suspension media, making them ideal for industrial scaling. Few viruses can propagate in CHO cells, lessening risk of contamination. CHO cells can grow in serum-free, chemically defined media, contributing towards batch-reproducibility and regulatory compliance. The protein glycosylation pattern observed in CHO cells resembles that of human proteins, for instance with absence of the immunogenic α-galactose epitope (Lai *et al.,* 2013). Strategies such as Dihydrofolate Reductase (*DHFR*)-linked and more recently Glutamine Synthetase (*GS*)-linked gene amplification have also been developed in CHO cells, allowing for high cell-specific productivity (Gaughan, 2016; Noh et al., 2018). Finally, the highly developed platform of CHO protein production is recognised by regulatory bodies, and can therefore streamline regulatory approval (Kim *et al.,* 2012).

1.1.4. Industry Landscape and Challenges

For most of the history of mammalian cell-based glycoprotein production, the industry has been dominated by the 'blockbuster' model. Patent-protected biopharmaceuticals faced no competition, and could dominate markets. Humira, an anti-tumour necrosis factor alpha (TNFα) monoclonal antibody (mAb), has led worldwide sales since its' FDA approval in 2002 (Jayapal *et al.,* 2007), continuing with global sales of $11bn in 2013 (Barnard *et al.,* 2015), $16bn in 2016 (Moorkens *et al.,* 2017), and is projected to remain the world's highest selling biopharmaceutical until 2024 (Evaluate Pharma, 2018).

This landscape, however, is beginning to shift. A number of patents for high selling drugs, such as Remicade, are set to expire (Sarantos and Cleo, 2013). Biosimilars, duplicates of already approved drugs, with subtly different manufacturing processes, are being granted streamlined approval pipelines by regulatory bodies, to decrease costs and time to market, and to increase competition (Morrison, 2018). In the face of these changes, biopharmaceutical companies must diversify from this blockbuster modality. Companies are already being undercut by introduction of biosimilars, such as Sanofi introducing Admelog, a biosimilar of Eli Lilly's Humalog (Morrison, 2018). Humira is only projected to remain at the top of biopharmaceutical sales because of legal challenges delaying the release of Adalimumab biosimilars until 2023 (Evaluate Pharma, 2018), and its manufacturer AbbVie's reliance on the blockbuster model, as shown in Figure 1.1, potentially leaves it vulnerable to Humira being challenged on the market.

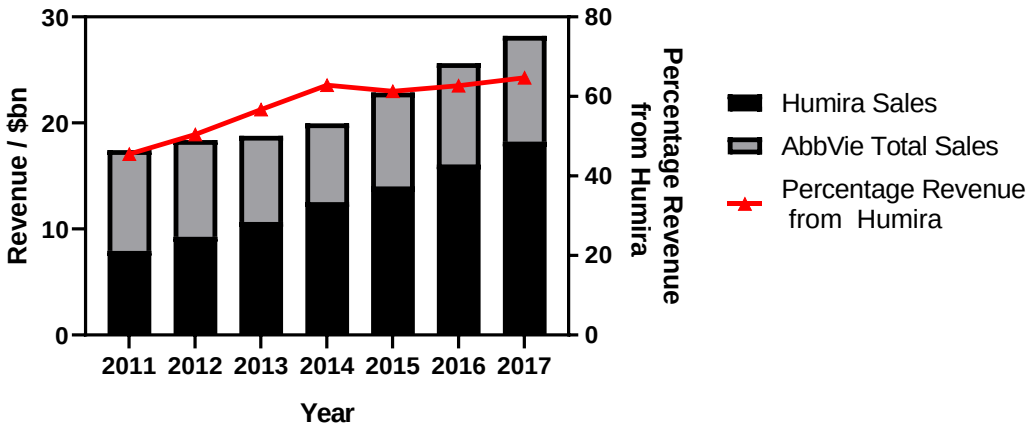


Figure 1.1 - Showing AbbVie's proportional reliance on its flagship blockbuster drug, Humira. The proportion of total revenue derived from sales of Humira has increased from 45.47% in 2011 to 64.67% in 2017, which may leave AbbVie vulnerable to being undercut by Adalimumab biosimilars. Data gathered from https://www.macrotrends.net/stocks/charts/ABBV/abbvie/revenue. bn: billion.

However, biopharmaceutical companies can neither financially, nor ethically, rely solely on these undercutting biosimilars to prosper. Orphan drugs, defined by the EMA as drugs that do not affect more than five in 10,000 people within the European Union, are already underrepresented in biopharmaceutical approvals, representing only 10 of 212 biopharmaceuticals approved by the EMA by 2013 (Sarantos and Cleo, 2013). Furthermore, with patient pressure to justify exorbitantly high prices for such rare drugs rising (Evaluate Pharma, 2018; Xie, 2018), biopharmaceutical companies cannot continue to rely on the general

price-inelasticity of drugs to recoup their costs. Relying entirely on production of biosimilars mimicking already released drugs ducks this ethical issue, and deprives patient groups of potentially effective therapies. Financially, due to their competitive nature, biosimilars cannot act as flagship drugs (Morrison, 2018), and companies with the healthiest outlooks are those which have a diverse pipeline of novel therapeutics (Evaluate Pharma, 2018; Moorkens *et al.,* 2017).

This innovation driven approach, though, is high in risk. Vector optimisation and cell line development for an industrial glycoprotein-producing cell line can take up to 20 weeks (Barnard *et al.,* 2015; Rajendra *et al.,* 2016), contributing to the cost of developing new biopharmaceuticals sometimes nearing $3bn (DiMasi *et al.,* 2016). Risk remains high even after production, where only an estimated 9.6% of biopharmaceuticals will make it from phase I trials to regulatory approval (David W. Thomas *et al.,* 2016). Some progress is being made, with a record 49 biopharmaceutical approvals in 2017, back up from a dip in 2016, although 5 of those approvals were for biosimilars (Morrison, 2018). However, the biopharmaceutical industry remains a highly volatile one, at constant risk of making losses, due to the idiosyncratic risk of R&D failure. In fact, even the risk that has been considered systematic in biopharmaceutical industry is largely composed of the anticipated and systematised losses from this stochastic R&D leverage (Thakor *et al.,* 2017).

Put briefly, the challenge facing the biopharmaceutical industry is to continue to innovate, investing in R&D in pursuit of a diverse, novel pipeline, and to simultaneously find ways to mitigate the cost, time, and risk associated with its current R&D practices.

1.2. Cell Line Development

Once a therapeutic glycoprotein has been identified and designated for development, for instance after humanization of a clinically relevant mAb (Gaughan, 2016), or design of next-generation antibody-drug conjugate (LoRusso *et al.,* 2011), it moves into the process of cell line development (CLD). This is the process by which a product is taken from a gene to a manufacturing cell line suitable for clinical and market supply, which has been extensively screened and selected for critical parameters most relevant to biopharmaceutical production: high production levels and cell growth, amenability to scale-up, correct post-translational

modifications, and good safety profile (Rita Costa *et al.*, 2010). This is an expensive and time-consuming R&D process, historically taking up to a year (Jayapal *et al.*, 2007; Lai *et al.*, 2013), although this has since improved to general timescales of around 6 months (Jostock and Knopf, 2012; Rajendra *et al.*, 2016). The archetypal CLD process will first be discussed, followed by the advances which have led to this progress, and the inefficiencies that remain.

1.2.1. The Archetypal CLD Process

The archetypal CLD process can be broken down into 8 steps:

1. An expression plasmid containing the coding sequence (CDS) of interest and a selection maker is delivered to the host cell nucleus. This can be performed by a variety of methods, such as lipofection or electroporation. The vector is integrated randomly into the host cell chromosomes, and can land in transcriptionally active or inactive sites.
2. The stably transfected cells are grown in selective media. For instance in the case of DHFR-deficient cells, cells are grown in medium deficient in glycine, hypoxanthine and thymidine, for which they are auxotrophic. Linkage of a *DHFR* gene with the vector allows selection of successfully transfected cells. *GS*-based selection can also be used, although this differs from DHFR selection in that most industrial CHO cell lines, such as DG44 are not *GS*-deficient, necessitating selection by inclusion of methionine sulfoximine (MSX), a GS inhibitor, in the medium (Jostock and Knopf, 2012). However, GS-deficient cell lines exhibiting high selection stringency have recently been developed to overcome this problem (Noh et al., 2018). Selection can also be mediated through antibiotic agents, such as a hygromycin resistance gene.
3. Selected cells are scaled-up and recovered.
4. Amplification is performed, in which an inhibitor of a vector-linked selection gene is added in increasing quantities, in order to encourage replication of the vector across the genome. Two options are commonly used for this step: *DHFR* and more recently *GS* (Lai *et al.*, 2013). The two systems are similar, in that they use an inhibitor of an essential biosynthetic pathway, methotrexate (MTX) for *DHFR* and MSX for *GS*, for amplification, as shown in Figure 1. *GS*-linked amplification tends to require fewer steps in concentration, shortening CLD timelines (Noh *et al.*, 2018).

5. A number of single-clone populations in the order of ~10^2 are isolated from these heterologous pools of amplified cells, and screened for favourable production characteristics in multi-well plates.
6. A number of clonal populations in the order of ~10^1 are chosen and taken forward through expansion.
7. Each clone is assessed in lab-scale bioreactors, designed to replicate conditions in industrial production, and screened again for favourable characteristics.
8. A 'winner' clonal cell line is selected, and cryopreserved in cell banks, to be revived for industrial production (Jayapal *et al.*, 2007).

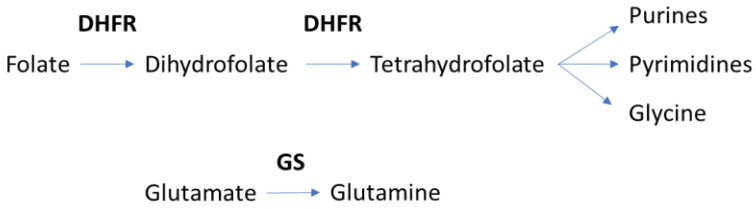


Figure 1.2 - Showing the two biosynthetic pathways and their essential products most often targeted by selection and amplification steps in CLD.

1.2.2. Developments in the CLD Process

Transforming yields of therapeutic glycoproteins from milligram to gram per litre scales (Jayapal *et al.*, 2007), whilst retaining the same generalised process, CLD has seen a great deal of development (Kim *et al.*, 2012). Acknowledging the important role that advancements in extrinsic process factors has played, such as with development of culture media (Kunert and Reinhart, 2016; Kuo *et al.*, 2018; Sellick *et al.*, 2011) and high-throughput cell-screening technologies (Kim *et al.*, 2012; Lai *et al.*, 2013; Nielsen and Borth, 2015), this review will focus on development of factors intrinsic to the molecular biology of the cell expression system.

During this review of developments in the CLD process, molecular biology tools will be assessed by several desirable attributes for efficient, rationalised CLD, which for the purposes of this thesis will be defined here:

- **Synthetic** will refer to manmade, as opposed to natural, molecular biology tools. It should be noted that very few molecular biology tools are 'truly' synthetic, i.e. designed from scratch with no basis in natural elements. With this in mind, care should be taken to understand which tool is being referred to as synthetic. For example, a 'synthetic' expression cassette could be constructed by artificial arrangement of 'natural' vector elements, such as promoters and terminators.
- **Predictable** will refer to tools whose effect in an experimental setting can be accurately user-defined and modelled beforehand, as shown in Figure 1.3a.
- **Titratable** will refer to tools whose effect can be predictably titrated across a range of strengths, as shown in Figure 1.3b.

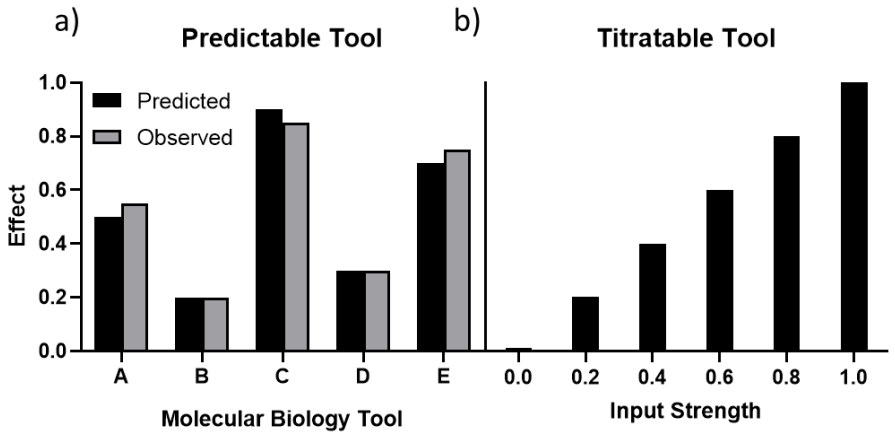


Figure 1.3 – Showing the generalised desirable qualities of synthetic molecular biology tools. Predictable tools, shown in 3a, have user-defined functionalities that accurately predict their observed experimental effect. Titratable tools, shown in 3b, have a broad range of predictable effect strengths, controlled by their input strength.

1.2.3. Vector Engineering

Taking the basal glycoprotein expression vector as a design space, containing a selection marker for molecular cloning via *E.coli* (e.g. ampicillin or kanamycin resistance), a marker for selection and amplification in CHO cells (e.g. *DHFR*, *GS*), and a cassette expressing the desired product, multiple improvements have been made, both adding new elements, as well as controlling and optimising those already in place, summarised in Figure 1.4.

1.2.3.1. Promoters

As shown by experiments measuring significant changes in titre and expression characteristics of GFP (green fluorescent protein) and EPO (erythropoietin) (Wang *et al.*, 2017),

promoters exercise a significant degree of control over the expression of recombinant proteins. Thus, predictable control of their transcriptional strength is desirable, not only for its maximisation, but also in cases such as mAbs, where two polypeptide chains have to be transcribed at a precise stoichiometry for optimal expression (Pybus *et al.*, 2014). For strong transcription of therapeutic glycoproteins, the most commonly used promoter is the human cytomegalovirus major (CMV) immediate early gene promoter (Wang *et al.*, 2017), followed by the elongation factor 1-alpha (EF1α) and simian virus 40 (SV40) promoters, the latter of which has a lower transcription activity than the other two, but may maintain its expression more robustly over long-term cultures (Ho *et al.*, 2014; Wang *et al.*, 2016).

Promoters commonly used in the biopharmaceutical industry can generally be divided into two regions: the core, containing the transcription start site, and the upstream proximal promoter region, containing transcription factor recognition elements (TFREs) (Lenhard *et al.*, 2012). Work towards creating synthetic promoters with differential strengths has largely focused on this proximal promoter region. For example, in Schlabach *et al.*, 2010, a library of every possible DNA 10-mer, concatemerised to 100 base pairs (bp), was screened for its ability to upregulate the transcriptional activity of a CMV core promoter when placed upstream. By combining activating 10-mers together, the transcriptional activity of wild-type (WT) CMV could be doubled in some cell lines. Crucially though, this change in transcriptional strength was neither predictable, nor titratable (Schlabach *et al.*, 2010).

Since this, an approach to design through modular placement of TFREs along a synthetic proximal promoter has yielded greater such control (Brown and James, 2015). In the first tranche of such work, TFREs known to be active in CHO cells were first concatemerised upstream of a CMV core promoter, then randomly combined, before expression data from these constructs was used to create a final library of 44 proximal promoters, showcasing a broad, titratable range of transcriptional strength, up to double the activity of the WT CMV promoter. These promoters were shown to be consistent across multiple cell types, and across different transient culture lengths (Brown *et al.*, 2014a). Further mechanistic understanding of these synthetic promoters has since been achieved, firstly by characterisation of TFREs in the CMV proximal promoter, in particular the functionally critical NF-κB and CRE elements (Brown *et al.*, 2015). Secondly, highly transcriptionally active TFREs have been screened for

their context and spacing independence, allowing for creation of synthetic, predictable, titratable promoter libraries with transcriptional strengths over two orders of magnitude, as well as design of synthetic promoters *in silico* with truly modular, predictably titratable elements (Brown *et al.*, 2017). Finally, omics datasets from different culture stages and conditions have allowed discovery of previously unidentified highly active TFREs, and creation of synthetic promoters insensitive to changing culture conditions, such as hypothermia (Johari *et al.*, 2019).

1.2.3.2. Codon Optimisation and Translational Control

Since mRNA codons are a degenerate code for the amino acid residues they recruit, myriad different CDSs can be used to produce the same protein. Codon optimisation refers to the practice of designing an mRNA CDS that will most efficiently be translated into protein. Most commonly, this involves the adaptation of transcripts to fit the relative abundance of donor transfer RNAs (tRNAs) within the particular cell line of production, ensuring an optimal rate and accuracy of translational elongation (Hung *et al.*, 2010). In Kotsopoulou *et al.*, 2010, it was found that titre of a mAb could be significantly increased by optimisation of the heavy chain transcription codon adaptation index (CAI), a measure of the proportion of codons within a transcript that are 'optimal' within the codon bias of the cell (Kotsopoulou *et al.*, 2010). Through this optimisation of codon adaption, the half-life of cytoplasmic mRNA is also increased (Presnyak *et al.*, 2015), due to the emerging mechanistic link between slow ribosome translocation and targeted mRNA degradation (Bicknell and Ricci, 2017; Łabno *et al.*, 2016). Other methods of codon optimisation have been developed, such as minimisation of secondary structure elements in the mRNA, and maximisation of GC content, although even a multiparameter optimisation of these factors, utilising a sliding scale along an mRNA, was unable to significantly outperform the holistic optimisation of CAI (Brown *et al.*, 2019).

Besides design of the CDS for elongation efficiency, relatively few options are available for controlling mRNA translation rate. Translation initiation, rather than elongation, is usually the rate limiting step of the process (Aitken and Lorsch, 2012). The near-ubiquitous Kozak consensus sequence (Kozak, 1987) is commonly used to enhance this rate, and some tools are available to predict the translation initiation rate of transcripts in *E.coli* (Reeve *et al.*, 2014), but no predictable or titratable control has yet been exercised in CHO cells. As such, and in

contrast to other elements, very few synthetic elements for the 5′ and 3′ untranslated region (UTR), which control translation initiation of the CDS, are available for vector design (Brown *et al.*, 2019).

1.2.3.3. Signal Peptides

A crucial, sometimes rate-limiting, stage in the expression of secreted recombinant proteins is the translocation of the nascent polypeptide to the endoplasmic reticulum (ER), through recognition of the signal peptide (SP), by the signal recognition particle (SRP), for the purposes of protein folding and secretion. For example, (Kober *et al.*, 2013), showed that a panel of natural SPs fused to both polypeptide chains of a mAb led to fold change in cell-specific productivity over an order of magnitude. Some progress towards synthetic SPs has been made, for instance with the creation of Secrecon, a SP optimised for expression by a hidden Markov model, discriminating for SP-like features (Barash *et al.*, 2002), which remains one of the highest-performing signal peptides to date (Brown *et al.*, 2019; Güler-Gane *et al.*, 2016). Some further mechanistic understanding has since been gained, such as the introduction of one and two alanine residues to the end of the signal peptide slightly improving protein titre (Güler-Gane *et al.*, 2016), and the ability to discriminate *in silico* between signal and non-signal peptides (Petersen *et al.*, 2011). However, no comprehensive explanation of cell line to cell line or product to product variation has yet been made, and no generally applicable predictable design principles have yet been produced. The best available tool is currently to screen a large library of randomly generated SPs, choosing those predicted to perform best. These SPs can slightly, if not significantly, outperform Secrecon, and their strengths are neither predictable nor titratable (Brown *et al.*, 2019).

1.2.3.4. IRES

Classically, pre-initiation complexes (PICs) recruit mRNAs for translation via their 5′ methyl-guanosine cap, in a mechanism which will be further discussed later in this review (Aitken and Lorsch, 2012). However, cis-elements, such as the internal ribosome entry site (IRES) enable binding of the PIC for translation inside the mRNA transcript (Hinnebusch *et al.*, 2016), through recruitment of initiation factors by a diverse family of secondary structure elements (Leppek *et al.*, 2018; Martinez-Salas *et al.*, 2018). In a recombinant production setting, these IRES features allow for creation of multi-cistronic mRNA transcripts, from which multiple

polypeptides can be translated. These IRES-mediated vectors have become a common tool in biotechnology studies (Dreesen and Fussenegger, 2011; Ho *et al.*, 2014; Wang *et al.*, 2017, 2018). In one study, an IRES-mediated tricistronic expression cassette was constructed, containing a mAb light chain (LC), heavy chain (HC), and selection gene. Lower mAb titres were achieved in transient transfection, compared to a multi-promoter plasmid control, since IRES sites perform less efficient PIC recruitment than 5'caps. However, stably transfected pools showed the tricistronic vector achieving markedly higher titres than the control. This was partially because polypeptide stoichiometry could be controlled by order of the CDSs inside the transcript. More importantly, since the LC and HC CDSs could no longer become genetically unlinked from the selection marker, or one another, a far greater proportion of clones that survived selection were producing both the LC and HC polypeptide, compared to the standard multi-promoter design. Thus, IRES systems can reduce the number of low/non-expressing clones in transfected populations, prevent product dislinkage from the selection marker, and allow for defined stoichiometric control over multiple polypeptides, though this can come at the cost of a reduced maximal titre (Ho *et al.*, 2012). A modicum of predictable, synthetic control has since been exercised over IRES, such as by mutation of functionally critical AUG triplets in the commonly used Encephalomyocarditis virus (EMCV) IRES, to create a library of 24 differential-strength sites, which were validated as consistent across multiple cell lines, and in their ability to produce LC and HC polypeptides at differing stoichiometries. Notably, none of these were able to increase translation efficiency, and therefore expression levels, compared to the wild-type IRES (Koh *et al.*, 2013).

1.2.3.5. Selection Marker Attenuation

The amplification step of CLD is predicated on the expectation that a higher concentration of selective agent (e.g. MSX, MTX) will necessitate cells increasing the expression of their selection marker to survive, thereby concurrently increasingly the expression of the genomically or cistronically-linked gene of interest. By attenuating the selection marker, making it less effective or more difficult to express, the cell is forced to amplify it to a greater extent, concomitantly increasing the gene of interest's expression to the same degree (Lai *et al.*, 2013). Various strategies of attenuation have been performed. In one study, a *DHFR* selection marker under the control of a strong Ef1α promoter was codon-deoptimised by minimisation of its CAI, leading to an up to 2-fold increase in the expression of a fusion

reporter protein, with a positive correlation observed between the extent of codon-deoptimisation, selection stringency, and resultant fold change in yield (Westwood *et al.*, 2010). In another, it was shown that expression of a recombinant gene could be attenuated by inclusion of negatively-regulated micro RNA (miRNA) binding sites in its 3'UTR, and that attenuation of a *DHFR* selection marker by this method led to an increase in mAb production after selection in 10nM MTX (Jossé *et al.*, 2018). Combination of selection marker attenuation with IRES expression systems is popular, as the relative resistance of IRES cassettes to selection marker dislinkage and gene fragmentation make them ideal systems for a tool predicated on linkage between the selection marker and gene of interest (Chin *et al.*, 2015; Ho *et al.*, 2012; Ng *et al.*, 2012). This effect has been predictably titrated, with differentially attenuated IRES sites placed upstream of a variety of selection markers leading to a broad range of resultant mAb titres from transfected stable pools of cells (Yeo *et al.*, 2017). One disadvantage of utilising selection marker attenuation is that effectively increasing the stringency of selection intrinsically means that fewer cells will survive the process (Westwood *et al.*, 2010). Though the resultant surviving cells may be express the gene of interest more highly, this reduction in selection survival may complicate clonal isolation and expansion, and reduce the overall expression level of heterologous transfected pools if the stringency is too high (Yeo *et al.*, 2017).

1.2.3.6. Chromatin Modulating Elements

In the archetypal CLD process, the glycoprotein expression vector is randomly integrated and dispersed through the host genome. This means it can land in either transcriptionally active euchromatin sites, or inactive heterochromatin sites, leading to reduced expression, and disrupting the correlation between gene copy number and expression levels required for predictable amplification (Lai *et al.*, 2013). Moreover, regulation of chromatin state is a dynamic system, and even transcriptionally active cassettes can become silenced over time by methylation (Yang *et al.*, 2010), especially when selection pressure is removed, as is often required for industrial production (Saunders *et al.*, 2015). This problem can be approached by the non-random integration of the vector into a predetermined genomic site, which will be discussed in section 1.2.7. of this review, or by the inclusion of cis elements in the vector to control the surrounding chromatin state. A number of different chromatin modulating elements are used for this purpose, which largely fall into two functional groups: those which

function as borders, preventing the spread of the surrounding chromatin state into the vector, and those which actively function through dominant chromatin remodelling mechanisms (Neville *et al.*, 2017).

The most commonly used element from the first category is the matrix attachment region (MAR): a sequence of DNA that acts as an attachment point to anchor chromatin to the nuclear matrix during interphase, and can prevent spread of heterochromatin (Zhao *et al.*, 2017). Introduction of MARs into vectors can sometimes yield mixed results, for example in one study, where popular MARs derived from chicken lysozyme and human interferon β were used to flank a GFP reporter under the control of SV40, EF1α, and CMV promoters. Both MARs were able to increase GFP expression in stably transfected pools, but only with the SV40 promoter, the weakest of the three. Furthermore, this MAR-SV40 construct retained significantly less of its GFP expression compared to an SV40 control after 8 weeks of culture in non-selective medium (Ho *et al.*, 2014). More positive results have since been found, such as flanking combinations of interferon β and *β-globin* MARs increasing expression of GFP in stably transfected pools, with both SV40 and CMV promoters, whilst also demonstrating strikingly higher numbers of successfully transfected clones, a higher proportion of positive GFP expressors within those clones, and higher retention of GFP expression than controls lacking a MAR (Zhao *et al.*, 2017). Development of MARs in a recombinant production setting is continuing, for instance with the recent discovery of a new MAR, that with a CMV has been shown to increase GFP expression 4.5-fold, with a markedly higher expression retention after 40 generations than non-MAR controls (Tian *et al.*, 2018).

In contrast to MARs, ubiquitous chromatin opening elements (UCOEs), characterised by methylation-free CpG islands, dominantly remodel chromatin to an active state, and do not need to flank both ends of an expression cassette. Discovered in the human HNRPA2B1-CBX3 locus, A2UCOE is the most commonly utilised variant, with fragments of 8, 4, and 1.5 kilobases (kb) conferring chromatin remodelling upon a construct when paired with a CMV promoter (Neville *et al.*, 2017). A series of studies pairing the 8kb A2UCOE initially with a GFP-expressing vector have well defined their attributes. Stable transfection of CHO-DG44 cells led to a higher average level of GFP expression than non-UCOE controls. Although UCOE-containing transfected cells could then be amplified using a higher initial

concentration of MTX (250nM), UCOE cell lines had lower expression than controls after this amplification process, whilst showing similar retention of expression with and without continued selection. Notably, this small difference in expression (<1.5 fold) occurred despite a marked decrease in gene copy number (>6 fold), implying more efficient transcription at integration sites (Betts *et al.*, 2015). Improved results were then found in a similar system with an EPO reporter, with UCOE cell lines maintaining an increased titre after amplification, once more despite a markedly lower gene copy number, whilst simultaneously showing a stronger correlation between copy number and titre, indicating the abolition of position-mediated effects (Betts and Dickson, 2015). Finally, it was observed that selected UCOE cell lines contained vectors integrated in a more dispersed manner across their karyotype, again suggesting abolition of position-dependent effects (Betts and Dickson, 2016).

A modicum of cross-element comparison has been made between chromatin modulating factors, such as a comparison of four elements, each with various positional variants, for production of a mAb in CHO cell culture. Of the four, including a MAR element, the 1.5kb core A2UCOE element performed the best by far, boosting expression by 6.5 and 7.5 fold in stable pools and clonal cell lines respectively, and demonstrating greater retention of expression after 120 generations than controls (Saunders *et al.*, 2015). Finally, some titratable synthetic systems utilising chromatin modulating elements have been developed, such as the integration of a minimised production-enhancing DNA element into an accumulative site specific integration system, demonstrating predictable increases in expression, and a clear correlation between gene copy number and recombinant protein production (Kawabe *et al.*, 2017).

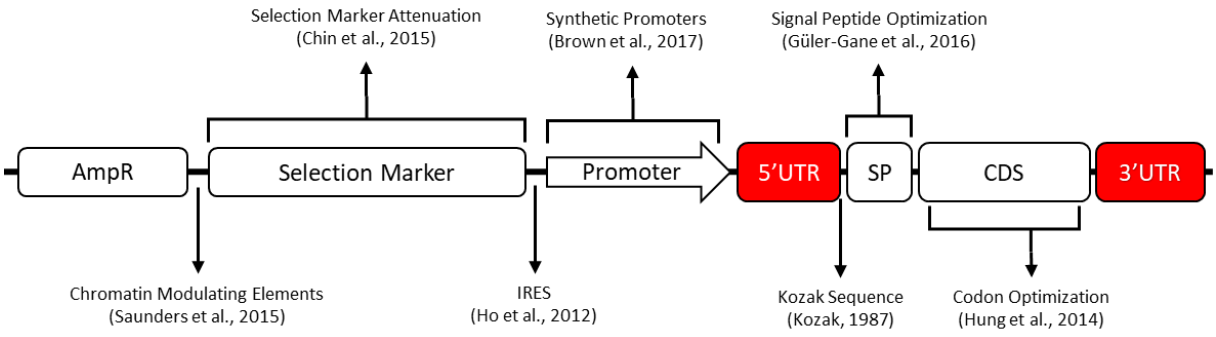


Figure 1.4 - Showing the basal therapeutic glycoprotein vector, and the developments that have been made to each of its' constituent parts. Regions critical to expression, the 5' and 3'UTRs, for which no synthetic systems of

control in glycoprotein-expressing CHO cells have been implemented, are highlighted. AmpR: ampicillin resistance; UTR: untranslated region; SP: signal peptide; CDS: coding sequence.

1.2.3.7. MPRA

As stated in Figure 4, no synthetic vector elements have yet been implemented in the 5'UTR and 3'UTR for industrial glycoprotein expression in CHO cells. However, the technology of massively parallel reporter assays (MPRAs) is rapidly becoming more established and could represent a powerful solution for the rational engineering of 5' and 3' UTRs for industrial control of biopharmaceutical expression. Made possible by advances in high-throughput oligonucleotide synthesis, sequencing, and reporter technologies, MPRA is an experimental method, involving the parallel testing of thousands of genetic variants. An example MPRA assay might involve the production and transfection of a library of reporters with thousands of randomly generated 3'UTR sequences, before analysis by deep sequencing, inferring from their relative abundance the effect of each sequence feature on mRNA accumulation (Melnikov *et al.,* 2014). In one example, a library of rationally designed 3'UTRs enabled the dissection of 3'UTR regulatory sequences into three categories of regulation, and the rational design of 58 constructs, whose stability to could be accurately predicted by regression analysis of the original dataset (Rabani *et al.,* 2017).

In terms of 5'UTR design, several studies have been carried out. For example, by partnering a library of 280,000 random 5'UTRs with polysome profiling in HEK-293T cells, the effect of various sequences on translational activation was analysed, facilitating 5'UTR design by a deep-learning algorithm, accurately prescribing different levels of polysome loading (Sample *et al.,* 2019). In another study, a library of all possible translation initiation site variants (from -6 to +5, 65,536 sequence variants), was partnered with fluorescence-associated cell sorting (FACS) to model and accurately predict genome-wide translation initiation efficiencies of human mRNAs (Noderer *et al.,* 2014). In a similar study in yeast, 2,000 sequence variants of the 10bp upstream of the translation start site were used in yeast, to inform a model which could predict 70% of the variation in translation initiation across 5'UTRs, including those with nucleotide compositions not included in the original dataset (Dvir *et al.,* 2013).

Similar studies have been carried out in the 3'UTR. In one study, 13,000 rationally designed 3'UTR variants were screened for their ability to effect expression dynamics in yeast, giving the observation that most mutations had minimal effect on expression, and allowing the

identification of a critical TA-rich element, which could then be molecularly characterized (Shalem *et al.*, 2015). In another study for functional annotation, >450kb of 3'UTR sequence from human genes were screened with a GFP reporter, allowing for the discovery of 87 novel 3'UTR-regulatory elements (Zhao *et al.*, 2014). Finally, a cell-based expression screen was used in one study to identify 8mers associated with regulation of protein expression. This facilitated the discovery of hundreds of mostly expression-activating sequences, which was a surprising result in the 3'UTR, a feature mostly associated with negative regulation (Wissink *et al.*, 2016).

As demonstrated above, the MPRA methodology represents a potentially powerful tool in the functional dissection, and thereby the design, of 5' and 3'UTRs for biopharmaceutical expression. The potential of applying sequences discovered by this platform to vector design, specifically sequences discovered by Oikonomou *et al.*, 2014, will be discussed and investigated in chapter 4 of this thesis. The wholesale application of MPRA technology to each new industrial product and vector, as part of the CLD process could be a potent tool to rationalise vector design. Its' implementation would also depend on the development of timescale and resource-appropriate CHO MPRA platforms, which may be a valuable investigation to undertake in future.

1.2.4. Transient Gene Expression

As the isolation of clonal high-producing cell lines is a time and labour intensive process, other methods, primarily transient gene expression (TGE), are often used to produce smaller quantities of recombinant protein for pre-production processes (Lalonde and Durocher, 2017). Quantities of protein on the g/l scale can be produced by a process spanning as little as 7 days (Barnard *et al.*, 2015; Daramola *et al.*, 2014), in contrast to months of CLD, giving an indication of steps to improve their eventual stable manufacturability (Mason *et al.*, 2012), and providing material for pre-clinical studies, such as toxicology investigations, all without confronting the potentially confounding effect of clonal cell line variation (Pybus *et al.*, 2014).

TGE comprises the delivery of intact plasmid DNA (pDNA), the basal vector containing only the expression cassette and *E.coli* selection marker, into a heterogenous pool of cells. This pDNA is harboured episomally in the nucleus, but not integrated into host chromosomes, and is transcribed, and thereby expressed, from this episomal space (Geisse and Fux, 2009). Historically, TGE has mostly been performed in HEK-293 cells, due to their higher yields and

earlier development of stimulatory systems such as *EBNA*-linked replication than CHO cells. However, more development has recently been performed on transient CHO systems, to ensure maximal relevance of the material gathered to stable production (Hacker *et al.,* 2013), and with developments, CHO cells have been shown to produce higher titres of industrially-relevant molecules such as mAbs, than HEK-293 cells (Zhong *et al.,* 2018). A variety of delivery methods are available, the three most widely used systems being lipofection, electroporation, and polyethyleneimine (PEI) – mediated transfection. Transfection efficiency, a measure of the percentage of cells expressing the transfected gene, is notably higher for lipofection and electroporation, with reported efficiencies as high as 93.65% (Wang *et al.,* 2018) and 96% (Steger *et al.,* 2015) in CHO cells respectively, with PEI-mediated transfection usually achieving efficiencies of 50-60% (Hacker *et al.,* 2013), and up to 80% (Rajendra *et al.,* 2015a). However, due to ease of scale-up compared to electroporation, a significant cost-reduction compared to lipofection, and optimisation of transfection protocols, PEI is the most common method of pDNA delivery for larger-scale industrial TGE (Gutiérrez-Granados *et al.,* 2018).

A significant limitation of basal TGE is that episomal DNA is not replicated upon cell division, leading to dilution of pDNA from cells, as cultures grow and divide. A solution to this problem that has been implemented in HEK-293 cells for a number of years is Epstein-Barr virus (EBV) nuclear antigen 1 (*EBNA-1*)-based stimulation of plasmid replication. *EBNA-1* is a gene that stimulates replication of plasmids carrying the EBV origin of replication, *oriP*. Transfection of pDNA containing *oriP* into a cell line stably expressing the *EBNA-1* gene leads to replication of the pDNA with cell division, maintaining its levels inside cells as cell density grows (Carpentier *et al.,* 2007; Yates *et al.,* 2002). A major development for TGE in CHO cells was the implementation of this system. Stable integration of *EBNA-1* into CHO cells led to significant increases in transient titre of a mAb when transfected in a vector containing *oriP* using PEI. Titre was further increased by the stable cotransfection of *EBNA-1* and *GS*, allowing for MSX-mediated selection. Combined with the optimisation of the PEI transfection protocols, and of the culture platform and feeding regime, transient mAb titres of 2g/l could be achieved (Daramola *et al.,* 2014).

Through extensive characterisation of TGE in CHO cells, further cell-line adaptations have sought to overcome bottlenecks identified at the transcriptional and post-transcriptional steps

in expression (Mason *et al.,* 2012; Rajendra *et al.,* 2015a). For instance, stable cotransfection of the transcription factor X-box Binding Protein 1 (*XBP1*) has been shown by multiple studies to increase TGE titre of varied reporter proteins (Brown *et al.,* 2019; Cain *et al.,* 2013; Johari *et al.,* 2015; Pybus *et al.,* 2014; Rajendra *et al.,* 2015b). A number of other proteins, such as Activating Transcription Factor 6 (*ATF6*) and the chaperone Binding Immunoglobulin Protein (*BiP*) (Brown *et al.,* 2019; Johari *et al.,* 2015; Pybus *et al.,* 2014), have also been identified for this purpose. In the converse methodology, knockout of two apoptotic-factor genes (*Bax* and *Bak*), has been shown to create a cell line capable of a 3-4 fold increase in TGE yield of a mAb, mediated through an increase in cell density and viability, and increased uptake of pDNA (Macaraeg *et al.,* 2013).

A number of advances in transfection and culture technique have also been made, and optimised protocols for TGE in CHO cells published (Hacker *et al.,* 2013). Further improvements are being made, such as the development of a high cell density PEI transfection system, allowing for >1g/l titres of a mAb (Rajendra *et al.,* 2015c), and subsequent pairing with a downstream purification system, enabling transient production of ~1mg quantities of mAb from a 24 deep-well plate (Barnard *et al.,* 2015). Development of other transfection methods continue apace, such as with the creation of a flow-electroporation based system, leading to ~1g/l mAb titres, in cell culture up to the litre scale (Steger *et al.,* 2015).

Some work has been performed for the predictable, titratable control of TGE. For example, in Johari *et al.,* 2015, a number of TGE-enhancing techniques (overexpression of chaperone proteins, chemical supplements, temperature shifts), were screened and combined, to create an optimised Biphasic 12-day production process, increasing titre of a difficult to express (DTE) fusion protein 6-fold. Similarly, in Brown *et al.,* 2019, different cell lines, overexpressed proteins, and vector features were combined, leading to a titratable range of up to 10-fold increase compared to controls of a Secreted Alkaline Phosphatase (SEAP) reporter.

1.2.5. Cell Chassis Development

Regardless of vector design and DNA delivery platform, the chassis through which therapeutic glycoproteins are ultimately manufactured and exported is the CHO cell. Therefore, control of their expression profile will ultimately be mediated through the cell's capacity to grow, survive, and express recombinant protein. As demonstrated by the

comparison of the typical maximum cell-specific production rate of CHO cells, 50-90 picograms per cell per day (pcd), and the output of secretory plasma cells producing Immunoglobulin M (IgM), up to 400pcd, this cell-chassis has not been maximally exploited, and thus much work has been invested to develop its abilities (Hansen *et al.*, 2017). The chassis is generally manipulated by either overexpression of favourable genes, or disruption of unfavourable genes, by knockout, or RNA interference-based methods (Fischer *et al.*, 2015). Many genes have been successfully targeted by this methodology, reviewed comprehensively by Fischer *et al.*, 2015, such as the translation-controlling Mammalian Target of Rapamycin (*mTOR*) (Dreesen and Fussenegger, 2011), and the transcription factor Yin Yang 1 (*YY1*) (Tastanova *et al.*, 2016). However, this review will focus on the two most commonly targeted functions: expression capacity, and anti-apoptosis engineering.

1.2.5.1. Expression Capacity Engineering

In CHO cell lines expressing recalcitrant recombinant proteins, dislinkage is often observed between heightened gene copy number, mRNA levels, intracellular polypeptide levels, and secreted protein titre, which fails to concomitantly increase (Chromikova *et al.*, 2015; Reinhart *et al.*, 2014). This indicates an expression bottleneck at the stages of protein folding, post-translational modification, trafficking and secretion. A key contributor towards this bottleneck is the unfolded protein response (UPR). Sensing unfolded protein accumulation via *BiP*, the UPR is a signalling cascade, which inhibits translation, degrades mRNA, and upregulates expression of chaperone proteins, in an attempt to bring levels of unfolded protein back to homeostatic levels. If unable to do so, the UPR will trigger autophagy or apoptosis (Hussain *et al.*, 2014). A cell's ability to upregulate UPR-responsive genes, thus mitigating these negative changes, has been shown to positively correlate with their ability to produce recombinant protein (Kober *et al.*, 2012). Accumulation of unfolded and misfolded protein in the endoplasmic reticulum also leads to ER-associated degradation (ERAD) of proteins by the proteasome, and aggregation of recombinant protein, further compromising expression (Zhou *et al.*, 2018).

One tactic to mitigate this response is to overexpress chaperone proteins, which bind to polypeptides and prevent the formation and aggregation of unfolded protein. The two most commonly targeted of these proteins are *BiP* and Protein Disulphide-Isomerase (*PDI*), and

overexpression of both has been shown to increase the expression of secreted proteins (Johari *et al.,* 2015; Pybus *et al.,* 2014), though for reasons that will shortly be discussed, this effect is not universal across cell lines and recombinant proteins (Mohan *et al.,* 2008). Another tactic is to overexpress proteins directly implicated in the UPR. *XBP1*, and its splicing variant *XBP1S*, are UPR-activated proteins, contributing towards endoplasmic reticulum biogenesis when overexpressed, and thereby increasing the capacity of the cell to process translating and folding protein (Hussain *et al.,* 2014). Overexpression of *XBP1*/*XBP1S* has been shown to increase recombinant protein titre in multiple studies, under both transient (Johari *et al.,* 2015; Pybus *et al.,* 2014; Rajendra *et al.,* 2015b) and stable expression (Cain *et al.,* 2013; Gulis *et al.,* 2014). *ATF6*, previously mentioned for its' ability to increase recombinant protein expression across multiple studies (Brown *et al.,* 2019; Johari *et al.,* 2015; Pybus *et al.,* 2014), is also a target of this strategy, being a UPR-inducible transcription factor associated with the expression of multiple foldases and chaperones. Finally, secretion factors can be overexpressed, such as Vesicle-associated membrane protein 8 (*VAMP8*) and Synaptosomal-associated protein 23 (*SNAP23*), which have been shown to increase production of a mAb, when overexpressed both episomally and stably, by altering the ratio of intracellular to extracellular (secreted) recombinant protein (Peng *et al.,* 2011). Overexpression of SRPs, primarily *SRP14*, has also been shown to increase titre of a difficult to express mAb, by a similar mechanism (Le Fourn *et al.,* 2014).

Despite substantial overlap in identified effector genes between different studies, very little predictable control over expression can be exercised by this method. For example, separate studies on the co-expression of *PDI* with a reporter protein have found changes in reporter specific productivity (qP) ranging from a 1.6-fold decrease, to a 1.37-fold increase. This is partially attributable to a lack of standardised methodology: differences in transfection method, origin of the effector gene (CHO, human, mouse), and effector gene dosage can all impact the resultant changes in expression (Hansen *et al.,* 2017). The main source of this unpredictability, however, is the genetic instability of immortalised cell lines. Because of this instability, pools of cells carry an intrinsically high genetic heterogeneity. When clonal cell lines are isolated from transfected pools, the cells may or may not be susceptible to manipulation of a certain effector gene, and may have high or low reporter expression for a number of genetic reasons. Therefore, to reliably test an effector gene's consequence on

expression, multiple clonal lines must be assessed as a population, or heterologous pools of transiently transfected cells can be used. In turn, these results may not be applicable to a particular isolated clonal cell line. Due to this seemingly intractable incongruity, effector genes must be assessed on a case-by-case basis, rather than in a systematic, predictable fashion (Hansen *et al.,* 2017).

1.2.5.2. Anti-Apoptosis Engineering

As the titre of a transfected culture is necessarily a factor of how many cells are producing protein for how long, substantial work has been invested into improving cell culture performance, particularly in delaying the cell death phase, through anti-apoptosis engineering (Kim *et al.,* 2012; Mohan *et al.,* 2008). This can be achieved through either overexpression of anti-apoptotic genes, or disruption of apoptotic genes. The two most common targets for the former approach have been the overexpression of B-cell lymphoma (BCL) proteins *BCL-XL* and *BCL2* (Fischer *et al.,* 2015), both of which have been shown to increase apoptosis resistance. This technique has also recently been advanced by the mutagenesis and directed evolution of *BCL-XL*, selecting for apoptosis resistance, to create a hyper-active form of the gene, conferring greater cell culture performance (Majors *et al.,* 2012). More novel targets for this strategy include Heat-shock protein 27 (*HSP27*), overexpression of which led to increased maximum cell density and apoptosis resistance by downregulation of various caspases (Tan *et al.,* 2015), and *mTOR*, which when overexpressed via an IRES cassette, led to increases in cell proliferation and viability (Dreesen and Fussenegger, 2011).

By far the most prevalent target of gene disruption for anti-apoptosis engineering are Bcl-2-associated X protein (*Bax*) and Bcl-2 homologous antagonist killer (*Bak*) genes, the knockout of which has been shown to increase cell culture performance and titre in both stable and transient expression systems (Fischer *et al.,* 2015; Macaraeg *et al.,* 2013). Comprehensive knockdown of apoptotic pathways has recently been made more tractable by the advent of CRISPR/*Cas9* gene editing technology, for instance in the single-step multiplexed disruption of *FUT8*, *Bax*, and *Bak*, leading to reduced apoptosis activity and increased productivity (Grav *et al.,* 2015).

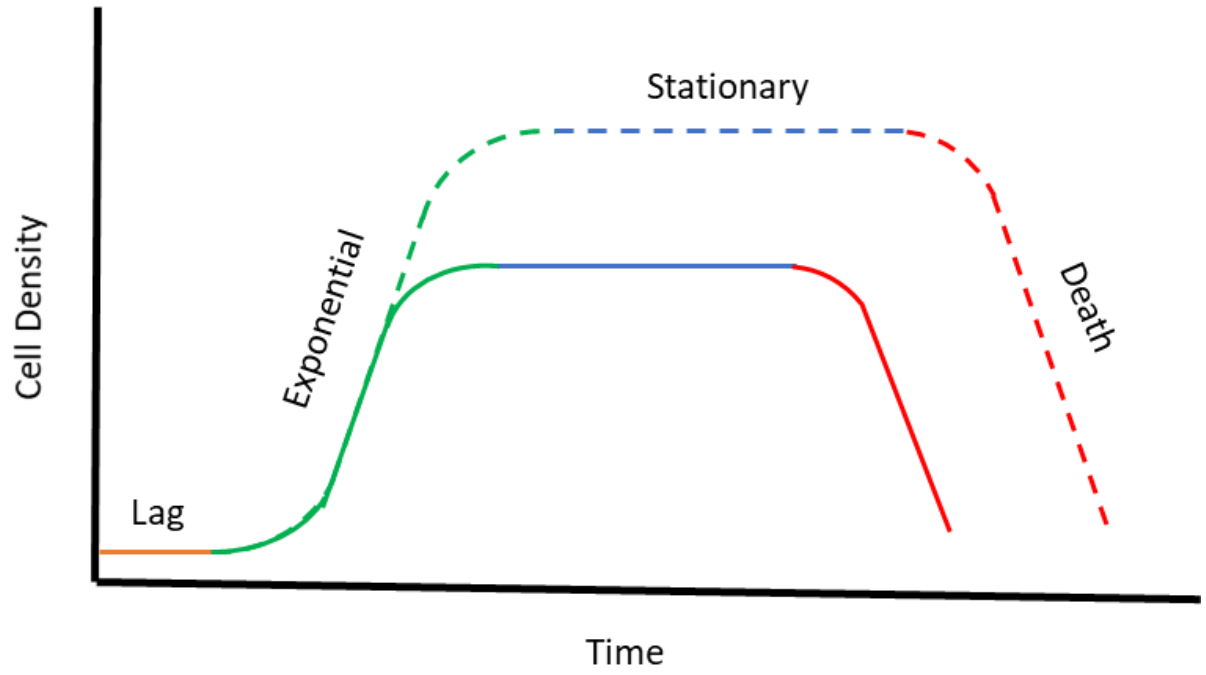


Figure 1.5 -Showing the generalised cell culture performance of non-engineered (solid line) and anti-apoptosis engineered (dotted line) CHO cells. The area under this curve, measuring the number of cell days in each culture is referred to as the integral of viable cell density (IVCD), and correlates with recombinant protein titre.

1.2.6. miRNA Engineering

miRNAs are small (~22nt) regulatory RNAs, which control the critical and energy-consuming process of translation. They are usually transcribed by RNA polymerase II, as either a dedicated pri-miRNA, or as part of an mRNA intron, a pri-mintron, before cleavage into a precursor-miRNA hairpin by either *Drosha*, or the spliceosome, respectively. After processing by *Dicer*, mature miRNAs bind *Ago2*, which can then bind mRNAs with regions of complementarity to the miRNA as part of the RISC complex, inhibiting translation, and upregulating deadenylation and decay of the mRNA (Hammond, 2015). As the obligatory region of complementarity is very small (~6nt seed sequence), miRNAs target multiple mRNAs, often within a functionally related family, making them attractive targets for cell engineering, since the effectiveness of single gene engineering is often mitigated by the genetic redundancy of CHO cells (Hackl *et al.*, 2012; Zhou *et al.*, 2018). Additionally, unlike protein-based cell engineering, they impose no translational burden on the cell (Hackl *et al.*, 2012; Stolfa *et al.*, 2018).

Identification of CHO miRNAs has been expediated by sequencing of the CHO genome (Xu *et al.*, 2011), in combination with other omics datasets, such as RNA-seq (Stolfa *et al.*, 2018). Public databases of these miRNAs are available, such as miRbase, which currently lists 351 mature CHO miRNAs, significantly less than the 2654 identified in the better annotated human genome, though this difference may be partially explained by genetic factors, such as the lower chromosome number in CHO cells (22 vs 46) (Fischer *et al.*, 2015). Target miRNAs can also be identified by omics datasets, such as by comparing the miRNA expression of high to low expressing cell lines (Maccani *et al.*, 2014). Numerous synthetic strategies exist to control miRNAs in CHO cells, such as transient overexpression by miRNA mimics, or the use of Antagomirs to inhibit endogenous miRNAs (Fischer *et al.*, 2015; Jadhav *et al.*, 2013).

In Fischer *et al.*, 2017, it was shown that transient transfection of mir-557 led to a significant, ~40% increase in product expression in seven established cell lines, producing a variety of products (mAbs, bispecific antibodies, scFv), amplified by both *GS* and *DHFR* methods, with both high and low levels of expression. A cell line stably overexpressing mir-557 was then created, and taken through a CLD process with a difficult to express mAb, against a control overexpressing a nontargeting miRNA. The mir-557 cell line produced higher-expressing

stable pools, a much higher proportion of high-expressing clones (41.5% vs 1.2%), and the top 3 selected clones produced 2-fold more of the mAb in a fed-batch culture, with no compromise in product quality (Fischer *et al.,* 2017). In an alternative approach to miRNA engineering, Jossé *et al.* introduced miRNA target sites into the 3' UTR of a *DHFR* cassette, as a means of selection marker attenuation (Jossé *et al.,* 2018). A great variety of miRNAs have been targeted as a means of engineering. Common strategies have been developed, including knockdown of miR-7a to reduce growth and enhance proliferation, and overexpression of let-7 to supress apoptosis (Fischer *et al.,* 2015; Jadhav *et al.,* 2013), rendering this synthetic tool somewhat, if not entirely, predictable.

1.2.7. Site-Specific Integration

In the canonical CLD process, the expression cassette is integrated randomly into the host-cell genome, before amplification, again into random genomic loci (Jayapal *et al.,* 2007). The underlying assumption of this methodology is that as gene copy number rises, expression levels will concurrently increase. However, due to a myriad of genomic factors, such as chromatin state and presence of upstream or distal enhancer regions, different integration sites in the genome show markedly different levels of cassette transcription. As such, integration and amplification steps are rendered heterologous and unpredictable, as the assumed correlation between copy number and expression levels does not always arise (Alves and Dobrowsky, 2017).

To circumvent this problem, various strategies have been developed to insert the expression cassette into transcriptionally active known genomic sites, or 'hotspots', ensuring more consistent transgene expression, and better correlation between copy number and expression. The first step in this process is to identify a genomic hotspot to be targeted. A number of traditional screening methods have been employed for this, such as in Cheng et al., 2016, where a GFP-expressing cassette was randomly integrated into mammalian HT1080 cells. After clonal isolation, the location of 10 high-producing clones was assessed by sequencing, identifying 8 distinct genomic sites. Site-specific integration of a *SEAP* reporter into one of these sites led to both more consistent and higher reporter expression than random integration. Such exhaustive methods may become unnecessary in the future, as the growing availability and depth of genomic, epigenomic, and transcriptomic datasets (to be discussed

later) begin to allow for *in silico* selection of hotspots (Alves and Dobrowsky, 2017; Stolfa *et al.*, 2018).

Following identification of an integration site, three leading options are available for targeted delivery of the cassette: site-specific recombination, genome editing, or transposase-mediated integration.

In site-specific recombinase systems, a recombination site in the genome and cassette are recognised and recombined by an enzyme. The two most common systems for this purpose are *Cre*/*loxP* and *Flp*/*FRT* (Lai *et al.*, 2013), both offering a highly specific integration event at a single predetermined site. The technology has been further developed by the creation of Recombination-Mediated Cassette Exchange (RMCE) systems, wherein a master cell line, containing a cassette flanked by recombination sites in a known genomic hotspot, is exchanged for the desired expression cassette. This cassette exchange can be used to screen for transfected clones, through removal of a fluorescent protein, or addition of an antibiotic resistance gene. Integration of cassettes at these sites has led to high expression, high expression stability across multiple generations, and better correlation between gene copy number and expression levels (Zhang *et al.*, 2015). The system has further evolved, with the development of an *Flp*/*FRT*-mediated binary RMCE system, comprising two independently integrated cassettes, with predictable, additive expression profiles, allowing for site-specific co-integration and expression of two reporters (Baser *et al.*, 2016). Similarly, an accumulative *Cre*/*loxP* integration system has been designed, whereby a number of independent RMCE events take place, as each successive cassette is flanked by a new distinct *loxP* site, specific for the recombination of the next cassette. This system allows for predictable and additive integration of up to 8 cassettes at a single genomic site, creating titratable control of expression (Kameyama *et al.*, 2010; Kawabe *et al.*, 2017).

One disadvantage of *Cre*/*loxP* and *Flp*/*FRT* technology is that integration events leave the recombination site unaltered, leaving open the possibility of reversibility, and that such constructs may experience genomic instability. Hence, the analogous Φ31 integrase system has been developed, utilising a viral integrase, and endogenous *attP*/*attB* sites within the genome, which are altered upon an integration event. This also allows for higher copy number

than recombinase-mediated systems, but at the cost of an inability to select a hotspot integration site (Alves and Dobrowsky, 2017; Campbell *et al.*, 2010).

The emerging technologies of genome editing are a potentially powerful tool for site-specific recombination. These technologies pair a DNA recognition mechanism with nuclease activity to perform site-specific double-strand breaks, and providing genetic material for homology-directed repair (HDR) or non-homologous end joining (NHEJ) allowing for concurrent knock-in of DNA (Gaj *et al.*, 2013). Optimised vectors for Transcription Activator-like Effector Nuclease (TALEN)-mediated genome editing have been developed, allowing for efficient integration and expression of a single-chain Fv-Fc (scFv-Fc) protein in CHO cells at a predetermined genomic site (Sakuma *et al.*, 2015). TALEN technology is limited its' mechanism of DNA recognition, by modular DNA-binding protein domains, the development and design of which is a difficult process. As such, more excitement is being drawn by the far more easily programmable CRISPR/*Cas9* genome editing system (Lee *et al.*, 2015a), which targets DNA via a small guide RNA (gRNA). Methods have been developed in CHO cells for the stable integration of expression cassettes, both by HDR (Lee *et al.*, 2015b, 2016b), and by NHEJ (Bachu *et al.*, 2015), both with levels of efficiency allowing for selection via one round of limiting dilution into a 96 well plate, resulting in a much more consistent expression level across multiple clones than in random integration (Lee *et al.*, 2015b).

Finally, transposon-mediated integration systems, such as sleeping-beauty and piggyBAC, are well established in molecular biology studies (Narayanavari *et al.*, 2017), but have only recently been applied to production of biopharmaceuticals. Promisingly, it has been demonstrated that stable cell pools generated using the piggyBAC system, as opposed to random integration, exhibited around 4-fold higher productivity of a panel of mAbs, up to an impressive 7.6g/l in a 16-day production process (Rajendra *et al.*, 2016). These encouraging results should merit further investigation of transposons for biopharmaceutical production.

1.2.8. Omics

'Omics' is an umbrella term, encompassing the various emerging techniques of gathering and analysing large datasets, broadly describing the state of a cell. A genomics dataset might be a sequenced genome, a transcriptomic dataset could be an RNA-seq file describing global RNA abundance, a metabolomic dataset could be a data-driven model of the metabolic flux

performed by a cell, and so on. Sequencing of the CHO-K1 genome (Xu *et al.*, 2011), combined with the advent of technologies such as RNA-seq, and mass spectrometry, and freely-available tools such as CHOgenome.org (containing an annotated CHO-K1 genome), CHOmine (a repository of publicly available CHO omics datasets), and CRISPy (a CRISPR/*Cas9* target finder for CHO-K1 cells), are facilitating the use of omics technology in biopharmaceutical research and development (Stolfa *et al.*, 2018).

The potential benefits of omics technology include the *in silico* screening of genomes for transcriptional hotspots, appropriate for site-specific integration of a transgene (Kuo *et al.*, 2018; Stolfa *et al.*, 2018), such as in Zhang *et al.*, 2015, where clonal cell lines exhibiting high product expression, expression stability over 100 generations, and strong correlation between copy number and expression levels, were created by insertion of a mAb-expressing cassette at *in silico* determined hotspots using RMCE (Zhang *et al.*, 2015).

Another application of omics technologies is to analyse the properties of cell lines with desirable production qualities, learning which attributes tend to make cells highly productive, and even utilising the power of emerging synthetic biology technologies, to apply these solutions with a more systems-based approach than had been attempted before, for example with the downregulation of an entire pathway via miRNA signalling (Fischer *et al.*, 2015; Kuo *et al.*, 2018), or multiplexed knockout of genes by CRISPR/*Cas9* (Grav *et al.*, 2015). Multiple studies with this premise have been published, such as the transcriptomic and proteomic analysis of cell lines with a wide range of productivities, allowing identification of genes most correlated with product secretion, which led to the counterintuitive insight that in some cases, intracellular expression of the product was not strongly linked to overall secreted productivity (Kang *et al.*, 2014). The CHO translatome has also been analysed, to identify genes most essential for cellular growth and proliferation, the data also suggesting that transcription and translation level are uncoupled for 95% of genes in the CHO genome (Courtes *et al.*, 2013). Omics data used to reconstruct the protein secretion pathway in mice has been successfully applied to CHO cells, predicted several novel targets for increasing secretion capacity (Lund *et al.*, 2017). Through use of omics datasets, all three of these studies offer striking insights to understanding glycoprotein production characteristics in CHO cells, unattainable by more conventional means. Omics datasets can also be used to inform more traditional molecular

biology practices, such as in the omics-based identification of a correlation between deletion of a telomeric region of CHO chromosome 8, and an increase in productivity (Ritter *et al.*, 2016a). Using this result to inform more traditional molecular biology, quantitative PCR (qPCR) and small interfering RNA (siRNA) screens were used to identify the causative *C12orf35* gene within this locus, and to show that its' knockdown reproduced the phenotype of increased productivity (Ritter *et al.*, 2016b).

Most ambitiously, some have aimed to use omics datasets, in combination with various tools for machine learning, to develop an *in silico* accurate, predictive model of CHO cell behaviour. This problem has already been approached on a slightly smaller scale, such as with reconstruction of the CHO protein secretory pathway (Lund *et al.*, 2017), and creation of a model of the CHO metabolome (Hefzi *et al.*, 2016). With such tools and the appropriate data, the application of synthetic biology parts to a certain bioproduction process could be made truly predictable and rational.

The major roadblock to this data-driven rationalisation of CHO-cell engineering remains the relative paucity of available datasets, for example with only a handful of publicly available CHO genomes (Lalonde and Durocher, 2017; Stolfa *et al.*, 2018). Furthermore, considering that the inherent genomic plasticity of CHO cells leads to high genetic heterogeneity between parental and clonal cell lines, it can become uninformative, or even misleading, to apply to them omics datasets obtained from ancestral cell lines, such as CHO-K1 (Datta *et al.*, 2013). With the continuing advance of technology, shown for instance in Figure 1.6 by the cost of sequencing DNA per megabase falling by several order of magnitude in less than 20 years, it may be only a matter of time before this roadblock is cleared, as omics datasets become more readily available, and affordably gathered.

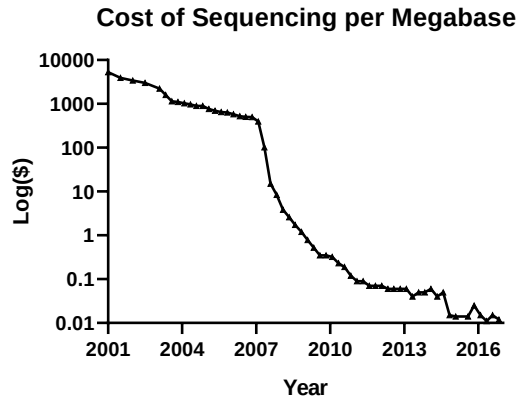


Figure 1.6 - Demonstrating the increasing availability and ease of gathering Omics datasets by the showing the decreasing cost per megabase of DNA sequencing from 2001-2017. Data gathered from www.genome.gov

1.3. Pitch: A Molecular Toolkit of Synthetic Expression Elements

As has been shown by the varied examples within this review, bottlenecks in biopharmaceutical expression can occur at almost any of the distinct molecular junctures through which genetic information must pass before it can be harvested as a secreted glycoprotein. Different studies have observed limitations in productivity stemming from inefficient transcription (Brown *et al.*, 2014a), inadequate processing and translation of mRNA (Mason *et al.*, 2012; Rajendra *et al.*, 2015a), polypeptide aggregation and ER stress (Reinhart *et al.*, 2014), inability to traffic protein to the Golgi apparatus (Mathias *et al.*, 2018), incapability of secreting protein from the Golgi apparatus (Kaneyoshi *et al.*, 2019), and limitation in cellular capacity (Dreesen and Fussenegger, 2011).

Production of glycoproteins is product, cell, and process-specific expression characteristics, and each example requires a specifically targeted solution. Synthetically maximised transcription could not alleviate a bottleneck stemming from protein aggregation, and conversely, enhanced secretion machinery could not effectively stimulate the titre of a product suffering from improper mRNA processing. In short, no generalised, one-size-fits-all solutions or platforms can be applied to producing therapeutic glycoproteins as a whole.

Rather, to shorten CLD timelines and costs, and to ensure that more therapeutic glycoproteins can be brought forward through production with as few inefficiencies as possible, biopharmaceutical companies should have access to a broad toolbox of synthetic molecular biology elements. These elements should provide control over every possible stage in glycoprotein expression, containing for example promoters for transcription, signal peptides for secretion, and effector genes for cell performance. Ideally these elements should be predictable, and titratable in their strength.

Paired with a rapid and high-throughput transient expression system, this toolbox would allow for a more efficient and streamlined CLD screening process, whereby expression characteristics are quickly identified, rationally designed against, and screened again, as shown in Figure 1.7. Expression platforms optimised via this system could then be taken

forward to stable cell line generation, carrying with them a far greater assurance that the labour-intensive process will yield clonal cell lines with favourable expression profiles.

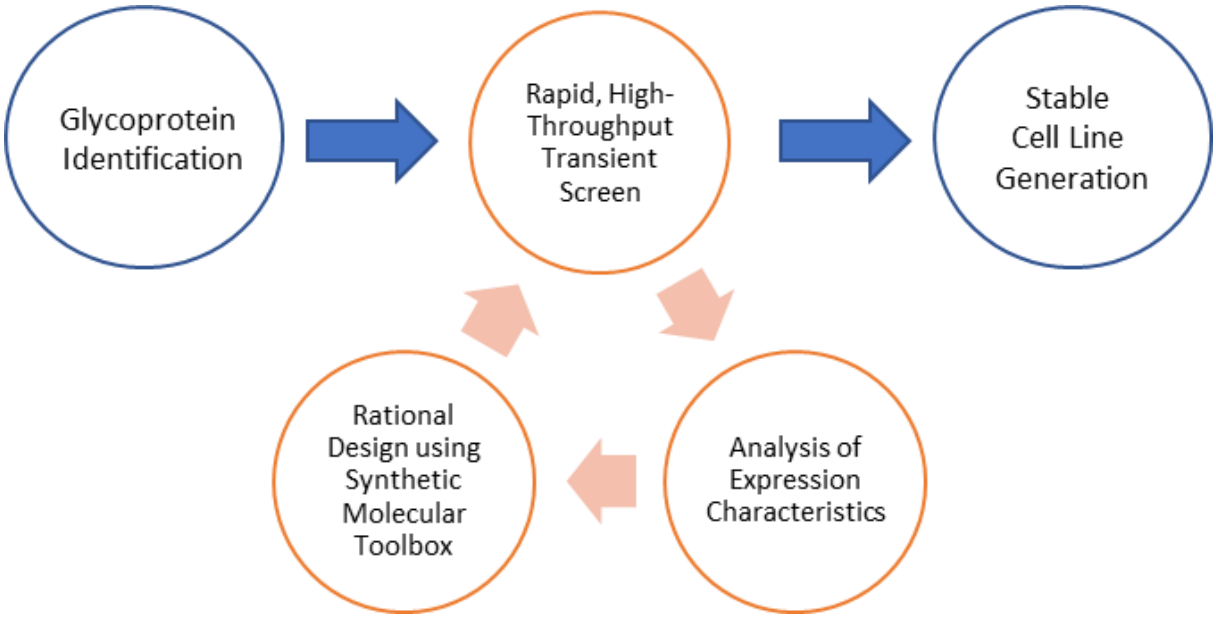


Figure 1.7 – Showing the generalised schematic of a CLD platform designed to reduce risk, and ensure a high rate of success in creating stable glycoprotein-expressing cell lines, by utilising rational design via a toolbox of synthetic molecular biology elements.

Work to create this molecular toolkit is well underway, with its most comprehensive realisation so far published by Brown *et al.*, 2019. In this study, synthetic variants of promoters, signal peptides, codon optimised transcripts, cell lines, and effector genes, were all screened for their effect on SEAP expression in a high-throughput transient system, compared to an 'industry standard' construct. It shows the many strengths that this approach currently brings: at least one variant of each synthetic element led to a significant increase in SEAP expression, when elements were combined together increases in titre were always predictable and often titratable, and combining the most favourable of every element created a construct with ~10-fold higher SEAP expression than the industry standard control (Brown *et al.*, 2019).

However, this study also demonstrates the current limitations of this platform. The study was exclusively performed using the well-characterised and easy to express SEAP reporter, and the 'design' of elements such as the variant signal peptides was based primarily in prior understanding and study of SEAP, which would not be available when optimising the expression of a new, novel glycoprotein. Moreover, multiple expression elements, such as the 5' and 3' UTRs had no synthetic variants. If SEAP expression suffered from inefficient

translation initiation, mRNA stability, or nuclear export, as these elements control, it could not have been relieved.

Rather than presenting a fully developed platform, Brown *et al.*, 2019 acts as an ideal template for how to perform next-generation CLD. As work continues into discovering new synthetic systems of expression control, and predictable understanding of those that exist grows, a landscape can be approached where this paradigm is applicable and useful for the development of every novel glycoprotein-producing cell line. In this thesis, I hope to further this paradigm, by the discovery and characterisation of synthetic expression elements for biopharmaceutical mRNA engineering.

1.4. The Molecular Journey of a Biopharmaceutical mRNA

In defiance of the historical 'central dogma' of molecular biology (Shapiro, 2009), diverse families of RNA have been discovered, exhibiting myriad functions beyond codon-based information transfer from DNA to protein: regulation of gene expression by miRNA, catalysis of splicing by snRNA, and sequence-specific guidance of endonucleases by crRNA and tracrRNA, to name a few (Doudna and Charpentier, 2014; Köhler and Hurt, 2007). Rather, this responsibility is given specifically to mRNA, which transports and translates genetic information from a storage capacity in DNA to a functional capacity in protein. As with all stages of gene expression, these processes are tightly regulated, and their synthetic control is as pertinent to optimal biopharmaceutical development and production as any other stage in expression. In the following section, the processing of a biopharmaceutical mRNA will be reviewed, tracing the flow of genetic information from the end of transcription to translation, and discussing potential junctures for the development of effective, novel, synthetic expression tools. The development of tools to address these proposed junctures will then form the experimental basis of this thesis.

1.4.1. Transfection

Whilst the transfection of recombinant DNA is far removed from the mRNA processing pathways the rest of this literature review will focus on, elements essential to those processes impact recombinant protein expression through their effect on transfection.

Gene expression cassette copy number is a fundamental factor in the efficient expression of recombinant protein. The more sites that are available for transcription of recombinant DNA, the more recombinant mRNA will be transcribed, leading to improved expression. As such, recombinant protein expression levels correlate positively with recombinant DNA copy number (Rajendra *et al.*, 2015c). This positive correlation is slightly complicated by factors such as the increase in transfection efficiency associated with addition of any DNA, including noncoding 'filler' DNA (Rajendra *et al.*, 2015a). Furthermore, like any other enhancing factor, increases in copy number may still be rendered ineffective by expression bottlenecks further downstream in the expression pathway, such as in translation or protein folding (Rajendra *et al.*, 2015b). Nonetheless, a synthetic system must optimise every element of its' expression construct, and ensuring a high copy number is an important part of this.

Despite copy number being the metric best suited to align recombinant DNA quantity with expression, cellular ability to uptake DNA across all transfection strategies (e.g. electroporation, lipofection, PEI) is fundamentally limited and defined by its total weight (Hacker *et al.*, 2013; Steger *et al.*, 2015; Wang *et al.*, 2018). Beyond a certain threshold of recombinant DNA mass per cell, for instance 1mg of pDNA per litre of culture at $1*10^6$ cells/ml (Daramola et al., 2014), no more DNA can be taken up by cells without incurring a significant cytotoxic penalty and resultant loss of titre.

As shown in Figure 1.8, DNA copy number per constant weight of DNA is negatively correlated with the size of the DNA cassette. As copy number is a fundamental factor in recombinant protein expression, it is desirable for glycoprotein expression cassettes to be as small as possible. Thus, the creation of minimised cassettes by the removal or modification of sizable mRNA elements, such as introns, without losing their expression benefits, may help to maximise copy number per transfection.

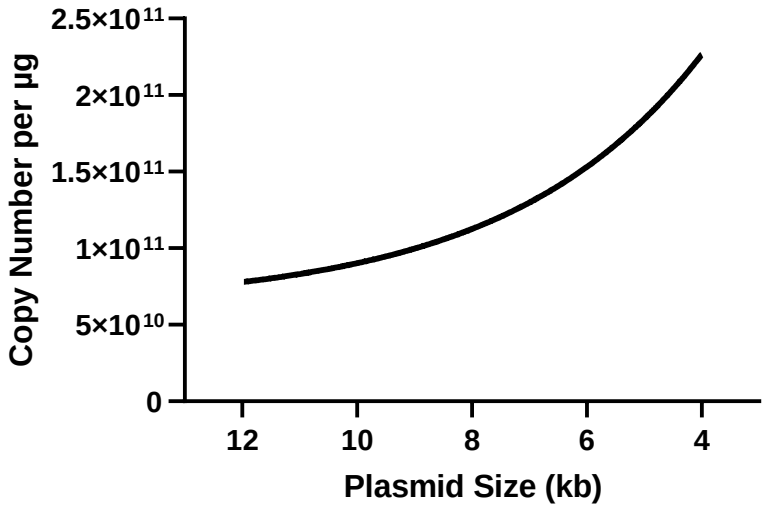


Figure 1.8 – Showing the change in recombinant DNA copy number per µg over a representative range of plasmid sizes.

1.4.2. Processing and Export

Before a nascently transcribed pre-mRNA can be translated into protein, it must first be processed, modified, and exported from the nucleus to the cytoplasm. There are multiple regulated steps in this process, each critical to efficient export of translation-competent mRNA (Heath *et al.*, 2016; Rodríguez-Navarro and Hurt, 2011).

1.4.2.1. 5′ Capping

The first stages of mRNA maturation occur co-transcriptionally (Hsin and Manley, 2012). One of these stages is 5′ capping, which involves the reversible addition of a 7-methylguanosine cap by 5-5 linkage to the 5′ end of the pre-mRNA (Bentley, 2014). This reaction occurs when the nascent transcript emerging from the RNA Polymerase II is only 20-30bp long (Rasmussen and Lis, 1993). Rather than being a ubiquitous, universal reaction, complicating evidence of such quality control mechanisms (Fernandez-Sanchez *et al.,* 2009) and cytoplasmic capping (Otsuka *et al.,* 2009) suggest that capping is regulated in a nuanced fashion (Bentley, 2014). The methyltransferase enzyme responsible for this reaction is recruited to nascent RNA via the CTD domain of RNA Polymerase II (Hsin and Manley, 2012), which is a common mechanism in transcriptional coupling. After this step, the Cap Binding Complex (CBC) binds the 5′ end of the mRNA (Heath *et al.,* 2016).

1.4.2.2. 3′ End Processing

Concurrently in pre-mRNA maturation is 3′ end cleavage and polyadenylation, giving mature mRNAs (and other long noncoding RNAs) their characteristic PolyA tail. These two processes, whereby transcribed RNA beyond the polyadenylation signal is cleaved and degraded, and a PolyA tail is added to the mRNA, are coupled, both to one another, and to transcription (Bentley, 2014; Proudfoot, 2011). The reaction is carried out by a multisubunit complex, comprising cleavage stimulation factor (*CstF*), cleavage and polyadenylation specificity factor (*CPSF*) bearing the necessary endonuclease, and cleavage factors I and II (*CFIm* and *CFIIm*) (Bentley, 2014). In order to prevent premature cleavage, coding sequence is protected by the *U1snRNP*, which inhibits cleavage and polyadenylation (Kaida *et al.,* 2010).

Since the 3′ UTR is used as a binding platform for a number of regulatory factors (Mazumder *et al.,* 2003), alternative sites of cleavage and polyadenylation, resulting in different lengths of the 3′ UTR, can have an impact on gene expression and regulation (Elkon *et al.,* 2013). For instance, proliferating cancer cells have been shown to increase cleavage at upstream polyadenylation sites, leading to reduced regulation by miRNA binding to the 3′ UTR (Sandberg *et al.,* 2008a). In another example, the long 3′ UTR isoform of the *CD47* mRNA has been shown to increase membrane localisation of *CD47*, by recruiting the 3′ UTR binding protein *HuR* (Berkovits and Mayr, 2015).

1.4.2.3. Splicing

Splicing is the process by which noncoding introns are excised from pre-mRNA, leaving ligated exons, which together form the mature mRNA, to be translated into protein. This is carried out by a large multimeric complex, known as the spliceosome (Sperling *et al.,* 2008). The core components of this complex are U1, U2, and U4-6 small nuclear RNAs (snRNAs). These RNPs require a complex biogenesis, with the RNA components of each undergoing 5′ capping and 3′ end processing, before U1, 2, 4 and 5 snRNAs are exported to the cytoplasm for processing, and re-imported to the nucleus (Köhler and Hurt, 2007). These snRNA components are essential, as through their rearrangements, they catalyse the splicing reaction itself (Madhani and Guthrie, 1994). The process of splicing begins with assembly of the spliceosome, as U1 snRNP binds the 5′ splice site (5′ss) via base-pairing reactions, and interaction with the Pol II CTD (Matera and Wang, 2014; Spiluttini *et al.,* 2010). The 3′ splice site (3′ss) is then bound by U2 snRNP, along with associated factors. A conformational change then occurs, bringing U2 snRNP into contact with the branchpoint adenosine. After this exon definition complex has been made, a pre-formed tri-snRNP consisting of U4-6 binds the complex. Conformational changes, catalysed by several Prp ATPase helicase proteins, and involving the loss of U1 and U4 snRNPs, then facilitate the two splicing reactions, shown in Figure 1.9. Whilst the reactions are directly carried out by U2 and U6 snRNPs, U5 is also essential to this process (Matera and Wang, 2014).

Splicing components are highly enriched in nuclear organelles, known as nuclear speckles (Lamond and Spector, 2003). It is thought that most splicing activity occurs on the borders between these compartments and the surrounding nuclear space (Girard *et al.,* 2012). Splicing has been shown to be substantially co-transcriptional, occurring as the nascent mRNA emerges from RNAP II (Bentley, 2014; Beyer and Osheim, 1988).

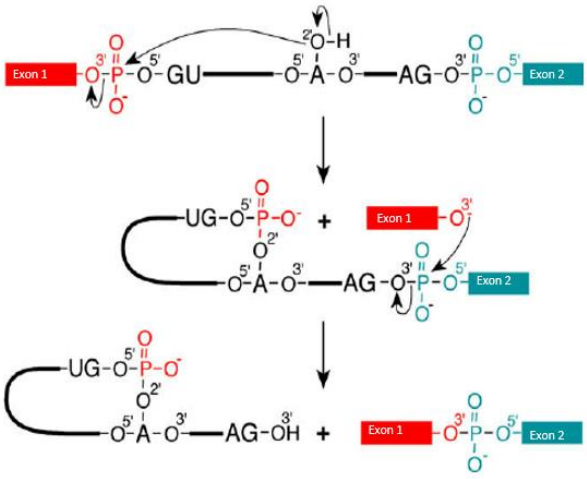


Figure 1.9 - Showing the mechanism of intron splicing, whereby an intron is cleaved out of an mRNA, leaving an exon-exon junction and an excised intron lariat. This reaction is catalysed by dynamic interactions between the pre-mRNA with U2snRNA and U6s

1.4.2.4. EJC

The Exon Junction Complex (EJC), consisting of the four proteins *Y14*, *Magoh*, *EI4A3* and *BTZ*, is a key protein signature of splicing, and has multiple roles in regulation of gene expression (Boehm and Gehring, 2016). *EIF4A3* is the main RNA-binding protein of the complex, binding via the phosphate backbone, ensuring a lack of sequence specificity (Bono *et al.*, 2006). However, it is only recruited to pre-mRNA in a splicing-dependent manner, via a transient association with the spliceosome, and therefore mostly appears in canonical sites 20-24 bases upstream of exon junctions. The EJC can also be deposited at non-canonical sites, in a mechanism that is not fully understood. It is speculated that this is related to the EJCs enrichment at purine-rich sequences of RNA, particularly a GAAGAA motif (Saulière *et al.*, 2012).

One of the strongest effects the EJC has on gene expression is through its' activation of nonsense mediated decay (NMD) of mRNA. The general model for this effect is that NMD is triggered when the EJC is detected downstream of translation termination, which could indicate incorrect splicing, or an early translation stop site (Boehm and Gehring, 2016; Nagy and Maquat, 1998). The EJC can also increase gene expression, by facilitating binding of RNA export factors, such as *ALYREF* and *DDX39B* to a messenger ribonucleoprotein (mRNP) (Gromadzka *et al.*, 2016; Heath *et al.*, 2016). This recruitment of mRNA export factors contributes towards the improved expression associated with intron-containing genes (Lu and Cullen, 2003). These are just two of the myriad of ways in which the EJC can affect gene

expression, including splicing regulation, and enhancement of translation initiation (Boehm and Gehring, 2016).

1.4.2.5. TREX

The bulk of mammalian mRNA export from the nucleus requires the transcription export (TREX) complex and the heterodimeric nuclear receptor *NXF1-NXT1* (Viphakone *et al.,* 2012). At each point during the maturation pathway key components of the TREX, such as THO1,2,5-7 (Masuda *et al.,* 2005) are recruited. Alternative forms of TREX consist of various interchangeable proteins in addition to this stoichiometric core, such as *ALYREF*, *DDX39B* or *CHTOP* (Chang *et al.,* 2013; Chi *et al.,* 2013). Unlike in yeast, where TREX recruitment is primarily coupled to transcription, metazoan TREX recruitment is primarily coupled to splicing (Masuda *et al.,* 2005), though some inefficient EJC and export factor deposition has been observed in intronless transcripts, possibly deposited via co-transcriptional scanning for introns (Viphakone *et al.,* 2019). Metazoan TREX recruitment is also coupled to mRNA processing through its' interactions with the RNAPII CTD, the CBC, the EJC, and *Prp19* (Gromadzka *et al.,* 2016; Heath *et al.,* 2016; Hsin and Manley, 2012). In particular, the critical export factors *ALYREF* and *CHTOP* are deposited preferentially at the 5′ and 3′ ends of pre-mRNA respectively, the former by utilising the CBC as a transient landing pad before associating more stably with EJCs and influencing splicing events, and the latter through preferential association with Ser2-phosphorylated RNA Polymerase 2 (which occurs more frequently toward the 3′ end of transcription), from where it can influence 3′ end processing (Viphakone *et al.,* 2019).

After early recruitment to these protein signals, TREX subunit deposition on the mRNA is mostly mediated through the *DDX39A/B* RNA helicases, which load export factors such as *ALYREF*, *CHTOP*, and *UIF* using it's helicase ATPase cycle (Dufu *et al.,* 2010). Unlike the THO core, all of these proteins strongly bind RNA. This, in combination with the substantially greater concentration of these proteins than THO components (Schwanhausser *et al.,* 2011) have led to speculation that THO's major role is as a chaperone for export factors binding mRNA, and it is unclear whether these export factors remain bound to THO after their deposition (Heath *et al.,* 2016).

Once export factors have bound to the mRNA, its' primary export ability comes from recruitment of the *NXF1* protein to the mRNA, which then facilitates nuclear export through the nuclear pore complex (NPC) (Okamura *et al.,* 2014). Through an NTF2-like (NT2FL) domain, *NXF1* binds the *NXT1* protein, forming a heterodimer required for its' stability and function (Liker *et al.,* 2000; Okamura *et al.,* 2014). Through various protein interactions, chiefly mediated by *ALYREF*, autoinhibition of *NXF1*'s RNA-binding capacity is relieved, facilitating mRNA binding and nuclear export (Viphakone *et al.,* 2012).

1.4.2.6. Nuclear Export

Once *NXF1-NXT1* has bound the mRNA, a conformational change in the mRNP occurs, known as remodelling, which is orchestrated by *DDX39B*, and it is taken to the nuclear membrane for export through the NPC (Okamura *et al.,* 2014). Through interaction with FG repeats, *NXF1* helps the mRNP to overcome the permeability barrier of the NPC, facilitating export into the cytoplasm (Köhler and Hurt, 2007).

Inefficient processing and export of recombinant transcripts can be a limiting factor in biopharmaceutical expression, with higher quantities of pDNA delivered into the nucleus not leading to higher quantities of exported mRNA (Rajendra *et al.,* 2015a). This is especially relevant in regard to solving the problem of excess vector size discussed in section 1.4.1. As much of the mRNA processing and export machinery is recruited by splicing (Heath *et al.,* 2016; Masuda *et al.,* 2005), the creation of minimised transcripts by removal of introns may impair efficient mRNA delivery into the cytoplasm (Lu and Cullen, 2003). Tools targeted toward this problem will be investigated in chapter 3 of this thesis.

1.4.3. Translation

Translation, in which mRNA codons are decoded by interactions with specific aminoacyl-tRNAs, facilitated by the polypeptide factory of the ribosome, is the final stage of gene expression with which a biopharmaceutical mRNA is involved. It is heavily energy consuming, and its' four stages are thus all tightly regulated (Jackson *et al.,* 2010; Valvezan and Manning, 2019).

1.4.3.1. Initiation

The first stage of translation is initiation, starting with a processed and exported mRNP in the cytoplasm, and ending when that mRNP has been bound by a full 80S ribosome, the start codon positioned at the peptidyl decoding site, and the methionyl-initiator transfer RNA (Met-tRNAi) bound to it, ready for polypeptide elongation to begin (Aylett and Ban, 2017). Initiation is the most strictly regulated of all four steps in translation, and is thus often a bottleneck in protein production (Aitken and Lorsch, 2012; Jackson *et al.,* 2010).

To begin translation initiation, an exported mRNP is bound by eukaryotic initiation factor 4F (*eIF4F*), a trimer consisting of *eIF4E*, *eIF4A1*, and *eIF4G*, and by *eIF4B* at its 5' terminus. The mRNP is also bound by PolyA-binding protein (*PABP*), at its PolyA tail. *PABP* binds to *eIF4G*, forming a circularised mRNP. Assisted by binding from *eIF4G* and *eIF4B*, *eIF4A1* then begins to unwind any 5'UTR mRNA secondary structures, via its ATP-dependent helicase activity, rendering an activated mRNP, ready for ribosome binding (Aitken and Lorsch, 2012; Jackson *et al.,* 2010). Evidence gathered from assays in the presence and absence of these various factors indicate that not every factor is strictly essential for this mRNA activation, rather, they form a network of redundant interactions, perturbations of which are used to regulate dynamics of initiation (Mayberry *et al.,* 2009), for instance in the observation that abrogation of the *PABP-eIF4G* interaction only significantly downregulates translation initiation under conditions of high competition for translational apparatus (Svitkin *et al.,* 2009).

This activating mRNP is then bound by the 43S pre-initiation complex (PIC), consisting of the 40S ribosome, *eIF1A*, *eIF1*, *eIF5*, *eIF3*, and Met-rRNAi-*eIF2*-GTP. Using the processive helicase activity of *eIF4A1*, the mRNA is then scanned for the AUG start codon, to which Met-tRNAi can bind. To enable scanning along the mRNA, *eIF1* and *eIF1a* bind to inhibit ribosomal RNA (rRNA)-mRNA interactions, which might otherwise inhibit the processivity of scanning. When the start codon, most efficiently within the ideal context of a Kozak sequence (GCC(A/G)CCAUGG), passes thought the P site, it is arrested by binding to Met-tRNAi, causing a conformational change that expels *eIF1*, and hydrolyses the GTP bound to *eIF2*, before it is expelled from the complex (Aylett and Ban, 2017), forming the 'closed' 48S PIC. Finally, aided by GTP-bound *eIF5B*, the 60S ribosome binds, and upon hydrolysis of GTP-

eIF5B, all initiation factors are ejected, and the complete 80S initiation complex, ready for elongation, is formed (Jackson *et al.,* 2010).

Translation initiation can be regulated by a number of mechanisms. Firstly, global regulation can occur by the phosphorylation or otherwise manipulation of translation initiation factors. For example, *eIF2a* is phosphorylated by PERK to downregulate translation in response to triggering of the UPR (Jackson *et al.,* 2010), and *mTOR* dynamically regulates the phosphorylation of *4E-BP1/2/3* in response to nutrients and growth signals, which when hyperphosphorylated, sequester *eIF4E*, inhibiting translation (Roux and Topisirovic, 2018; Valvezan and Manning, 2019). Initiation can also be regulated by sequence-specific protein-RNA binding events. Examples of this include the regulation of adenine/uridine-rich elements (ARE) in the 3'UTR by specific binding of *HuR* (Mazan-Mamczarz *et al.,* 2003; Wu *et al.,* 2015), or the nutrient/growth factor dependant regulation of ribosomal proteins containing a 5'TOP sequence, specifically bound by *LARP1*, rendering transcripts hypersensitive to regulation by *mTOR* (Fonseca *et al.,* 2018; Hong *et al.,* 2017).

Initiation is the most tightly regulated stage of translation, acting as the primary bottleneck of its rate, outside of cases with notably suboptimal codons (Aitken and Lorsch, 2012; Jackson *et al.,* 2010). Abundance of protein in mammalian cells is fundamentally controlled by translation (Schwanhausser *et al.,* 2011), and expression deficits can occur at translation, failing to efficiently convert abundant mRNA into recombinant protein (Rajendra *et al.,* 2015a). Despite this, beyond the ubiquitous Kozak sequence, there are no synthetic, predictable, or titratable elements for controlling the efficiency of translation initiation. Potential tools targeting this step in expression will be explored in Chapter 5 of this thesis.

1.4.3.2. Elongation

Translation elongation involves the processive ratcheting of mRNA through three sites in 80S ribosome (Aminoacyl site, Peptidyl site, Exit site), matching each codon with its cognate aminoacyl-tRNA to form peptide bonds between each residue of the open reading frame (ORF), and ends when a stop codon is reached. It is a cycle, composed of three steps: tRNA selection, peptide-bond formation, and translocation of the mRNA-tRNA complex. During tRNA selection, aminoacyl-tRNAs are delivered to the A site by *eEF1a*, in complex in GTP. Upon binding of a cognate codon and tRNA, *eEF1a* utilises its GTPase activity to hydrolyse

its bound GTP and fully accommodates the tRNA at the A site. As the peptide bond is formed by nucleophilic attack, a conformational change occurs, such that the two bound tRNAs occupy both the P/E and A/P sites respectively: the so-called hybrid-state. This hybrid state is the substrate for *eEF2*, which hydrolyses GTP to translocate the tRNA-mRNA complex through the ribosome, to the E and P sites (Schuller and Green, 2018).

Several mechanisms are used to regulate elongation. For example, the peptide bond formation kinetics are less favourable for some residue pairs than others, for instance, a proline-proline pair can cause translation to stall. This stalling can be rescued by *eIF5a*, originally thought to be an initiation factor, now with growing calls to be renamed as an elongation factor, which stimulates peptide bond formation (Browning and Bailey-Serres, 2015). Ribosome stalling can also occur due to poor rates of A site occupancy, caused by low levels of aminoacyl-tRNAs cognate to a specific codon. This stalling is most often resolved by either a frameshift in codon reading, or misincorporation of a near-cognate tRNA (Schuller and Green, 2018). These resolutions are obviously problematic for the expression of a supposedly homogenous biopharmaceutical protein, meaning that codon optimality is important both for increasing product expression, and product quality (Hanson and Coller, 2018; Lalonde and Durocher, 2017).

1.4.3.3. Termination and Recycling

Translation termination begins when the stop codon enters the A site of the 80S ribosome. This stop codon is recognised by *eRF1*, a protein which resembles an amino-acyl tRNA in structure, binding through a highly conserved NIKS domain, with codon-anticodon like interactions. In this position, eRF1 extends into the site previously occupied by tRNA acceptor stems, promoting peptide hydrolysis of the nascent polypeptide. *eRF1* also recruits *eRF3* to the ribosome, which facilitates peptide release through its GTPase activity. Termination is typically regulated by the detection that the stop codon is the appropriate distance from the appropriate 3'UTR binding proteins, particularly *PABP*. Inappropriate distances between proteins, as well as detection of proteins positioned incorrectly, most canonically an EJC downstream of the stop codon, contribute toward inefficient termination, and even NMD of the transcript (Schuller and Green, 2018).

After this termination, the mRNA and *eRF1* remain bound to the 80S ribosome. This complex is separated and recycled by the *ABCE1* protein, which utilises energy generated by ATP hydrolysis to separate the constituent parts, recycling them for the next round of translation (Dever and Green, 2012). *ABCE1* and even the mRNA can sometimes remain bound to the 40S ribosome after this process, facilitating reinitiation of translation, utilising mRNA circularisation to re-bind and re-scan the transcript (Schuller and Green, 2017).

1.4.4. mRNA Stability

Highly regulated rates of turnover and mRNA half-lives are critical to the concentration of an mRNA in the cytoplasm, as stable mRNAs may be translated many times, leading to high protein expression, whereas certain transcripts may even be co-translationally degraded upon their first round of translation (Bicknell and Ricci, 2017). Therefore, mRNA stability is a critical determinant of expression characteristics (Fukao and Fujiwara, 2017; Presnyak *et al.*, 2015).

1.4.4.1. Global Decay Pathways

Degradation of mRNA is generally performed by either 5′-3′ decay (which in some cases can occur co-translationally (Bicknell and Ricci, 2017)), or by exosome-mediated 3′-5′ decay.

5′-3′ decay begins with deadenylation of the PolyA tail, mediated through the *PAN2*/*PAN3* and *CCR4*/*NOT* complexes, partially dissociating *PABP* from the transcript, thereby inhibiting translation. The decapping complex of *DCP1*/*DCP2* is recruited by this process, via interactions with decapping coactivators *DDX6* and *Edc3*. Upon binding, a conformational change in the catalytic *DCP2* and binding of cofactors such as *Edc1* allow hydrolysis of the 5′ cap, thus rendering the 5′ end of the transcript vulnerable to rapid, processive exonucleolytic degradation by *Xrn1* (Mugridge *et al.*, 2018). This process is inhibited by both translation initiation and elongation, with *PABP* association and codon optimality thought to be the main factors linking the two processes (Chan *et al.*, 2018; Hanson and Coller, 2018; Presnyak *et al.*, 2015). In the case of 3′-5′ degradation, deadenylated transcripts are bound by the 10^{44} exosome, consisting of a 9-subunit core, which binds RNA and regulates the exoribonuclease activity of *Rrp44*, which processivity degrades the mRNA in a 3′-5′ direction (Januszyk and Lima, 2014).

1.4.4.2. mRNA Surveillance

To prevent the production of aberrant protein, mRNA in the cytoplasm is subject to three main quality control mechanisms: nonsense-mediated, nonstop, and no-go decay (NMD, NSD, NGD), which are known collectively as mRNA surveillance (Bicknell and Ricci, 2017).

NMD detects transcripts with premature stop codons (PTCs), canonically identified by detection of an EJC downstream of a stop codon. The efficient NMD pathways of organisms containing largely intronless genes have since called the simplicity of this model into question, and a new model has emerged, in which proximity of the stop codon to *PABP* and abundance of 3'UTR-binding proteins such as Upf1 are also integrated into the detection signal (Shoemaker and Green, 2012), reinforced by evidence that 3'UTRs can influence mRNA stability by forming secondary structures to change physical stop codon proximity to the PolyA sequence (Wu and Bartel, 2017).

NSD and NGD detect transcripts lacking a stop codon, and transcripts on which the ribosome permanently stalls whilst inside the CDS, respectively. In all cases the ribosome will stall: on the end of a truncated transcript, at a site of stable mRNA secondary structure, at a site of inefficient peptide-linkage, or at a PolyA tail, all of which are detected by *Dom34*/*Hsb1*. The shared mechanism of ribosome stalling has somewhat blurred the sharp line between NSD and NGD, for instance in cases of stalling on a PolyA tail, which may be due as much to formation of energetically unfavourable Poly-Lys peptides (causing NGD) as to reaching the end of the transcript (causing NSD) (Radhakrishnan and Green, 2016; Shoemaker and Green, 2012).

Upon detection by any of the mRNA surveillance pathways, transcripts are split by an endonuclease upstream of the detection site, and the 5' and 3' fragments degraded by the exosome and *Xrn1*, respectively (Bicknell and Ricci, 2017).

1.4.4.3. mRNA Stability Regulation

The stability of cytoplasmic mRNA is regulated by a number of factors, such as 5' cap methylation (Mauer *et al.*, 2017), binding of AREs by RNA-binding proteins (Łabno *et al.*, 2016; Schoenberg and Maquat, 2012), the potential use of which will be discussed in chapter 3, and targeted regulation by miRNAs (Hammond, 2015), which has already been discussed.

However, the most relevant mechanism of stability regulation for a biopharmaceutical transcript is the link between mRNA half-life and codon optimality.

A striking genome-wide correlation has been observed between codon optimality and mRNA stability, with transcripts containing >70% optimal codons having around four-fold longer half-lives than those containing <40% in budding yeast. This mechanistic link was strengthened by the observation that substitution of WT codons with optimised and unoptimised codons had significant effects on mRNA half-life, and an implication of codon optimality as a secondary genetic code made by observation of the optimal codon enrichment in functional groups such as ribosomal protein mRNAs, and depletion in other groups, such as tRNA modification genes (Presnyak *et al.*, 2015). This correlation has been replicated both in higher eukaryotic zebrafish (Mishima and Tomari, 2016), and in mammalian NIH3T3 cells (Radhakrishnan and Green, 2016). This link is thought to be mediated through slower ribosomal processivity on suboptimal transcripts, sensed by the DEAD-box protein *DDX6* (functionally associated with mRNA decapping), which preferentially associates with slow-moving ribosomes to induce enhanced degradation (Hanson and Coller, 2018; Heck and Wilusz, 2018).

The efficiency of recombinant protein expression is tied to the availability of translation-competent mRNA in the cytoplasm, and high rates of mRNA turnover can compromise this, with studies showing deficiencies in production stemming from low concentrations of mRNA, not caused by transcription deficiencies (Mason *et al.*, 2012). In cases that led to high mRNA turnover that cannot be resolved by codon optimisation, such as undesirable peptide linkages, it is desirable to control mRNA stability through other means. Potential solutions targeted towards this stage in expression will be explored in chapter 4 of this thesis.

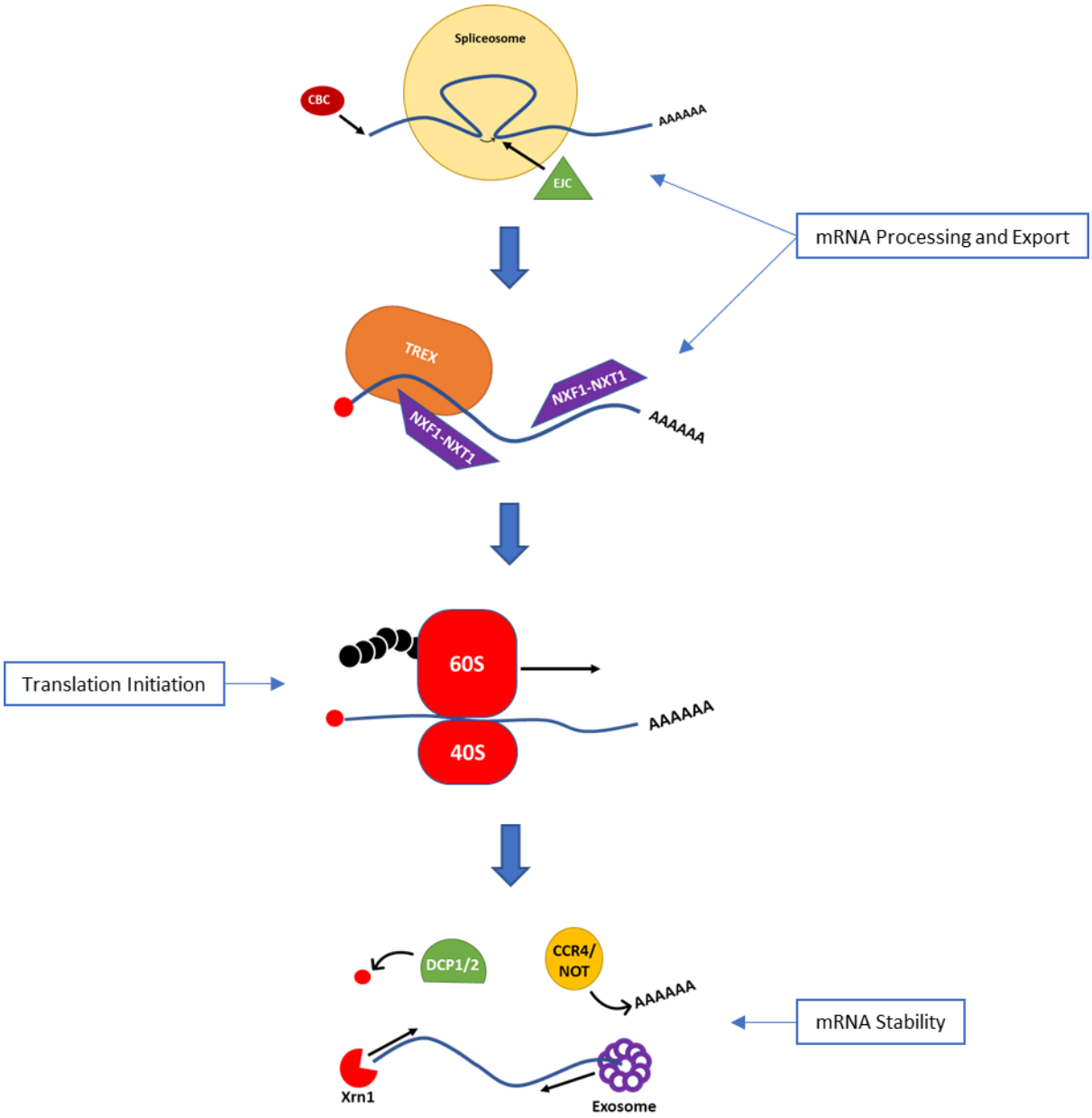


Figure 1.10 – Showing the four main stages of, and most key factors in, the journey of a biopharmaceutical mRNA. First is mRNA maturation, including the addition of a 5′ cap, splicing of introns, and 3′ end maturation, followed by loading with factors and exportation from the nucleus (it should be noted that there is no clear chronological divide between these first two stages, and RNA export factors may even influence RNA maturation (Viphakone *et al.*, 2019)*). Next is translation in the cytoplasm, followed finally by degradation. The four molecular junctures for the development of molecular biology tools to be explored in this thesis are also marked at their time of action.*

1.5. Conclusions

In response to changing ethical and economic needs, the CLD process for production of therapeutic glycoproteins has undergone a myriad of synthetic developments in molecular biology, from various vector engineering strategies, to cell line manipulation, to transfection and culture strategy. In order to meet the current demands, a streamlined development platform, with reduced risk and increased screening capabilities must be developed, whereby rapidly transiently screened solutions can be rationally engineered toward optimality. In order to create this system, a molecular toolkit of synthetic biology parts, covering every step of product expression in both the vector and cell line must be developed. The work presented in this thesis will contribute towards this unmet need, by investigating synthetic tools for controlling biopharmaceutical titre by manipulation of mRNA, forming opportunities for synthetic control.

1.6. References

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. 19, 568–576.

Alves, C.S., and Dobrowsky, T.M. (2017). Strategies and Considerations for Improving Expression of "Difficult to Express" Proteins in CHO Cells. In Methods in Molecular Biology, P. Meleady, ed. (New York, NY: Springer New York), pp. 1–23.

Aylett, C.H.S., and Ban, N. (2017). Eukaryotic aspects of translation initiation brought into focus. Philos. Trans. R. Soc. B Biol. Sci. 372.

Bachu, R., Bergareche, I., and Chasin, L.A. (2015). CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining. Biotechnol. Bioeng. 112, 2154–2162.

Barash, S., Wang, W., and Shi, Y. (2002). Human secretory signal peptide description by hidden Markov model and generation of a strong artificial signal peptide for secreted protein expression. Biochem. Biophys. Res. Commun. 294, 835–842.

Barnard, G.C., Hougland, M.D., and Rajendra, Y. (2015). High-throughput mAb expression and purification platform based on transient CHO. Biotechnol. Prog. 31, 239–247.

Baser, B., Spehr, J., Büssow, K., and van den Heuvel, J. (2016). A method for specifically targeting two independent genomic integration sites for co-expression of genes in CHO cells. Methods 95, 3–12.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. 15, 163–175.

Berkovits, B.D., and Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature 522, 363–367.

Betts, Z., and Dickson, A.J. (2015). Assessment of UCOE on Recombinant EPO Production and Expression Stability in Amplified Chinese Hamster Ovary Cells. Mol. Biotechnol. 57, 846–858.

Betts, Z., and Dickson, A.J. (2016). Ubiquitous Chromatin Opening Elements (UCOEs) effect on transgene position and expression stability in CHO cells following methotrexate (MTX) amplification. Biotechnol. J. 11, 554–564.

Betts, Z., Croxford, A.S., and Dickson, A.J. (2015). Evaluating the interaction between UCOE and DHFR-linked amplification and stability of recombinant protein expression. Biotechnol. Prog. 31, 1014–1025.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. 2, 754–765.

Bicknell, A.A., and Ricci, E.P. (2017). When mRNA translation meets decay. Biochem. Soc. Trans. 45, 339–351.

Boehm, V., and Gehring, N.H. (2016). Exon Junction Complexes: Supervising the Gene Expression Assembly Line. Trends Genet. xx, 1–12.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The Crystal Structure of the Exon Junction Complex Reveals How It Maintains a Stable Grip on mRNA. Cell 126, 713–725.

Brown, A.J., and James, D.C. (2015). Precision control of recombinant gene transcription for CHO cell synthetic biology. Biotechnol. Adv. 34, 492–503.

Brown, A.J., Sweeney, B., Mainwaring, D.O., and James, D.C. (2014). Synthetic promoters for CHO cell engineering. Biotechnol. Bioeng. 111, 1638–1647.

Brown, A.J., Sweeney, B., Mainwaring, D.O., and James, D.C. (2015). NF-κB, CRE and YY1 elements are key functional regulators of CMV promoter-driven transient gene expression in CHO cells. Biotechnol. J. 10, 1019–1028.

Brown, A.J., Gibson, S.J., Hatton, D., and James, D.C. (2017). In silicodesign of context-responsive mammalian promoters with user-defined functionality. Nucleic Acids Res. 45, 10906–10919.

Brown, A.J., Gibson, S.J., Hatton, D., Arnall, C.L., and James, D.C. (2019). Whole synthetic pathway engineering of recombinant protein production. Biotechnol. Bioeng. 116, 375–387.

Browning, K.S., and Bailey-Serres, J. (2015). Mechanism of Cytoplasmic mRNA Translation. Arab. B. 13, e0176.

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

Campbell, M., Corisdeo, S., McGee, C., and Kraichely, D. (2010). Utilization of site-specific recombination for generating therapeutic protein producing cell lines. Mol. Biotechnol. 45, 199–202.

Carpentier, E., Paris, S., Kamen, A.A., and Durocher, Y. (2007). Limiting factors governing protein expression following polyethylenimine-mediated gene transfer in HEK293-EBNA1 cells. J. Biotechnol. 128, 268–280.

Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. Elife 7, 1–32.

Chang, C., Hautbergue, G.M., Walsh, M.J., Viphakone, N., van Dijk, T.B., Philipsen, S., and Wilson, S.A. (2013). Chtop is a component of the dynamic TREX mRNA export complex. EMBO J. 32, 473–486.

Cheng, J.K., Lewis, A.M., Kim, D.S., Dyess, T., and Alper, H.S. (2016). Identifying and retargeting transcriptional hot spots in the human genome. Biotechnol. J. 11, 1100–1109.

Chi, B., Wang, Q., Wu, G., Tan, M., Wang, L., Shi, M., Chang, X., and Cheng, H. (2013). Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. Nucleic Acids Res. 41, 1294–1306.

Chin, C.L., Chin, H.K., Chin, C.S., Lai, E.T., and Ng, S.K. (2015). Engineering selection stringency on expression vector for the production of recombinant human alpha1-antitrypsin using Chinese Hamster ovary cells. BMC Biotechnol. 15, 1–15.

Chromikova, V., Mader, A., Steinfellner, W., and Kunert, R. (2015). Evaluating the bottlenecks of recombinant IgM production in mammalian cells. Cytotechnology 67, 343–356.

Courtes, F.C., Lin, J., Lim, H.L., Ng, S.W., Wong, N.S.C., Koh, G., Vardy, L., Yap, M.G.S., Loo, B., and Lee, D.Y. (2013). Translatome analysis of CHO cells to identify key growth genes. J. Biotechnol. 167, 215–224.

Daramola, O., Stevenson, J., Dean, G., Hatton, D., Pettman, G., Holmes, W., and Field, R. (2014). A high-yielding CHO transient system: Coexpression of genes encoding EBNA-1 and GS enhances transient protein expression. Biotechnol. Prog. 30, 132–141.

Datta, P., Linhardt, R.J., and Sharfstein, S.T. (2013). An 'omics approach towards CHO cell engineering. Biotechnol. Bioeng. 110, 1255–1271.

David W. Thomas, Justin Burns, John Audette, Adam Carroll, Corey Dow-Hygelund, and Michael Hay (2016). Clinical Development Success Rates. BioMedTracker June.

Delday, M., Mulder, I., Logan, E.T., and Grant, G. (2018). Bacteroides thetaiotaomicron Ameliorates Colon Inflammation in Preclinical Models of Crohn's Disease. Inflamm. Bowel Dis. XX, 1–12.

Dever, T.E., and Green, R. (2012). The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. Cold Spring Harb. Perspect. Biol. 4, a013706–a013706.

DiMasi, J.A., Grabowski, H.G., and Hansen, R.W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. J. Health Econ. 47, 20–33.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science (80-.). 346, 1258096–1258096.

Dreesen, I.A.J., and Fussenegger, M. (2011). Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. Biotechnol. Bioeng. 108, 853–866.

Dufu, K., Livingstone, M.J., Seebacher, J., Gygi, S.P., Wilson, S.A., and Reed, R. (2010). ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. Genes Dev. 24, 2043–2053.

Dvir, S. et al. (2013) 'Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast.', Proceedings of the National Academy of Sciences of the United States of America. National Academy of Sciences, 110(30), pp. E2792-801

Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. 14, 496–506.

EMA (2019). Glossary. Retrieved September 20, 2019, from https://www.ema.europa.eu/en/about-us/about-website/glossary.

Evaluate Pharma (2018). World Preview 2018 , Outlook to 2024. 11th Editi.

Fernandez-Sanchez, M.E., Gonatopoulos-Pournatzis, T., Preston, G., Lawlor, M.A., and Cowling, V.H. (2009). S-Adenosyl Homocysteine Hydrolase Is Required for Myc-Induced mRNA Cap Methylation, Protein Synthesis, and Cell Proliferation. Mol. Cell. Biol. 29, 6182–6191.

Fischer, S., Handrick, R., and Otte, K. (2015). The art of CHO cell engineering: A comprehensive retrospect and future perspectives. Biotechnol. Adv. 33, 1878–1896.

Fischer, S., Marquart, K.F., Pieper, L.A., Fieder, J., Gamer, M., Gorr, I., Schulz, P., and Bradl, H. (2017). miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development. Biotechnol. Bioeng. 114, 1495–1510.

Fonseca, B.D., Lahr, R.M., Damgaard, C.K., Alain, T., and Berman, A.J. (2018). LARP1 on TOP of ribosome production. Wiley Interdiscip. Rev. RNA 1–16.

Le Fourn, V., Girod, P.A., Buceta, M., Regamey, A., and Mermod, N. (2014). CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. Metab. Eng. 21, 91–102.

Fu, M.H., Li, C.L., Lin, H.L., Chen, P.C., Calkins, M.J., Chang, Y.F., Cheng, P.H., and Yang, S.H. (2015). Stem cell transplantation therapy in Parkinson's disease. Springerplus 4, 1–8.

Fukao, A., and Fujiwara, T. (2017). The coupled and uncoupled mechanisms by which trans-acting factors regulate mRNA stability and translation. J. Biochem. 161, 309–314.

Gaj, T., Gersbach, C.A., and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31, 397–405.

Gaughan, C.L. (2016). The present state of the art in expression, production and characterization of monoclonal antibodies. Mol. Divers. 20, 255–270.

Geisse, S., and Fux, C. (2009). Chapter 15 Recombinant Protein Production by Transient Gene Transfer into Mammalian Cells (Elsevier Inc.).

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nat. Commun. 3, 994.

Grav, L.M., Lee, J.S., Gerling, S., Kallehauge, T.B., Hansen, A.H., Kol, S., Lee, G.M., Pedersen, L.E., and Kildegaard, H.F. (2015). One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. Biotechnol. J. 10, 1446–1456.

Gromadzka, A.M., Steckelberg, A.L., Singh, K.K., Hofmann, K., and Gehring, N.H. (2016). A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. Nucleic Acids Res. 44, 2348–2361.

Güler-Gane, G., Kidd, S., Sridharan, S., Vaughan, T.J., Wilkinson, T.C.I., and Tigue, N.J. (2016). Overcoming the Refractory Expression of Secreted Recombinant Proteins in Mammalian Cells through Modification of the Signal Peptide and Adjacent Amino Acids. PLoS One 11, e0155340.

Gulis, G., Simi, K.C.R., de Toledo, R.R., Maranhao, A.Q., and Brigido, M.M. (2014). Optimization of heterologous protein production in Chinese hamster ovary cells under overexpression of spliced form of human X-box binding protein. BMC Biotechnol. 14, 1–12.

Gutiérrez-Granados, S., Cervera, L., Kamen, A.A., and Gòdia, F. (2018). Advancements in mammalian cell transient gene expression (TGE) technology for accelerated production of biologics. Crit. Rev. Biotechnol. 38, 918–940.

Hacker, D.L., Kiseljak, D., Rajendra, Y., Thurnheer, S., Baldi, L., and Wurm, F.M. (2013). Polyethyleneimine-based transient gene expression processes for suspension-adapted HEK-293E and CHO-DG44 cells. Protein Expr. Purif. 92, 67–76.

Hackl, M., Borth, N., and Grillari, J. (2012). MiRNAs - pathway engineering of CHO cell factories that avoids translational burdening. Trends Biotechnol. 30, 405–406.

Hammond, S.M. (2015). An overview of microRNAs. Adv. Drug Deliv. Rev. 87, 3–14.

Hansen, H.G., Pristovšek, N., Kildegaard, H.F., and Lee, G.M. (2017). Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions. Biotechnol. Adv. 35, 64–76.

Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. Nat. Rev. Mol. Cell Biol. 19, 20–30.

Heath, C.G., Viphakone, N., and Wilson, S.A. (2016). The role of TREX in gene expression and disease. Biochem. J. 473, 2911–2935.

Heck, A.M., and Wilusz, J. (2018). The interplay between the RNA decay and translation machinery in eukaryotes. Cold Spring Harb. Perspect. Biol. 10.

Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., et al. (2016). A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. Cell Syst. 3, 434-443.e8.

Hinnebusch, A.G., Ivanov, I.P., and Sonenberg, N. (2016). Translational control by 5′-untranslated regions of eukaryotic mRNAs. Science (80-.). 352, 1413–1416.

Ho, S.C.L., Bardor, M., Feng, H., Mariati, Tong, Y.W., Song, Z., Yap, M.G.S., and Yang, Y. (2012). IRES-mediated Tricistronic vectors for enhancing generation of high monoclonal antibody expressing CHO cell lines. J. Biotechnol. 157, 130–139.

Ho, S.C.L., Mariati, Yeo, J.H.M., Fang, S.G., and Yang, Y. (2014). Impact of Using Different Promoters and Matrix Attachment Regions on Recombinant Protein Expression Level and Stability in Stably Transfected CHO Cells. Mol. Biotechnol. 57, 138–144.

Hong, S., Freeberg, M.A., Han, T., Kamath, A., Yao, Y., Fukuda, T., Suzuki, T., Kim, J.K., and Inoki, K. (2017). LARP1 functions as a molecular switch for mTORC1-mediated translation of an essential class of mRNAs. Elife 6.

Hsin, J., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. 2119–2137.

Hung, F., Deng, L., Ravnikar, P., Condon, R., Li, B., Do, L., Saha, D., Tsao, Y.S., Merchant, A., Liu, Z., et al. (2010). mRNA stability and antibody production in CHO cells: Improvement through gene optimization. Biotechnol. J. 5, 393–401.

Hussain, H., Maldonado-Agurto, R., and Dickson, A.J. (2014). The endoplasmic reticulum and unfolded protein response in the control of mammalian recombinant protein production. Biotechnol. Lett. 36, 1581–1593.

Jackson, R.J., Hellen, C.U.T., and Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. 11, 113–127.

Jadhav, V., Hackl, M., Druz, A., Shridhar, S., Chung, C.Y., Heffner, K.M., Kreil, D.P., Betenbaugh, M., Shiloach, J., Barron, N., et al. (2013). CHO microRNA engineering is growing up: Recent successes and future challenges. Biotechnol. Adv. 31, 1501–1513.

Januszyk, K., and Lima, C.D. (2014). The eukaryotic RNA exosome. Curr. Opin. Struct. Biol. 24, 132–140.

Jayapal, K., Wlaschin, K., Hu, W., and Yap, G. (2007). Recombinant protein therapeutics from CHO cells-20 years and counting. Chem. Eng. Prog. 103, 40–47.

Johari, Y.B., Estes, S.D., Alves, C.S., Sinacore, M.S., and James, D.C. (2015). Integrated cell and process engineering for improved transient production of a “difficult-to-express” fusion protein by CHO cells. Biotechnol. Bioeng. 112, 2527–2542.

Johari, Y.B., Brown, A.J., Alves, C.S., Zhou, Y., Wright, C.M., Estes, S.D., Kshirsagar, R., and James, D.C. (2019). CHO genome mining for synthetic promoter design. J. Biotechnol. 294, 1–13.

Joshi, S.R., Parikh, R.M., and Das, A.K. (2007). Insulin--history, biochemistry, physiology and pharmacology. J. Assoc. Physicians India 55 Suppl, 19–25.

Jossé, L., Zhang, L., and Smales, C.M. (2018). Application of microRNA Targeted 3′UTRs to Repress DHFR Selection Marker Expression for Development of Recombinant Antibody Expressing CHO Cell Pools. Biotechnol. J. 13, 1–8.

Jostock, T., and Knopf, H.-P. (2012). Mammalian Stable Expression of Biotherapeutics. In Life Sciences, pp. 227–238.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468, 664–668.

Kameyama, Y., Kawabe, Y., Ito, A., and Kamihira, M. (2010). An accumulative site-specific gene integration system using Cre recombinase-mediated cassette exchange. Biotechnol. Bioeng. 105, 1106–1114.

Kaneyoshi, K., Kuroda, K., Uchiyama, K., Onitsuka, M., Yamano-Adachi, N., Koga, Y., and Omasa, T. (2019). Secretion analysis of intracellular “difficult-to-express” immunoglobulin G (IgG) in Chinese hamster ovary (CHO) cells. Cytotechnology 71, 305–316.

Kang, S., Ren, D., Xiao, G., Daris, K., Buck, L., Enyenihi, A.A., Zubarev, R., Bondarenko, P. V., and Deshpande, R. (2014). Cell line profiling to improve monoclonal antibody production. Biotechnol. Bioeng. 111, 748–760.

Kawabe, Y., Inao, T., Komatsu, S., Huang, G., Ito, A., Omasa, T., and Kamihira, M. (2017). Improved recombinant antibody production by CHO cells using a production enhancer DNA element with repeated transgene integration at a predetermined chromosomal site. J. Biosci. Bioeng. 123, 390–397.

Kesik-Brodacka, M. (2018). Progress in biopharmaceutical development. Biotechnol. Appl. Biochem. 65, 306–322.

Kim, J.Y., Kim, Y.G., and Lee, G.M. (2012). CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. Appl. Microbiol. Biotechnol. 93, 917–930.

Kinch, M.S. (2015). An overview of FDA-approved biologics medicines. Drug Discov. Today 20, 393–398.

Kober, L., Zehe, C., and Bode, J. (2012). Development of a novel er stress based selection system for the isolation of highly productive clones. Biotechnol. Bioeng. 109, 2599–2611.

Kober, L., Zehe, C., and Bode, J. (2013). Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnol. Bioeng. 110, 1164–1173.

Koh, E.Y.C., Ho, S.C.L., Mariati, Song, Z., Bi, X., Bardor, M., and Yang, Y. (2013). An internal ribosome entry site (IRES) mutant library for tuning expression level of multiple genes in mammalian cells. PLoS One 8.

Köhler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. 8, 761–773.

Kotsopoulou, E., Bosteels, H., Chim, Y.T., Pegman, P., Stephen, G., Thornhill, S.I., Faulkner, J.D., and Uden, M. (2010). Optimised mammalian expression through the coupling of codon adaptation with gene amplification: Maximum yields with minimum effort. J. Biotechnol. 146, 186–193.

Kozak, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15, 8125–8148.

Kunert, R., and Reinhart, D. (2016). Advances in recombinant antibody manufacturing. Appl. Microbiol. Biotechnol. 100, 3451–3461.

Kuo, C.C., Chiang, A.W., Shamie, I., Samoudi, M., Gutierrez, J.M., and Lewis, N.E. (2018). The emerging role of systems biology for engineering protein production in CHO cells. Curr. Opin. Biotechnol. 51, 64–69.

Łabno, A., Tomecki, R., and Dziembowski, A. (2016). Cytoplasmic RNA decay pathways - Enzymes and mechanisms. Biochim. Biophys. Acta - Mol. Cell Res. 1863, 3125–3147.

Lai, T., Yang, Y., and Ng, S.K. (2013). Advances in Mammalian cell line development technologies for recombinant protein production. Pharmaceuticals (Basel). 6, 579–603.

Lalonde, M.E., and Durocher, Y. (2017). Therapeutic glycoprotein production in mammalian cells. J. Biotechnol. 251, 128–140.

Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. Nat. Rev. Mol. Cell Biol. 4, 605–612.

Lee, J.S., Grav, L.M., Lewis, N.E., and Kildegaard, H.F. (2015a). CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. Biotechnol. J. 10, 979–994.

Lee, J.S., Kallehauge, T.B., Pedersen, L.E., and Kildegaard, H.F. (2015b). Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. Sci. Rep. 5, 8572.

Lee, J.S., Grav, L.M., Pedersen, L.E., Lee, G.M., and Kildegaard, H.F. (2016). Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. Biotechnol. Bioeng. 9999, 1–17.

Leech, A.A., and Dusetzina, S.B. (2018). Cost-Effective But Unaffordable: The CAR-T Conundrum. JNCI J. Natl. Cancer Inst. 111, 2018–2019.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. 13, 233–245.

Leppek, K., Das, R., and Barna, M. (2018). Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat. Rev. Mol. Cell Biol. 19, 158–174.

Liker, E., Fernandez, E., Izaurralde, E., and Conti, E. (2000). The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. EMBO J. 19, 5587–5598.

Long, C., McAnally, J.R., Shelton, J.M., Mireault, A.A., Bassel-Duby, R., and Olson, E.N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. Science 345, 1184–1188.

LoRusso, P.M., Weiss, D., Guardino, E., Girish, S., and Sliwkowski, M.X. (2011). Trastuzumab emtansine: A unique antibody-drug conjugate in development for human epidermal growth factor receptor 2-positive cancer. Clin. Cancer Res. 17, 6437–6447.

Lu, S., and Cullen, B.R. (2003). Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. RNA 618–630.

Lund, A.M., Kaas, C.S., Brandl, J., Pedersen, L.E., Kildegaard, H.F., Kristensen, C., and Andersen, M.R. (2017). Network reconstruction of the mouse secretory pathway applied on CHO cell transcriptome data. BMC Syst. Biol. 11, 1–17.

Macaraeg, N.F., Reilly, D.E., and Wong, A.W. (2013). Use of an anti-apoptotic CHO cell line for transient gene expression. Biotechnol. Prog. 29, 1050–1058.

Maccani, A., Hackl, M., Leitner, C., Steinfellner, W., Graf, A.B., Tatto, N.E., Karbiener, M., Scheideler, M., Grillari, J., Mattanovich, D., et al. (2014). Identification of microRNAs specific for high producer CHO cell lines using steady-state cultivation. Appl. Microbiol. Biotechnol. 98, 7535–7548.

Madhani, H.D., and Guthrie, C. (1994). Dynamic RNA-RNA interactions in the spliceosome. Annu. Rev. Genet. 28, 1–26.

Maeder, M.L., and Gersbach, C.A. (2016). Genome-editing technologies for gene and cell therapy. Mol. Ther. 24, 430–446.

Majors, B.S., Chiang, G.G., Pederson, N.E., and Betenbaugh, M.J. (2012). Directed evolution of mammalian anti-apoptosis proteins by somatic hypermutation. Protein Eng. Des. Sel. 25, 27–38.

Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., and Embarek, A.M. (2018). Insights into structural and mechanistic features of viral IRES elements. Front. Microbiol. 8, 1–15.

Mason, M., Sweeney, B., Cain, K., Stephens, P., and Sharfstein, S.T. (2012). Identifying bottlenecks in transient and stable production of recombinant monoclonal-antibody sequence variants in chinese hamster ovary cells. Biotechnol. Prog. 28, 846–855.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev. 19, 1512–1517.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. 15, 108–121.

Mathias, S., Fischer, S., Handrick, R., Fieder, J., Schulz, P., Bradl, H., Gorr, I., Gamer, M., and Otte, K. (2018). Visualisation of intracellular production bottlenecks in suspension-adapted CHO cells producing complex biopharmaceuticals using fluorescence microscopy. J. Biotechnol. 271, 47–55.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.J., Chen, Q., et al. (2017). Reversible methylation of m 6 A m in the 5′ cap controls mRNA stability. Nature 541, 371–375.

Mayberry, L.K., Allen, M.L., Dennis, M.D., and Browning, K.S. (2009). Evidence for Variation in the Optimal Translation Initiation Complex: Plant eIF4B, eIF4F, and eIF(iso)4F Differentially Promote Translation of mRNAs. Plant Physiol. 150, 1844–1854.

Mazan-Mamczarz, K., Galbán, S., López de Silanes, I., Martindale, J.L., Atasoy, U., Keene, J.D., and Gorospe, M. (2003). RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation. Proc. Natl. Acad. Sci. U. S. A. 100, 8354–8359.

Mazumder, B., Seshadri, V., and Fox, P.L. (2003). Translational control by the 3′-UTR: The ends specify the means. Trends Biochem. Sci. 28, 91–98.

Melnikov, A., Zhang, X., Rogov, P., Wang, L., & Mikkelsen, T. S. (2014). Massively Parallel Reporter Assays in Cultured Mammalian Cells. Journal of Visualized Experiments, (90).

Mishima, Y., and Tomari, Y. (2016). Codon Usage and 3′ UTR Length Determine Maternal mRNA Stability in Zebrafish. Mol. Cell 61, 874–885.

Mohan, C., Kim, Y.G., Koo, J., and Lee, G.M. (2008). Assessment of cell engineering strategies for improved therapeutic protein production in CHO cells. Biotechnol. J. 3, 624–630.

Moorkens, E., Meuwissen, N., Huys, I., Declerck, P., Vulto, A.G., and Simoens, S. (2017). The market of biopharmaceutical medicines: A snapshot of a diverse industrial landscape. Front. Pharmacol. 8.

Morrison, C. (2018). Fresh from the biotech pipeline—2017. Nat. Biotechnol. 36.

Mugridge, J.S., Coller, J., and Gross, J.D. (2018). Structural and molecular mechanisms for the control of eukaryotic 5′–3′ mRNA decay. Nat. Struct. Mol. Biol. 25, 1077–1085.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. Trends Biochem. Sci. 23, 198–199.

Naldini, L. (2015). Gene therapy returns to centre stage. Nature 526, 351–360.

Narayanavari, S.A., Chilkunda, S.S., Ivics, Z., and Izsvák, Z. (2017). Sleeping Beauty transposition: from biology to applications. Crit. Rev. Biochem. Mol. Biol. 52, 18–44.

Neville, J.J., Orlando, J., Mann, K., Mccloskey, B., and Antoniou, M.N. (2017). Ubiquitous Chromatin-opening Elements (UCOEs): Applications in biomanufacturing and gene therapy. Biotechnol. Adv. 35, 557–564.

Ng, S.K., Tan, T.R.M., Wang, Y., Ng, D., Goh, L.T., Bardor, M., Wong, V.V.T., and Lam, K.P. (2012). Production of Functional Soluble Dectin-1 Glycoprotein Using an IRES-Linked Destabilized-Dihydrofolate Reductase Expression Vector. PLoS One 7.

Nielsen, L., and Borth, N. (2015). Editorial: On the cusp of rational CHO cell engineering. Biotechnol. J. 10, 929–930.

Noderer, W. L. et al. (2014) ' Quantitative analysis of mammalian translation initiation sites by FACS -seq ', Molecular Systems Biology. EMBO, 10(8), p. 748.

Noh, S.M., Shin, S., and Lee, G.M. (2018). Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies. Sci. Rep. 8, 1–11.

O'Toole, P.W., Marchesi, J.R., and Hill, C. (2017). Next-generation probiotics: The spectrum from probiotics to live biotherapeutics. Nat. Microbiol. 2, 1–6.

Oikonomou, P., Goodarzi, H. and Tavazoie, S. (2014) 'Systematic identification of regulatory elements in conserved 3′ UTRs of human transcripts', Cell Reports. Elsevier, 7(1), pp. 281–292.

Okamura, M., Inose, H., and Masuda, S. (2014). RNA export through the NPC in eukaryotes. Genes (Basel). 6, 124–149.

Otsuka, Y., Kedersha, N.L., and Schoenberg, D.R. (2009). Identification of a Cytoplasmic Complex That Adds a Cap onto 5′-Monophosphate RNA. Mol. Cell. Biol. 29, 2155–2167.

Patel, A.K., Kaczmarek, J.C., Bose, S., Kauffman, K.J., Mir, F., Heartlein, M.W., DeRosa, F., Langer, R., and Anderson, D.G. (2019). Inhaled Nanoformulated mRNA Polyplexes for Protein Production in Lung Epithelium. Adv. Mater. 1805116, 1805116.

Peng, R.W., Abellan, E., and Fussenegger, M. (2011). Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells. Biotechnol. Bioeng. 108, 611–620.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. Cell 160, 1111–1124.

Proudfoot, N.J. (2011). Ending the message : poly (A) signals then and now. Genes Dev. 25, 1770–1782.

Puck, T.T., and Kao, F.T. (1967). Genetics of somatic mammalian cells. V. Treatment with 5-bromodeoxyuridine and visible light for isolation of nutritionally deficient mutants. Proc. Natl. Acad. Sci. 58, 1227–1234.

Pybus, L.P., Dean, G., West, N.R., Smith, A., Daramola, O., Field, R., Wilkinson, S.J., and James, D.C. (2014). Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells. Biotechnol. Bioeng. 111, 372–385.

Rabani, M., Pieper, L., Chew, G., & Schier, A. F. (2017). A Massively Parallel Reporter Assay of 3′ UTR Sequences Identifies In Vivo Rules for mRNA Degradation. Molecular Cell, 68(6).

Rader, R.A. (2008). (Re)defining biopharmaceutical. Nat. Biotechnol. 26, 743–751.

Radhakrishnan, A., and Green, R. (2016). Connections Underlying Translation and mRNA Stability. J. Mol. Biol. 428, 3558–3564.

Raftis, E.J., Delday, M.I., Cowie, P., McCluskey, S.M., Singh, M.D., Ettorre, A., and Mulder, I.E. (2018). Bifidobacterium breve MRx0004 protects against airway inflammation in a severe asthma model by suppressing both neutrophil and eosinophil lung infiltration. Sci. Rep. 8, 1–13.

Rajendra, Y., Kiseljak, D., Baldi, L., Wurm, F.M., and Hacker, D.L. (2015a). Transcriptional and post-transcriptional limitations of high-yielding, PEI-mediated transient transfection with CHO and HEK-293E cells. Biotechnol. Prog. 31, 541–549.

Rajendra, Y., Hougland, M.D., Schmitt, M.G., and Barnard, G.C. (2015b). Transcriptional and post-transcriptional targeting for enhanced transient gene expression in CHO cells. Biotechnol. Lett. 37, 2379–2386.

Rajendra, Y., Hougland, M.D., Alam, R., Morehead, T.A., and Barnard, G.C. (2015c). A high cell density transient transfection system for therapeutic protein expression based on a CHO GS-knockout cell line: Process development and product quality assessment. Biotechnol. Bioeng. 112, 977–986.

Rajendra, Y., Peery, R.B., and Barnard, G.C. (2016). Generation of stable Chinese hamster ovary pools yielding antibody titers of up to 7.6 g/L using the piggyBac transposon system. Biotechnol. Prog. 32, 1301–1307.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci. U. S. A. 90, 7923–7927.

Reeve, B., Hargest, T., Gilbert, C., and Ellis, T. (2014). Predicting Translation Initiation Rates for Designing Synthetic Biology. Front. Bioeng. Biotechnol. 2, 1–6.

Reinhart, D., Sommeregger, W., Debreczeny, M., Gludovacz, E., and Kunert, R. (2014). In search of expression bottlenecks in recombinant CHO cell lines - A case study. Appl. Microbiol. Biotechnol. 98, 5959–5965.

Riedel, S. (2005). Edward Jenner and the History of Smallpox and Vaccination. Baylor Univ. Med. Cent. Proc. 18, 21–25.

Rita Costa, A., Elisa Rodrigues, M., Henriques, M., Azeredo, J., and Oliveira, R. (2010). Guidelines to cell engineering for monoclonal antibody production. Eur. J. Pharm. Biopharm. 74, 127–138.

Ritter, A., Voedisch, B., Wienberg, J., Wilms, B., Geisse, S., Jostock, T., and Laux, H. (2016a). Deletion of a telomeric region on chromosome 8 correlates with higher productivity and stability of CHO cell lines. Biotechnol. Bioeng. 113, 1084–1093.

Ritter, A., Rauschert, T., Oertli, M., Piehlmaier, D., Mantas, P., Kuntzelmann, G., Lageyre, N., Brannetti, B., Voedisch, B., Geisse, S., et al. (2016b). Disruption of the gene C12orf35 leads to increased productivities in recombinant CHO cell lines. Biotechnol. Bioeng. 113, 2433–2442.

Rodríguez-Navarro, S., and Hurt, E. (2011). Linking gene regulation to mRNA production and export. Curr. Opin. Cell Biol. 23, 302–309.

Roux, P.P., and Topisirovic, I. (2018). Signaling pathways involved in the regulation of mRNA translation. Mol. Cell. Biol. MCB.00070-18.

Sakuma, T., Takenaga, M., Kawabe, Y., Nakamura, T., Kamihira, M., and Yamamoto, T. (2015). Homologous recombination-independent large gene cassette knock-in in CHO cells using TALEN and MMEJ-directed donor plasmids. Int. J. Mol. Sci. 16, 23849–23866.

Sample, P. J. et al. (2019) 'Human 5′ UTR design and variant effect prediction from a massively parallel translation assay', Nature Biotechnology. Springer US, 37(7), pp. 803–809.

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P. a, and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3′ UTRs and fewer microRNA target sites. Science (80-.). 320, 1643–1647.

Sarantos, K., and Cleo, K. (2013). Analysis of the landscape of biologically-derived pharmaceuticals in Europe: Dominant production systems, molecule types on the rise and approval trends. Eur. J. Pharm. Sci. 48, 428–441.

Saulière, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Crollius, H.R., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. Nat. Struct. Mol. Biol. 19, 1124–1131.

Saunders, F., Sweeney, B., Antoniou, M.N., Stephens, P., and Cain, K. (2015). Chromatin function modifying elements in an industrial antibody production platform - Comparison of UCOE, MAR, STAR and cHS4 elements. PLoS One 10, 1–20.

Schlabach, M.R., Hu, J.K., Li, M., and Elledge, S.J. (2010). Synthetic design of strong promoters. Proc. Natl. Acad. Sci. 107, 2538–2543.

Schoenberg, D.R., and Maquat, L.E. (2012). Regulation of cytoplasmic mRNA decay. Nat. Rev. Genet. 13, 246–259.

Schuller, A.P., and Green, R. (2017). The ABC(E1)s of Ribosome Recycling and Reinitiation. Mol. Cell 66, 578–580.

Schuller, A.P., and Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. Nat. Rev. Mol. Cell Biol. 19, 526–541.

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

Sellick, C.A., Croxford, A.S., Maqsood, A.R., Stephens, G., Westerhoff, H. V., Goodacre, R., and Dickson, A.J. (2011). Metabolite profiling of recombinant CHO cells: Designing tailored feeding regimes that enhance recombinant antibody production. Biotechnol. Bioeng. 108, 3025–3031.

Shalem, O. et al. (2015) 'Systematic Dissection of the Sequence Determinants of Gene 3' End Mediated Expression Control', PLOS Genetics. Edited by H. D. Madhani. Public Library of Science, 11(4), p. e1005147.

Shapiro, J.A. (2009). Revisiting the central dogma in the 21st century. Ann. N. Y. Acad. Sci. 1178, 6–28.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nat. Struct. Mol. Biol. 19, 594–601.

Sperling, J., Azubel, M., and Sperling, R. (2008). Structure and Function of the Pre-mRNA Splicing Machine. Structure 16, 1605–1615.

Spiluttini, B., Gu, B., Belagal, P., Smirnova, A.S., Nguyen, V.T., Hébert, C., Schmidt, U., Bertrand, E., Darzacq, X., and Bensaude, O. (2010). Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. J. Cell Sci. 123, 2085–2093.

Steger, K., Brady, J., Wang, W., Duskin, M., Donato, K., and Peshwa, M. (2015). CHO-S Antibody Titers >1 Gram/Liter Using Flow Electroporation-Mediated Transient Gene Expression followed by Rapid Migration to High-Yield Stable Cell Lines. J. Biomol. Screen. 20, 545–551.

Stern, A.M., and Markel, H. (2005). The history of vaccines and immunization: Familiar patterns, new challenges - If we could match the enormous scientific strides of the twentieth century with the political and economic investments of the nineteenth, the world's citizens might be much heal. Health Aff. 24, 611–621.

Stolfa, G., Smonskey, M.T., Boniface, R., Hachmann, A.B., Gulde, P., Joshi, A.D., Pierce, A.P., Jacobia, S.J., and Campbell, A. (2018). CHO-Omics Review: The Impact of Current and Emerging Technologies on Chinese Hamster Ovary Based Bioproduction. Biotechnol. J. 13, 1–14.

Svitkin, Y. V., Evdokimova, V.M., Brasey, A., Pestova, T. V., Fantus, D., Yanagiya, A., Imataka, H., Skabkin, M.A., Ovchinnikov, L.P., Merrick, W.C., et al. (2009). General RNA-binding proteins have a function in poly(A)-binding protein-dependent translation. EMBO J. 28, 58–68.

Tan, J.G.L., Lee, Y.Y., Wang, T., Yap, M.G.S., Tan, T.W., and Ng, S.K. (2015). Heat shock protein 27 overexpression in CHO cells modulates apoptosis pathways and delays activation of caspases to improve recombinant monoclonal antibody titre in fed-batch bioreactors. Biotechnol. J. 10, 790–800.

Tastanova, A., Schulz, A., Folcher, M., Tolstrup, A., Puklowski, A., Kaufmann, H., and Fussenegger, M. (2016). Overexpression of YY1 increases the protein production in mammalian cells. J. Biotechnol. 219, 72–85.

Thakor, R.T., Anaya, N., Zhang, Y., Vilanilam, C., Siah, K.W., Wong, C.H., and Lo, A.W. (2017). Just how good an investment is the biopharmaceutical sector? Nat. Biotechnol. 35, 1149–1157.

Tian, Z.W., Xu, D.H., Wang, T.Y., Wang, X.Y., Xu, H.Y., Zhao, C.P., and Xu, G.H. (2018). Identification of a potent MAR element from the human genome and assessment of its activity in stably transfected CHO cells. J. Cell. Mol. Med. 22, 1095–1102.

Tjio, J.H., and Puck, T.T. (1958). Genetics of somatic mammalian cells. II. Chromosomal constitution of cells in tissue culture. J. Exp. Med. 108, 259–268.

Urlaub, G., and Chasin, L.A. (1980). Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc. Natl. Acad. Sci. 77, 4216–4220.

Urlaub, G., Käs, E., Carothers, A.M., and Chasin, L.A. (1983). Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. Cell 33, 405–412.

Valvezan, A.J., and Manning, B.D. (2019). Molecular logic of mTORC1 signalling as a metabolic rheostat. Nat. Metab. 1, 321–333.

Viphakone, N., Hautbergue, G.M., Walsh, M., Chang, C.-T., Holland, A., Folco, E.G., Reed, R., and Wilson, S.A. (2012). TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nat. Commun. 3, 1006.

Viphakone, N., Sudbery, I., Griffith, L., Heath, C.G., Sims, D., and Wilson, S.A. (2019). Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. Mol. Cell 1–14.

Wang, W., Jia, Y.L., Li, Y.C., Jing, C.Q., Guo, X., Shang, X.F., Zhao, C.P., and Wang, T.Y. (2017). Impact of different promoters, promoter mutation, and an enhancer on recombinant protein expression in CHO cells. Sci. Rep. 7, 1–10.

Wang, W., Guo, X., Li, Y. mei, Wang, X. yin, Yang, X. jun, Wang, Y. fang, and Wang, T. yun (2018). Enhanced transgene expression using cis-acting elements combined with the EF1 promoter in a mammalian expression system. Eur. J. Pharm. Sci. 123, 539–545.

Wang, X.-Y., Zhang, J.-H., Zhang, X., Sun, Q.-L., Zhao, C.-P., and Wang, T.-Y. (2016). Impact of Different Promoters on Episomal Vectors Harbouring Characteristic Motifs of Matrix Attachment Regions. Sci. Rep. 6, 26446.

Westwood, A.D., Rowe, D.A., and Clarke, H.R.G. (2010). Improved recombinant protein yield using a codon deoptimized DHFR selectable marker in a CHEF1 expression plasmid. Biotechnol. Prog. 26, 1558–1566.

Wissink, E. M., Fogarty, E. A. and Grimson, A. (2016) 'High-throughput discovery of post-transcriptional cis-regulatory elements', BMC Genomics. BioMed Central Ltd., 17(1), p. 177.

Wu, X., and Bartel, D.P. (2017). Widespread Influence of 3′-End Structures on Mammalian mRNA Processing and Stability. Cell 169, 905-917.e11.

Wu, X., Lan, L., Wilson, D.M., Marquez, R.T., Tsao, W.C., Gao, P., Roy, A., Turner, B.A., McDonald, P., Tunge, J.A., et al. (2015). Identification and Validation of Novel Small Molecule Disruptors of HuR-mRNA Interaction. ACS Chem. Biol. 10, 1476–1484.

Xie, F. (2018). Highly Priced Gene Therapies: A Wake-Up Call for Early Price Regulation. Pharmacoeconomics 36, 883–888.

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., et al. (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat. Biotechnol. 29, 735–741.

Yang, Y., Mariati, Chusainow, J., and Yap, M.G.S. (2010). DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. J. Biotechnol. 147, 180–185.

Yates, J.L., Camiolo, S.M., and Bashaw, J.M. (2002). The Minimal Replicator of Epstein-Barr Virus oriP. J. Virol. 74, 4512–4522.

Yeo, J.H.M., Ho, S.C.L., Mariati, M., Koh, E., Tay, S.J., Woen, S., Zhang, P., and Yang, Y. (2017). Optimized Selection Marker and CHO Host Cell Combinations for Generating High Monoclonal Antibody Producing Cell Lines. Biotechnol. J. 12, 1–11.

Yip, A., and Webster, R.M. (2018). The market for chimeric antigen receptor T cell therapies. Nat. Rev. Drug Discov. 17, 161–162.

Yuille, S., Reichardt, N., Panda, S., Dunbar, H., and Mulder, I.E. (2018). Human gut bacteria as potent class I histone deacetylase inhibitors in vitro through production of butyric acid and valeric acid. PLoS One 13, e0201073.

Zhang, L., Inniss, M.C., Han, S., Moffat, M., Jones, H., Zhang, B., Cox, W.L., Rance, J.R., and Young, R.J. (2015). Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. Biotechnol. Prog. 31, 1645–1656.

Zhao, C.P., Guo, X., Chen, S.J., Li, C.Z., Yang, Y., Zhang, J.H., Chen, S.N., Jia, Y.L., and Wang, T.Y. (2017). Matrix attachment region combinations increase transgene expression in transfected Chinese hamster ovary cells. Sci. Rep. 7, 1–7.

Zhao, W. et al. (2014) 'Massively parallel functional annotation of 3′ untranslated regions', Nature Biotechnology. Nature Publishing Group, 32(4), pp. 387–391.

Zhong, X., Ma, W., Meade, C.L., Tam, A.S., Llewellyn, E., Cornell, R., Cote, K., Scarcelli, J.J., Marshall, J.K., Tzvetkova, B., et al. (2018). Transient CHO expression platform for robust antibody production and its enhanced N-glycan sialylation on therapeutic glycoproteins. Biotechnol. Prog. 1–12.

Zhou, Y., Raju, R., Alves, C., and Gilbert, A. (2018). Debottlenecking protein secretion and reducing protein aggregation in the cellular host. Curr. Opin. Biotechnol. 53, 151–157.

2. Materials and Methods

2.1. Buffers and Solutions

Buffer/Solution	Composition
5x TBE – Tris/Borate/EDTA	0.45M Tris Base, 0.45M Boric Acid, 10mM EDTA pH8.0.
LB Media	10g Bacto-tryptone, 5g yeast extract, 170mM NaCl. pH 7.5 with NaOH. Water to 1l.
Nutrient Agar	2.3% Agar.
10x PBS –Phosphate Buffered Saline	1.4M NaCl, 25mM KCl, 0.1M Na_2HPO_4, 18mM KH_2HPO_4, pH to 7.4 with NaOH.
Lysis Buffer	50mM HEPES pH7.5, 100mM NaCl.
4x Upper Buffer	100mM Tris-HCl (pH 6.8), 4% SDS, 0.2% bromophenol blue, 20% glycerol, 200mM β-mercaptoethanol.
4x Lower Buffer	1.15M Tris HCl pH8.8, 14mM SDS.
Transfer Buffer	25mM Tris, 190mM glycine, 20% glycine.
Blocking Buffer	0.5x TBS, 2% Tween, 2.5g Milk Powder, Water to 50ml.
Wash Buffer	1x TBS, 0.2% Tween.
ECL1	0.1M Tris pH 8.5, 2.5mM Luminol, 0.4M p-Coumaric acid.
ECL2	0.1M Tris pH 8.5, 18μM H_2O_2.
Colouring substrate	2M Diethanolamine, 1mM $MgCl_2$, 0.5mM $ZnCl_2$.
10x TBS – Tris Buffered Saline	1M Tris, 6.84M NaCl. pH to 7.6 with HCl.
5x Running Buffer	0.125M Tris, 1.25M Glycine, 17mM SDS.
FIX Solution	4% Formaldehyde, 1% PBS.
TX Solution	0.5% Triton X-100 (vol/vol), 1% PBS.
ELISA Wash Buffer	1X PBS, 0.05% Tween (vol/vol).

The generic chemicals in these buffers and solutions were purchased from:

- Thermo Fisher, Massachusetts, USA
- Sigma Aldrich, Missouri, USA

2.2. Vector Construction

2.2.1. PP7d-Fusion Proteins

In order to create PP7d-fusion proteins, a CHO codon optimised Kozak-PP7d sequence was inserted by Gibson assembly (NEB, Massachusetts, USA) into a pcDNA5 FRT TO expression vector (Thermo Fisher, Massachusetts, USA, USA), directly downstream of the CMV promoter. Fusion proteins were then created by Gibson assembly, removing the PP7d stop codon in the process, by mismatching of primer-template base pairing.

2.2.2. TU

Expression of CHO codon optimised reporter and non-fusion test genes were driven from an in-house AstraZeneca proprietary transcription-unit (TU) vector, under a CMV promoter (Patel *et al*., in press). For studies using the Ef1α promoter, the CMV promoter and 5'UTR were removed from TU, and replaced by their counterpart Ef1α sequences.

2.2.3. MGVT

For expression of reporters utilising OriP-mediated amplification (Daramola *et al.*, 2014), TU expression cassettes were transferred by golden gate cloning to an in-house multi-gene vector backbone (MGVT), containing the required OriP sequence (Patel *et al.*, in press).

2.3. Molecular Cloning

2.3.1. PCR

PCR was performed with the NEB Q5 2X Master Mix kit (NEB, Massachusetts, USA), as per the manufacturer's instructions, with the following generic reaction mixture:

- 12.5µl Q5 2X Master Mix
- 9µl Nuclease-free water
- 1.25µl 10µM Forward Primer
- 1.25µl 10µM Reverse Primer
- 1µl 10ng/µl Template DNA

Appropriate annealing temperatures were identified using the NEB Tm Calculator online tool.

2.3.2. SDM

To perform Site-Directed Mutagenesis, the NEB Q5 Site-Directed Mutagenesis kit (NEB, Massachusetts, USA) was used, following manufacturer instructions, with primers designed using the NEBaseChanger online tool. After PCR amplification, reaction mixes were taken forward into the Kinase-Ligase-DpnI reaction, with the following generic reaction mixture:

- 2.5µl 2X KLD Buffer
- 1.5µl Nuclease-free Water
- 0.5µl PCR mix
- 0.5µl KLD enzyme mix

Each 5µl KLD reaction mix was then transformed into competent DH5*α* *E.coli* (NEB, Massachusetts, USA), as described below.

2.3.3. Gibson Assembly

To perform Gibson Assembly, DNA backbone and inserts were first amplified by PCR, adding linkers with primers designed using the NEBuilder online tool. Template DNA was then degraded by addition of 1µl DpnI restriction enzyme (NEB, Massachusetts, USA) and incubation at 37°C for 1 hour. PCR mixes were then run on an agarose electrophoresis gel, and the appropriate bands extracted using the QIAquick Gel Extraction kit (Qiagen, Hilden, Germany). Finally, each fragment was taken forward to NEB Gibson Assembly (NEB, Massachusetts, USA), with the following generic reaction mixture:

- ~100ng of DNA fragments (3:1 insert:backbone molar ratio)
- 5µl Gibson Assembly 2X Master Mix
- Nuclease-free water to 10µl

This reaction mixture was incubated at 50°C for 1 hour, before being transformed into competent DH5*α* *E.coli* (NEB, Massachusetts, USA), as described in section 2.3.7.

2.3.4. Colony PCR

Transformed *E.coli* were screened for the correctly constructed plasmid by colony PCR. This was performed with Biotaq Polymerase (Bioline, London, UK). A typical reaction mix with Biotaq Polymerase:

- 2µl 10x Biotaq Buffer
- 0.6µl 50mM $MgCl_2$
- 2µl dNTPs
- 0.6µl 10µM For Primer
- 0.6µl 10µM Rev Primer
- 0.5µl Biotaq Polymerase
- 13.7µl dH_2O

The template was added by picking a numbered *E.coli* colony with a P10 pipette tip, and dipping it inside the PCR tube. The reaction scheme involved an initial denaturation at 95°C for 3 minutes, extending at 72°C for 1 minute/kb with 30 cycles, and a final extension at 72°C for 5 minutes.

2.3.5. Golden Gate Cloning

For insertion of a transcription unit into the in-house MGEV backbone, the NEB Golden Gate Assembly kit (NEB, Massachusetts, USA) was used, as per the manufacturer's instructions, with the following generic reaction mix:

- 75ng MGEV backbone
- 75ng insert
- 2µl 10X NEB Golden Gate Buffer
- 1µl NEB Golden Gate Assembly
- Nuclease-free water to 20µl

The reaction mixture was incubated at 37°C for 1 hour, followed by 60°C for 5 minutes, before 10µl was transformed into competent DH5α *E.coli* (NEB, Massachusetts, USA), as described in section 2.3.7.

2.3.6. Ligation

In order to clone by ligation, DNA was first digested by the appropriate NEB restriction enzyme (NEB, Massachusetts, USA), with the following generic reaction mixture.

- 3µl CutSmart 10X Buffer
- 1µl Restriction Enzyme 1
- 1µl Restriction Enzyme 2
- 1-3µg DNA
- Nuclease-free water to 30µl

Following digestion, reaction mixes were run on an agarose electrophoresis gel, before being gel extracted with the MinElute Gel Extraction Kit (Qiagen, Hilden, Germany). Gel extracted fragments were then ligated using the NEB Quick Ligation Kit (NEB, Massachusetts, USA), as per the manufacturer's instructions, with the following generic reaction mix:

- 10µl 2X Quick Ligation Buffer
- ~100ng DNA fragments (3:1 insert:backbone molar ratio)
- 1µl Quick Ligase
- Nuclease-free water to 20µl

This reaction mixture was incubated at room temperature for 10 minutes, before 10µl was transformed into competent DH5α *E.coli* (NEB, Massachusetts, USA), as described in section 2.3.7.

2.3.7. DH5α Transformation

Competent DH5α were transformed with DNA by heat shock. To do this, 50µl frozen DH5α *E.coli* (NEB, Massachusetts, USA) were mixed with the appropriate volume of DNA, and left on ice for 30 minutes. They were then heated to 42°C for 30 seconds, before being returned to ice for 2 minutes. 950µl of SOC media (NEB, Massachusetts, USA) was then added, and the cells incubated in a shaking incubator at 37°C for 1 hour. After this, the cells were spun at 6000rcf for 3 minutes, the supernatant removed, and the cells resuspended in 100µl LB medium (Thermo Fisher, Massachusetts, USA) for plating. *E.coli* were plated on agar plates, containing 100µg/ml Ampicillin (Thermo Fisher, Massachusetts, USA) or 50µg/ml Kanamycin (Thermo Fisher, Massachusetts, USA), as appropriate.

2.3.8. Plasmid Preparation

E.coli transformed with DNA were inoculated in liquid LB media (Thermo Fisher, Massachusetts, USA), containing either 100µg/ml Ampicillin (Thermo Fisher, Massachusetts, USA) or 50µg/ml Kanamycin (Thermo Fisher, Massachusetts, USA), and incubated with shaking at 37°C overnight. The cells were harvested by spinning at 6000rcf, for a time appropriate to the volume of medium. Plasmids were then isolated using Qiagen Mini/Midi/Maxiprep kits (Qiagen, Hilden, Germany), with protocols detailed by the manufacturer.

2.3.9. Sequencing

Correct molecular cloning was verified by the Sheffield Medical School Sanger sequencing service. DNA to be sequenced was diluted to 100ng/µl, primers to be used were diluted to 1µM and both were sent in an envelope to be sequenced.

2.4. Cell Culture

2.4.1. HEK-293T

HEK293T cells were subcultured every 3-4 days in T-75 and T-25 flasks (Corning, New York, USA) and kept in a static incubator at 37°C and 5% CO_2. Cells were washed in warm 1x PBS and trypsinised (Thermo Fisher, Massachusetts, USA), before being passaged into new medium, at approximately a 1:4 ratio, depending on their confluence. Cells were counted using a haemocytometer when plating cells for experiments.

Appropriate volumes of cell culture were used when incubated in various culture platforms, as displayed below:

Culture Platform	Culture Volume / ml
T-75 Flask (Corning, New York, USA)	20
T-25 Flask (Corning, New York, USA)	6
6cm Dish (Corning, New York, USA)	5
24-Well Plate (Corning, New York, USA)	0.5

2.4.2. CHO

CHO cells were subcultured every 3-4 days in chemically defined CD-CHO medium (Gibco, Massachusetts, USA), their viable cell densities measured using a ViCell XR machine (Beckman Coulter, California, USA), and seeded at $0.2*10^6$cells/ml. Cells were incubated at 37^0C and 5% CO_2, being shaken at 140RPM in Erlenmeyer flasks (Corning, New York, USA), 240RPM in cultiflasks (Sartorius, Göttingen, Germany) and 320RPM in plates. Different CHO cell lines were additionally grown in supplemented media, as shown below.

Cell Line	CD-CHO Supplement
Transient Host	25μM MSX, 100μg/ml Hygromycin B (Sigma Aldrich, Missouri, USA)
ETE mAb Stable Producer	50μM MSX

Appropriate volumes of cell culture were used when incubated in various culture platforms, as displayed below:

Culture Platform	Culture Volume / ml
250ml Erlenmeyer Flask (Corning, New York, USA)	60
125ml Erlenmeyer Flask (Corning, New York, USA)	30
Cultiflask (Sartorius, Göttingen, Germany)	10
24 Deep-Well Plate (Greiner Bio-One, Kremsmünster, Austria)	3
24 Shallow-Well Plate (Thermo Fisher, Massachusetts, USA)	0.6
96 Deep-Well Plate (Greiner Bio-One, Kremsmünster, Austria)	0.5

2.4.3. Cell Culture Feeding

For the transient fed-batch process described in Chapter 5 of this thesis, cells were treated with proprietary AstraZeneca feeds (AstraZeneca, Cambridge, UK). Day 0 feeds were applied 4 hours post-transfection. Each Feed was administered individually by direct pipetting into the culture vessel.

2.4.4. Multi-well Cell Growth Measurement

For experiments performed in multi-well plates, presenting too small a culture volume for measurement by ViCell, one of three methods were performed to monitor cell culture growth.

2.4.4.1. Bradford Assay

To measure cellular protein abundance as a proxy for cell growth, cell lysates of HEK-293T cells were first collected, as described in section 2.7.1. 10μl of this cell lysate was mixed with 990μl Bradford 1X dye reagent (Bio-Rad, California, USA). After 1 minute incubation, the absorbance of the sample was measured at 595nm, on a spectrophotometer blanked with 1ml Bradford 1x dye reagent.

After substitution of the of the blank sample's aborbance, fold change in protein concentration was directly inferred from fold change in absorbance.

2.4.4.2. Prestoblue Assay

To measure fold-change in viable cell density by Prestoblue assay, a 1:1 volume:volume mix of Prestoblue Cell Viability Reagent (Thermo Fisher, Massachusetts, USA) and CD-CHO medium (Gibco, Massachusetts) was created. After shaking, 100µl of cell culture was aspirated from the culture vessel, transferred to a clear-bottom 96-well plate (Thermo Fisher, Massachusetts, USA), and 100µl CD-CHO added to two 'blank' wells. 22.5µl of Prestoblue:CD-CHO mix was then added to each well of the plate, which was shaken for 20 seconds, and placed in a static 37°C 5% CO_2 incubator. After incubation for 30 minutes, fluorescence of each well was measured in a plate reader, with a 560nm excitation wavelength, and 590nm emission wavelength.

After substitution of the average of the blank well's fluorescence, fold change in cell growth was directly inferred from fold change in fluorescence across the plate.

2.4.4.3. Iprasense

To measure viable cell density and viability in multi-well culture platforms, the Iprasense Norma HT (Iprasense, Montpellier, France) cell counter was used, as per manufacturer's instructions. Where necessary, CD-CHO was used to dilute cell culture to the appropriate density for each slide, as described below.

Iprasense Slide Depth	Minimum Cell Density	Maximum Cell Density
100µm	0 cells/ml	$4*10^6$ cells/ml
20µm	$4*10^6$ cells/ml	$8*10^6$ cells/ml

2.5. Transfection

2.5.1. HEK-293T PEI

HEK-293T cells were transiently transfected using PEI (Thermo Fisher, Massachusetts, USA, USA) at 1µg/ml. In 24 well plates, $1*10^5$ cells per well were used, and $1*10^6$ cells per 6cm dish.

A weight ratio of 1:5 DNA:PEI was mixed with serum-free medium (Sigma Aldrich, Missouri, USA), to a volume equal to 1/6 of the total medium volume of the dish being transfected. This was mixed, centrifuged down using a pulse, and incubated for 20 minutes. The mixture was then added to the cells in medium, dropwise.

2.5.2. CHO Electroporation

To transfect CHO cells using electroporation, cells were first split into unsupplemented CD-CHO medium (Gibco, Massachusetts, USA) for the two passages prior to transfection. Electroporation was performed with the Nucleofector 2B Device, with both electroporation cuvettes, and a 96-well shuttle (Lonza, Basel, Switzerland). On the day of transfection, the appropriate volume of unsupplemented medium was added into the culture platform, and incubated in a 37°C 5% CO_2 shaking incubator.

DNA mixes were created by diluting the appropriate quantity of DNA (800ng for 96 deep-well plates, 10µg for cultiflasks) 1:4 in nucleofector solution. 5*10^6 cells per ml of desired transfected culture were spun down at 200G for 5 minutes, and the supernatant aspirated and discarded. The cells were resuspended in the appropriate volume of Nucleofector solution (15µl for 96 deep-well plates, 100µl for cultiflasks). This cell suspension was then added to the DNA mix, and electroporated.

Electroporated cells were diluted in the appropriate volume of unsupplemented CD-CHO for the desired cell density, seeded in the pre-gassed culture platform, and placed in the incubator.

2.5.3. CHO PEI

To transfect CHO cells using PEI, cells were first split into unsupplemented CD-CHO medium (Gibco, Massachusetts, USA) for the two passages prior to transfection. On the day of transfection, cells were seeded at 1*10^6 cells/ml in unsupplemented medium, in the appropriate culture platform. For the 7-day fed-batch process described in Chapter 5, cells were seeded at 4*10^6cells/ml. Cells were transfected with DNA with a maximum quantity of 1.2µg per 1*10^6 cells. The appropriate volume of DNA was mixed at a 1:1 ratio with 300mM NaCl (Sigma Aldrich, Missouri, USA). 25kDa PEI (Thermo Fisher, Massachusetts, USA, USA) was also mixed 1:1 with 300mM NaCl, with a 5:1 DNA:PEI ratio of PEI to DNA transfected.

The DNA/NaCl and PEI/NaCl mixes were then combined, and mixed by pipetting up and down. After 1-minute incubation, the transfection mixes were added to cell culture.

2.6. SEAP Assays

To assay for concentration of SEAP, two different methodologies were used: an in-house method for supernatant collected from HEK-293T cells, and the Sensolyte pNPP SEAP detection kit (Anaspec, Fremont, USA) for supernatant collected from CHO cells.

2.6.1. In-House

One tablet of P-nitrophenyl phosphate (Sigma Aldrich, Missouri, USA) was added to 20ml of colouring substrate, and mixed on a shaking plate for 20 minutes.

Medium was aspirated from adherent cells by pipette. It was then spun at 500G at 4°C for 4 minutes. The supernatant was then aspirated, leaving at least 50µl medium along with the pellet at the bottom of the tube. The appropriate amount of supernatant was then added to a 1:10 mixture of 0.05% CHAPS in PBS (Thermo Fisher, Massachusetts, USA, USA) with mixed colouring substrate. The reaction was incubated in the dark for 2-4 hours. After this, the absorbance of the mixture was measured in a spectrophotometer at 405nm.

If supernatant was not assayed on the day of collection, it was flash-frozen in liquid nitrogen, followed by storage in a -80°C freezer, before being thawed on ice on the day of assaying.

2.6.2. Sensolyte

Cell culture was harvested, and centrifuged at 1000G for 5 minutes, before supernatant was aspirated. If necessary, harvested supernatant was diluted using CD-CHO (Gibco, Massachusetts, USA). 50µl of supernatant was aliquoted into a 96-well plate (Thermo Fisher, Massachusetts, USA), before 50µl of Sensolyte component A (Anaspec, Fremont, USA) was added. Absorbance at 405nm was measured after 15 minutes incubation in a plate reader, using a well containing only CD-CHO and component A as a blank.

2.7. Western Blot

2.7.1. Cell Lysis

Medium was aspirated from adherent cells, or from centrifuged cell pellets. The cells were washed with 1% PBS (Thermo Fisher, Massachusetts, USA, USA) of equivalent volume to the medium, before 1/6 of this volume of Lysis buffer was added. The buffer/cell mixture was then collected and centrifuged at 500G for 5 minutes, before the cell lysate supernatant was taken. The protein concentration in the cell lysate was then assessed by Bradford assay (Bio-Rad, California, USA).

2.7.2. Protein Gel

Proteins were run on a Polyacrylamide gel, of which the percentage depended on the size of the protein being probed for. The Polyacrylamide gel consisted of a 5% upper gel, and a 10-15% lower gel. Constituents for different gel percentages are listed below:

5% Loading Gel

4x Upper Buffer	4ml
30% Acrylamide/0.8% Bisacrylamide (Bio-Rad, California, USA)	1.2ml
10% APS (Sigma Aldrich, Missouri, USA)	110µl
TEMED (Sigma Alrdich, Missouri, USA)	20µl
Water	Up to 10ml

10-15% Running Gel

4x Lower Buffer	4ml
30% Acrylamide/0.8% Bisacrylamide	3.3-5ml
10% APS	110µl
TEMED	20µl
Water	Up to 10ml

Proteins were denatured before gel loading by boiling at 95°C for three minutes and exposure to 1/6 diluted 2-mercaptoethanol (Sigma Aldrich, Missouri, USA). Samples were loaded onto gel using a gel-loading tip, and Thermo Scientific PAGEruler (Thermo Fisher, Massachusetts,

USA, USA). Gels were run in running buffer at 30mA, until the PAGEruler had sufficiently progressed down the gel.

2.7.3. Transfer and Blotting

Transfer to nitrocellulose membrane (Bio-Rad, California, USA) was performed by a Trans Blot Turbo machine (Bio-Rad, California, USA). The nitrocellulose and filter papers were soaked in transfer buffer, before the sandwich was stacked, and the protein was transferred at 25V for 7 minutes. After transfer, the nitrocellulose was briefly stained with Ponceau S (Sigma Aldrich, Missouri, USA), to check that proteins had transferred, before being washed with water.

Nitrocellulose blots were blocked with blocking buffer on a shaker, for one hour. They were then exposed to the primary antibody in blocking buffer at the appropriate concentration, and left on a shaker for at least an hour. After three short (10 seconds) and three long (10 minutes) washes with Wash Buffer, the secondary antibody is added in blocking buffer at the appropriate concentration, on a shaker for 1 hour. After three more short and long washes, the gel was visualised by chemiluminescence, after addition of a 1:1 mixture of ECL1 and ECL2 solutions.

2.8. Immunofluorescence

To perform immunofluorescence, cells were first seeded in a well of a 24-well plate containing a coverslip.

To fix the cells, medium was removed from the well, before the well was washed twice in 1X PBS. After this, 200µl FIX solution was added, and left at room temperature for 30 minutes.

To permeabilise the cells, the FIX solution was removed, before the well was washed with 1X PBS (Sigma Aldrich, Massachusetts, USA). 200µl of TX solution was then added, before the wells were incubated on ice for 10 minutes. TX solution was then removed, before the cells were washed twice more in 1X PBS.

After this, the cells were incubated with 100µl 1% BSA (Sigma Aldrich, Massachusetts, USA) in 1X PBS for 1 hour. This solution was removed, before the well was incubated with the primary antibody, diluted in 100µl 1% BSA in 1X PBS for 1 hour. The well was then washed a

further three times in 1X PBS. The well was then incubated with secondary fluorescent antibody, diluted 1/800 with 100µl 1% BSA in 1X PBS, covered in foil, for 30 minutes. Following this, the well was washed three final times with 1X PBS. The coverslip was then removed from the well, dried, and mounted with DAPI. Nail varnish was used to seal the coverslip, which could be stored at 4°C in the dark.

2.9. RT-qPCR

2.9.1. HEK-293T

In order to extract RNA from cells, medium was first aspirated from cells, before they were washed with 1X PBS. Cells were then lysed in a volume of lysis mix equivalent to 5% of the total medium in the dish. The components for 1.2ml lysis mix are as follows:

- 1150µl Lysis Buffer
- 48µl Protease Inhibitors (Thermo Fisher, Massachusetts, USA, USA)
- 1µl DTT (Thermo Fisher, Massachusetts, USA, USA)
- 1µl Ribosafe RNase Inhibitor (Bioline, London, UK)

After lysis cells were scraped off the plate, and pipetted into an Eppendorf tube, they were centrifuged at 16,100G for 5 minutes at 4°C, the supernatant aspirated, and the pellet discarded. A 3:1 ratio of TRIzol LS (Thermo Fisher, Massachusetts, USA, USA) was added to the lysis mix, and the solution was homogenised by pipetting up and down. The samples were left at room temperature for 10 minutes, before 200µl chloroform (Sigma Aldrich, Missouri, USA) was added, the tubes were hand-shaken for 30 seconds, and left at room temperature for a further 10 minutes.

Samples were centrifuged at 12,000G for 15 minutes at 4°C, before the colourless supernatant was added to a CrystalClear tube (Starlab, Milton Keynes, UK) containing 1µl of Glycogen (Sigma Aldrich, Missouri, USA). A 1:1 ratio of isopropanol (Thermo Fisher, Massachusetts, USA, USA) was then added, before the sample was mixed by inversion and left at room temperature for 10 minutes. Following this, samples were centrifuged at 12,000G for 15 minutes at 4°C. The supernatant was aspirated and discarded, leaving a white RNA pellet. This pellet was washed with 1.2ml of ice-cold 75% ethanol (Thermo Fisher, Massachusetts, USA, USA), by vortexing, and centrifuged at 7,500G for 5 minutes at 4°C. The supernatant was removed, leaving a white RNA pellet, which was left to air-dry in an open tube in a fume

cupboard. Once the white RNA pellet became colourless and translucent, it was stored at -80°C.

To treat RNA samples with DNase, RNA pellets were first re-suspended in 43µl in $H_2O_{RNase\text{-}free}$, by vortexing, incubating on ice for 10 minutes, and vortexing again. The following Turbo DNase (Thermo Fisher, Massachusetts, USA, USA) mix was then added per sample:

- 5µl 10X DNase Buffer
- 2µl Turbo DNase
- 1µl Ribosafe RNase Inhibitor

This reaction mix was incubated at 37°C for 1 hour, shaking gently, at no more than 350RPM. 50µl $H_2O_{RNase\text{-}free}$, and 100µl acidic phenol (Thermo Fisher, Massachusetts, USA, USA) were then added, before the sample was hand-shaken for 30 seconds, and left at room temperature for 5 minutes. The samples were centrifuged at 12,000G for 5 minutes at 4°C. 90µl supernatant was then aspirated, added to 1µl Glycogen (Sigma Aldrich, Missouri, USA) and 10µl Na-Acetate (Sigma Aldrich, Missouri, USA) (3M, pH5.8), and mixed by inversion. 250µl 100% ethanol was then added, and mixed again by inversion. The sample was then incubated for at least 20 minutes at -24°C in a freezer.

Samples were then centrifuged at 12,000G for 30 minutes at 4°C, and supernatant removed, revealing a white RNA pellet. This pellet was washed with 1.2ml of ice-cold 75% ethanol, by vortexing, and centrifuged at 7,500G for 5 minutes at 4°C. The supernatant was removed, leaving a white RNA pellet, which was left to air-dry in an open tube in a fume cupboard.

The pellets were then re-suspended in 40-50µl $H_2O_{RNase\text{-}free}$, by vortexing, incubating on ice for 10 minutes, and vortexing again. RNA concentration of the sample was then assessed by Nanodrop (Thermo Fisher, Massachusetts, USA, USA).

cDNA synthesis was performed on 1µg of RNA, using a Bioscript reverse transcription kit (Bioline, London, UK), as instructed by the manufacturer. The resultant cDNA was then stored at -24°C.

To perform qPCR, cDNA was diluted 5X in $H_2O_{RNase\text{-}free}$. If comparative qPCR against 18S rRNA was used, a further 100X dilution in $H_2O_{RNase\text{-}free}$ of this diluted cDNA was prepared in a separate tube for the 18S rRNA qPCR tubes. A Sensimix kit (Bioline, London, UK) was then

used to performed qPCR, as per manufacturer instructions, with the following run parameters:

- Anneal at 59°C or more for 15 seconds.
- Extend for 25 seconds.
- Cycle 45 times.

Relative RNA levels were calculated by the -ΔΔCt method, against 18S rRNA as a housekeeping gene.

2.9.2. CHO

To begin RT-qPCR, 1*10^6 cells were centrifuged at 200G for 5 minutes, before being resuspended and passed through a QIAshredder column (Qiagen, Hilden, Germany). RNA was extracted from the pellets using an RNeasy Miniprep kit (Qiagen, Hilden, Germany), as per manufacturer's instruction, and eluted in 30µl RNAse-free water. The concentration of this RNA was then measured using a Nanodrop One machine (Thermo Fisher, Massachusetts, USA, USA). 800ng of each RNA sample was then brought forward to cDNA synthesis using the QuantiTect Reverse Transcription kit (Qiagen, Hilden, Germany), as per manufacturer's instructions. cDNA was diluted to the appropriate concentration in RNAse-free water, followed by qPCR, using the Quantifast SYBR Green PCR kit (Qiagen, Hilden, Germany), using the following generic reaction mix:

- 8ul Nuclear-free water
- 2.5ul Primer Mix (1µM of each primer)
- 2ul cDNA
- 12.5ul SYBR Green Master Mix 2X

All qPCR reactions were run with the following settings:

- Anneal at 60°C for 15 seconds
- Extend for 20 seconds
- Cycle 40 times

Relative RNA levels were calculated by the -ΔΔCt method, against *Fkbp1a* RNA as a housekeeping gene, as recommended by Brown *et al.*, 2018.

Primers were designed using the Primer-BLAST online tool, and validated by serial dilution of cDNA before qPCR, allowing calculation of primer efficiency using the Thermo Fisher qPCR Efficiency Calculator online tool.

2.10. Titre Assays

2.10.1. IgG

IgG concentration in cell supernatant was assessed by firstly centrifuging CHO cells at 200G for 5 minutes, before removing the supernatant from the pellet. Concentration of IgG within this supernatant was then measured using the FastELISA Human IgG Quantification kit (2BScientific, Upper Heyford, UK), using the manufacturer's instructions.

2.10.2. EPO

To measure concentration of EPO in cell supernatant, CHO cells were centrifuged at 200G for 5 minutes, and the supernatant removed from the pellet. Concentration of EPO within this supernatant was then assessed with the Human Erythropoietin simplestepELISA Kit (ab211647) (Abcam, Cambridge, UK), as per the manufacturer's instructions.

2.10.3. Sc-FP

To measure titre of sc-FP proteins, CHO cells were centrifuged at 1000G for 5 minutes, and the supernatant removed from the pellet. Concentration of sc-FP within this supernatant was then assessed using the Octet HTX Machine (Molecular Devices, California, USA), as per the manufacturer's instructions.

2.11. Culture Supplementation

For experiments in which cell culture was supplemented with chemical agents, cells in 96 deep-well plates (Greiner Bio-One, Kremsmünster, Austria) were treated 24 hours post-transfection. Plates were removed from the shaking incubator, and left to stand for 5 minutes in a static 37°C 5% CO_2 incubator, allowing cells to settle in the bottom of each well. 100µl of conditioned medium was then aspirated slowly from the meniscus of each well. Finally, 100µl of one of Leucine/Arginine/Glutamine (Thermo Fisher, Massachusetts, USA, USA), Efficient Feed (Gibco, Massachusetts, USA), or Rapamycin (Sigma Aldrich, Missouri, USA) dissolved to the appropriate concentration in pre-warmed CD-CHO medium (Gibco, Massachusetts, USA), was added to each well. Control cells were treated in the same way, and supplemented with CD-CHO alone.

Master stocks of chemicals were prepared by dissolving the appropriate concentration of each supplement in CD-CHO medium, and sterilised by passing through a 0.22μM filter (Cole-Palmer, Illinois, USA). For chemicals requiring DMSO as a diluent (e.g. Rapamycin), control cells were treated with CD-CHO containing the equivalent concentration of DMSO (Sigma Aldrich, Missouri, USA).

2.12. mRNA Stability Assay

To measure stability of mRNA transcripts, cells were first transfected with the gene to be measured, using the PEI method detailed in section 2.5.3.

For timepoint 0h, $1*10^6$ cells were spun down at 200G for 5 minutes, and cell pellets harvested as per the CHO RT-qPCR protocol detailed in section 2.9.2. 48h-post transfection, cells were treated with 20μg/ml Actinomycin D (Sigma Aldrich, Missouri, USA), administered via addition to cell culture of 1mg/ml Actinomycin D stock in DMSO (Sigma Aldrich, Missouri, USA). Cell pellets were similarly harvested at every appropriate timepoint following Actinomycin D treatment. RT-qPCR was then performed on each sample, in accordance with the method detailed above, with the quantity of each transcript of interest measured by relative abundance compared to the housekeeping gene *fkbp1a*, as recommended by Brown *et al.*, 2018. mRNA half-life was calculated by non-linear regression of each transcripts' relative abundance against time points post-Actinomycin D treatment, utilising a one-phase exponential decay curve.

2.13. Sequence Elements

The following is a list of all cis sequence elements used in this thesis, with a short summary of how they were sourced, and incorporated into their vectors.

2.13.1. Coat Protein Binding Sites

Coat protein binding sites sequence are well established in literature (Chao *et al.,* 2008, 2012). Modules, containing 3 repeats of sequences of 2 stem-loops (totalling 6 binding sites), were isolated by PCR from a pre-existing in-house laboratory plasmid based on these sequences, and placed downstream of the 3'UTR, directly upstream of the PolyA signal in the AstraZeneca in-house TU1 expression vector.

Name	Sequence
2xMS2bs	TCGAGGAATTCCGATCCTCTAGAGTCGACCTGCAGACATGGGTG ATCCTCATGTTTTCTAGAGTCGACCTGCAGACATGGGTGATCCTC ATGTTTTCTAGAGTCGACCTGCAGGCATGCAAGCTGCCCGGGGG ATCCACTAGTTCTAGCCGGAATTCCTCGA
2xPP7bs	GATCCGTAGAAATAAGGAGTTTATATGGAAACCCTTACTGCTGG GAGATCCGTAGAAATAAGGAGTTTATATGGAAACCCTTACTGCT GGGA

2.13.2. Triple Helices

CHO MALAT1 and MENβ triple helices sequences were identified by aligning established human triple helices sequences (Brown *et al.*, 2012, 2014b), with the CHO-K1 genome, taking the region of homology as the CHO triple helices. The PAN triple helix sequence was taken from literature (Mitton-Fry *et al.*, 2010; Tycowski *et al.*, 2012). Triple helices were placed directly downstream of the 3'UTR, in the AstraZeneca in-house TU1 vector, inserted between the SbfI and NotI restriction sites, both with and without a PolyA sequence downstream, between the NotI and KpnI restriction sites, by restriction/ligation cloning.

Name	Sequence
MALAT1	CAGTAGGGCTGTAAAGGTTTTTCTTTTCCTGAGAAAACAAACTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTAAAAAAAAAGCAAAAGACACTGGTGGCCGGCACTCCTGGTCTCCAGGACGGGGTTCAAATCCCTGCGGTGTCT
MENβ	GTAGGGCTGTAAAGGTTTTTCTTTTCCTGAGAAAACAAACTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTAAAAAAAAAGCAAAAGACACTGGTGGCCGGCACTCCTGGTCTCCAGGACGGGGTTCAAATCCCTGCGGTGTCT
PAN	TGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATT

2.13.3. Stability Elements

34nt stability elements were extracted from the data of Oikonomou *et al.,* 2014, and placed directly downstream of the reporter CDS-bordering SbfI restriction site in the AstraZeneca in-house TU1 vector, upstream of the 3'UTR, by site-directed mutagenesis. How stability elements were selected from the large database is described in detail in Chapter 4.

Name	Oikonomou *et al.,* 2014 SequenceID	Sequence
SE1	C3U-seq15901	TTTGTTTTAGATGGAATAGCACAAGGAGAAAAAT
SE2	C3U-seq10578	GTTTTTTGAGGAATCTCAAGATGTGATATATTGG
SE3	C3U-seq4185	ATGTCTCCAGTTACAACTCCGCAGTGGATGTGAA
SE4	C3U-seq8401	GCAATTTAGCATGTTGGAACGTCTAGGGAGAAGG
SE5	C3U-seq12424	TCTCATTCCAGTAAGGCAGTTAGACACTTGAGTT
SE6	C3U-seq5465	CAGTTGAGATGAAGCACGTCGTTAGAACGTTGTT
SE7	C3U-seq212	AAAATGTAAAAATGTAACTATAGCATATGAATTG
SE8	C3U-seq1204	AAGTGGAGGTCTGGTTTGTAACTTTCCTTGTACT
SE9	C3U-seq8665	GCCTTAGGAGACTGGAAGTTTAAAAATGTACAAG
SE10	C3U-seq3598	ATAGCTGTACAAATATAAGAATAAAATGTTGAAA
SE11	C3U-seq9065	GGAGAAAGCTTCTCTATTTTGGATGCATTTCAGA
SE12	C3U-seq477	AAAGTTGCAAGATAAACAGCTGTAATTCGGACAA
SE13	C3U-seq10512	GTTTTAAGTAACTTTTTATAGCAAGATGATACAA
SE14	C3U-seq16091	TTTTGACTATTTTTATATATAAAGAAGAACTCAA
SE15	C3U-seq14491	TTATTGTGGATAACAAAGATATCTTTTCTTTAGA
SE16	C3U-seq13367	TGGCAGGTATTCCCATGATTCACAGAGTTACATT
SE17	C3U-seq248	AAACAAAAGCCTGGCTGAGTTGATGTTTTACATT
SE18	C3U-seq5093	CACAGTATTCGTGAATAAGTTGATTCTGTCCCCC
SE19	C3U-seq5184	CACTGAAGAGGTGGAAAAATAATCGTGTCAATCT
SE20	C3U-seq2244	ACTATAAATGCTTTGCAAAAATGGTTTCACGTTT
SE21	C3U-seq13645	TGTAGATCATAGGATAGCTGACTTTGACAGTCAC
SE22	C3U-seq1502	AATGGAACACAGACAGTGTAGAAGAATTCCTGAG
SE23	N/A	AGTAAACTGACGTTGTCCAACGTGCATATGGATT

2.13.4. 5'TOP Motifs

The criteria by which 5'TOP-containing genes were identified is described in detail in Chapter 5. Once genes were selected, 5'TOP sequences were identified by scanning for motifs in their CHO-K1 genome sequence (2014), with the following process:

1. Identify a sequence that satisfies the following criteria:
    a. Upstream of the start codon.
    b. Nearest to the start of the annotated gene.
    c. Begins with a Cytidine.
    d. Comprises an exclusive tract of 5+ pyrimidines.
2. If no such sequence can be identified, restart this process with the highest ranked gene of the next list.

5'TOP motifs were substituted with the natural sequence from the Ef1α 5'TOP sequence in both the Ef1α and modified CMV-TCT promoters, by site-directed mutagenesis. The motif was removed from both of these promoters as a NON-TOP control.

Source	Name	Sequence
EF1a	TOP1	CTTTTTC
NON-TOP	TOP2	
ACTB	TOP3	CTCTTTCTTC
RPLP2	TOP4	CTTCCTTTC
RPL11	TOP5	CCTCCT
RPS14	TOP6	CCTCCCCT
GAPDH	TOP7	CTCTCT
RPL13A	TOP8	CCTCCTCCTTTCCC
RPL14	TOP9	CCTTCTTCCTTCTC
RPS10	TOP10	CTTCTC
RPS25	TOP11	CTCCTCCC
RPS23	TOP12	CCTTCCT
RPS24	TOP13	CCTTCC
RPS7	TOP14	CCTCTTTCT
RPL31	TO*P15*	CTTCCCTTCCC
RPL18A	TOP16	CTTCCTTTT
RPS3A	TOP17	CTCCCC

RPS3A#2	TOP18	CCCTTTT
HSP90AB1	TOP19	CTTCTC
TUBB	TOP20	CCTTCCCTCCT
RPS27A	TOP21	CCTCTCTTCTC
VIM	TOP22	CCTCT
RPS8	TOP23	CCTTCCC
EEF1G	TOP24	CCTTTT
EEF2	TOP25	CTCTTCTCCT
HSPD1	TOP26	CCCTCCC
IPO5	TOP27	CTCCCTCCTCCTTCTCTCTCTC
RPL7	TOP28	CTTCCTCTCTCT
RPL10	TOP29	CTTTTCCTCC
EIF2S3	TOP30	CCTTCCTCTCT
HNRNPF	TOP31	CCTCTTCCTCCTC
RPL3	TOP32	CCTCT
RPL30	TOP33	CTTCCTTTCT
RPS2	TOP34	CCTTCCCC
RPL23A	TOP35	CCTTTT
EEF2	TOP36	CTCTTCTCC
ENO2	TOP37	CTTTCTCCTTCCTCC
DDX21	TOP38	CTTTCTTCCTCTCTCTTTT
IPO7	TOP39	CTTCTCTTTCCTTTC
GARNL3	TOP40	CCTTTTTTTTTTTTCTC
LAMB1	TOP41	CCCCTTCCT
DDX39B	TOP42	CTCTTCT
GNAI3	TOP43	CCCCTCTCCC
CS	TOP44	CCCTTCCT
MATR3	TOP45	CCTCCTT
EIF2S3	TOP46	CCTTCCTCTCT
EIF3A	TOP47	CTTTCC
KPNA3	TOP48	CTCTTT
-	sTOP	CTTTCT
-	TOP36-1	CTCTTCTC

2.13.5. Synthetic Proximal Promoters

Synthetic proximal promoters were taken from literature (Brown and James, 2015; Brown *et al.*, 2014a, 2017), and substituted the proximal promoters of both CMV and Ef1α, upstream of their cores, in the AstraZeneca in-house TU1 vector, using EcoRI/NheI and EcoRI/AflII restriction/ligation cloning for the CMV and Ef1α core promoters, respectively.

Name	Sequence
P-CMV	AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGC CCGCCTCGTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTA TGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAA GTCCGGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATG CCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTA GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAAT AACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT
100RPU	TGGGACTTTCCACCTTAGATGACACAGCAATCAGATTTGCTTGCGTGAGAAG ATATAGGATGACACAGCAATCTAGACTGGGACTTTCCACTGATATTTTGCGC AATTGACCTAATGACACAGCAATAGTATGTGGGGCGGGGATCTAACTGGGA CTTTCCAAAGGTCTTACCGGAAGTTGTTAGAATGACACAGCAATGGATTCAT ATCCTGGGACTTTCCAGTATACTGCTTGCGTGAGAAGATGATCATGGGACTTT CCATGTACAAAAGGTC
80RPU	TGGGGCGGGGAAGTATGATGACACAGCAATTGATCATGGGACTTTCCACTAG ACTGCTTGCGTGAGAAGAAAGGTCTTACCGGAAGTTGACCTAATGACACAG CAATGTTAGATGCTTGCGTGAGAAGACTGATATGGGACTTTCCAGTATACTG GGGCGGGGATCTAACTGGGACTTTCCACAGATTATGACACAGCAATTGTACA AAAGGT
20RPU	TTACCGGAAGTTGACCTATGCTTGCGTGAGAAGAGTTAGATGGGGCGGGGA AAGGTCTTTTGCGCAATTCAGATTTTACCGGAAGTTTATAGGATGACACAGC AATTGTACAAAAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATC GCCTAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGAC

10RPU	TTACCGGAAGTTAAGGTCTTTTGCGCAATTCAGATTGACCTATTACCGGAAGT TTATAGGTGGGGCGGGGAGTTAGATTTTGCGCAATTTGATCATTACCGGAAG TTTGTACAAAAGGT
5RPU	TATAGGAAGGTCTTACCGGAAGTTCCTTAGCTGATAGTATACCAGATTTTTTG CGCAATTCTAGACTGATCATCTAACGACCTATTACCGGAAGTTAGTATGTGT ACAAAAGGT
10xERSE	ACCAATGGCCAGCCTCCACGAAAGGTCACCAATGGCCAGCCTCCACGAAGT ATGACCAATGGCCAGCCTCCACGACAGATTACCAATGGCCAGCCTCCACGA CCTTAGACCAATGGCCAGCCTCCACGAGACCTAACCAATGGCCAGCCTCCA CGAGTATACACCAATGGCCAGCCTCCACGATCTAACACCAATGGCCAGCCTC CACGATGATCAACCAATGGCCAGCCTCCACGACTAGACACCAATGGCCAGC CTCCACGAAGGT
10xNFκB	TGGGACTTTCCAAAGGTCTGGGACTTTCCAAGTATGTGGGACTTTCCACAGAT TTGGGACTTTCCACCTTAGTGGGACTTTCCAGACCTATGGGACTTTCCAGTAT ACTGGGACTTTCCATCTAACTGGGACTTTCCATGATCATGGGACTTTCCACTA GACTGGGACTTTCCAAGGT

2.14. Graphing and Statistics

- Error bars displayed on all bar and line graphs are of the standard error of the mean.
- Box and whisker plots show the minimum and maximum values of the dataset.
- Statistical significance was measured by paired t.test in the case of datasets with directly comparable biological replicates, and unpaired t.test for datasets lacking them.

3. RNA-Binding Proteins

3.1. Introduction

3.1.1. Overcoming export deficiencies of intronless transcripts by tethering to mRNA binding proteins

Whilst minimisation of DNA vectors is desirable, removal of introns is well documented to decease gene expression (Callis *et al.,* 1987; Lu and Cullen, 2003; Rathus *et al.,* 1993). The extent of this effect is gene-dependent, ranging from a 35-fold to 2-fold decrease in expression upon intron removal, but is universal across intron-containing genes (Lu and Cullen, 2003). This effect occurs because mRNA recruitment into processing and export machinery in mammalian cells is splicing dependent (Heath *et al.,* 2016; Masuda *et al.,* 2005). The splicing-associated proteins in the exon junction complex (EJC), as well as *DDX39B* (Heath *et al.,* 2016; Viphakone *et al.,* 2019) and *Prp19* (Chanarat and Sträßer, 2013) are required for the stable deposition of the TREX complex on nascent mRNA. Since mRNA export is an extensively coupled process, for instance in the passing of mRNA from TREX to *NXF1* (Viphakone *et al.,* 2012), and from *NXF1* to the nuclear pore complex (Okamura *et al.,* 2014), lack of TREX recruitment via splicing has a knock on effect downstream, decreasing expression of genes with removed introns.

Natural intronless genes have evolved strategies to compensate for this expression deficit. A study into expression of several intronless mammalian genes found that a conserved portion of their coding region conferred TREX and *NXF1* dependent cytoplasmic accumulation and expression. These cytoplasmic accumulation regions (CARs) also conferred cytoplasmic localisation on an unrelated cDNA when inserted into its coding sequence (Lei *et al.,* 2011). In further studies, these CARs were found to associate with TREX and *Prp19* components. When a multimerised CAR was inserted downstream of an intronless *β-globin* reporter, it rescued expression. This rescue was reversed upon TREX knockdown, or mutation of the CAR sequence (Lei *et al.,* 2013). These data suggest that naturally intronless genes contain sequence specific elements that recruit them into the TREX processing pathway, compensating for a lack of co-splicing recruitment.

Another group of intronless proteins compensating for their expression deficit with this strategy can be found in viruses. The Hepatitis B Virus (HBV) core protein (*HBc*) has been shown to specifically interact with *NXF1* and with TREX components by immunoprecipitation. Furthermore, non-spliced HBV RNAs were found to decrease in cytoplasmic accumulation upon siRNA treatment of *NXF1* (Yang *et al.,* 2014). Taken together, these data suggest that HBV compensates for lack of expression in its intronless genes by 'hijacking' its way into the TREX/*NXF1* export pathway via the HBc protein. Further evidence of this strategy in HBV is found in a post-transcriptional RNA element, *SEP1*. The RNP that forms on this element in the nucleus contains TREX and other export factors. The element was found to bind *ZC3H18*, a TREX associated protein, and enhance cytoplasmic localisation in a manner dependent on both sequence, and *ZC3H18* (Chi *et al.,* 2014). This strategy can be found in a variety of viruses, with Herpesvirus Saimiri protein *ORF57* linking viral mRNA to the export adaptor *ALYREF* (Tunnicliffe *et al.,* 2014), and with the same protein in Kaposi's Sarcoma–Associated Herpesvirus, in which cytoplasmic localisation of viral mRNA is disrupted by *ORF57* mutation and by *ALYREF* or *NXF1* mutation (Boyne *et al.,* 2008). Export of Herpes Simplex Virus 1 mRNA is dependent upon the ICP27 protein, which binds viral mRNAs and *ALYREF* (Tian *et al.,* 2013). Finally, the Epstein-Barr virus early protein *EB2* has been shown to contain a nuclear export signal that binds both viral mRNAs and *NXF1*, and is required for cytoplasmic localisation of viral mRNA (Juillard *et al.,* 2009).

Together, these results show that a generalised strategy for efficient intronless RNA export is by hijacking, by either protein-protein or protein-RNA interactions, into mRNP processing and export pathways. Therefore, if an intronless expression cassette could be similarly hijacked into the export pathway, it could overcome the export deficit given by its' minimisation.

3.1.2. PP7/MS2 Tethering

PP7 and MS2 tethering are two analogous technologies for RNA-protein recruitment. Each system comprises of a bacteriophage coat protein, and an RNA hairpin loop, to which the coat protein binds in a sequence-specific manner (Chao *et al.,* 2008). In order to bind these RNA hairpins, the coat proteins must first dimerise. This is usually a rate limiting process, with the

K_d associated with coat protein dimerisation around an order of magnitude higher for dimerisation than RNA binding, as shown in Table 3.1 (Wu et al., 2012).

	PP7	MS2
Dimerisation K_D	20nM	410nM
RNA Binding K_D	1.6nM	5nM

Table 3.1 – The dissociation constants associated with each stage of PP7/MS2 RNA-protein tethering.

In order to circumvent this rate limiting step, translationally fused PP7 and MS2 coat protein dimers (PP7d, MS2d) have been created (Wu *et al.,* 2012), thus removing the rate-limiting dimerisation step of binding. This technology is powerful and versatile, for instance being used to measure cell-wide mRNA noise level and transcriptional activity of a single gene (Hocine *et al.,* 2012), or, by tweaking specificity via a construct requiring both PP7 and MS2 binding, to measure single mRNA molecules (Wu *et al.,* 2014).

It has also been shown that mRNA export and expression can be stimulated by tethering to proteins relating to this function. Tethering of RNA to a fusion protein containing an *NXF1*-associated nuclear export sequence led to increased nuclear export (Juillard *et al.,* 2009), and the same effect is replicated when the RNA is directly tethered to *NXF1* (Tintaru *et al.,* 2007). Tethering of *GFP* mRNA to an MS2 fusion with *HuR*, a 3' UTR binding protein, led to increased localisation of GFP at the cell membrane (Berkovits and Mayr, 2015).

Bringing these conclusions together, it was decided to emulate the generalised strategy for intronless mRNA expression described above. By association of intronless mRNAs with proteins involved in processing and export using a PP7 tethering system, a molecular biology tool could be created for efficient export of mRNA. Ideally, this tool would facilitate vector minimization by intron removal, whilst minimizing the export deficit associated with this change.

3.2. Results

3.2.1. Construction of PP7d-effector gene fusion Library

In order to tether an intronless reporter mRNA to RNA-binding export proteins, a library of 14 fusion proteins was created. These fusion proteins comprised an SV40 nuclear-localisation

sequence (NLS), a HA peptide tag, the PP7 dimer (PP7d) CDS, a short leucine-rich linker, and the gene of interest CDS, as detailed in section 2.2.1, and were created by modification of the pcDNA5 FRT TO vector by Gibson assembly as in section 2.3.3. Correct construction was verified by Sanger sequencing, and a representative plasmid map can be found under 'FRT TO – PP7d Fusion' in Appendix A, and a schematic shown in Figure 3.2. Thus, the resulting protein would be detectable by HA western blot, localised in the nucleus, and able to tether to an mRNA bearing PP7bs.

To examine whether the fusion proteins encoded by these plasmids were being properly expressed, $1*10^5$ HEK-293T cells in a 24-well plate format were transiently transfected by PEI at a 1:5 DNA:PEI ratio with 250ng of each construct, and their cell lysates collected 72 hours post-transfection for analysis by western blot with a HA-tag antibody, as described in section 2.7. As shown in Figure 3.1, testing of untransfected cell lysate led to detection of four off-target bands. Therefore, any bands appearing in addition to these four could be attributed to the transfected fusion protein. All the remaining lanes in Figure 3.1 are marked with the expected position of each of their fusion proteins, based on their predicted molecular weight. In all lanes an additional band is visible in the expected position. For the PP7d-*CIRP* and PP7d-*RBM3* (cold-inducible genes associated with increased mRNA stability (Tan et al., 2008; Xia et al., 2012; Zhu et al., 2016)) lanes, the predicted position of each protein was very close to that of the second-lowest off-target bar. However, in both lanes the additional protein is discernible both by the increased intensity of this band, and by the visibility of two overlapping peaks of intensity within the band. The blot in the PP7d-*NXF1* lane shows cells transfected with a PP7d-NXF1 fusion, with the addition of its' co-dimer P15, separated by a P2A self-cleaving peptide. If expressed correctly, the P15 will be cleaved from the fusion protein at translation, creating a stoichiometric increase in P15, to match NXF1 overexpression in the cell. As expected, the band in this lane is the predicted size of the PP7d-NXF1 fusion, accounting for P15 cleavage. The data shown in Figure 3.1 therefore suggests that all the fusion proteins created for these experiments were correctly expressed.

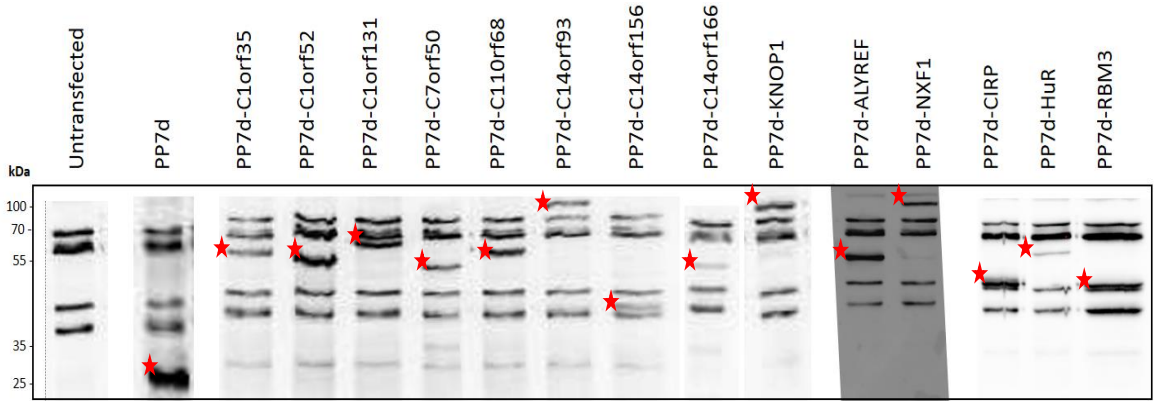


Figure 3.1 – Showing the testing for correct transient expression of a library of PP7d-fusion proteins, by western blot against a HA peptide tag after transfection by PEI at a 1:5 DNA:PEI ratio into HEK-293T cells. The off-target bands given by western blot of untransfected cell lysate are shown in lane 1. The expected position of each fusion protein is shown by a red star in each lane. Images taken at separate times were aligned in this figure, based on the molecular ladder present on each individual blot. Images were taken using automatically set exposure time, using 'Optimal' mode on the Chemidoc MP (Bio-Rad, California, USA) and 'Strong Bands' mode on the Chemidoc XRS (Bio-Rad, California, USA). No further brightness/contrast alterations were made to the images. Representative blots from two biological replicates are displayed.

3.2.2. Characterisation of SEAP Assay and Reporters

In order to accurately measure effects on SEAP expression, the assay to measure its' titre was characterised. Firstly, a linear relationship between SEAP titre and absorbance measured in the assay had to be demonstrated. Different volumes of supernatant taken 72h post-transfection from $1*10^5$ HEK-293T cells transiently transfected by PEI at a 1:5 DNA:PEI ratio with 250ng TU1-*SEAP* in 24-well plates were assayed by the in-house SEAP assay described in section 2.6.1. TU1-*SEAP* represents the in-house AstraZeneca proprietary TU vector (section 2.2.2.), with the *SEAP* CDS inserted by restriction/ligation cloning, as in section 2.3.6., and a representative plasmid map can be found under 'TU1' in Appendix A. Schematics of the TU1-SEAP-PP7bs reporter and effector gene PP7d fusions constructs used in this chapter are shown in Figure 3.2.

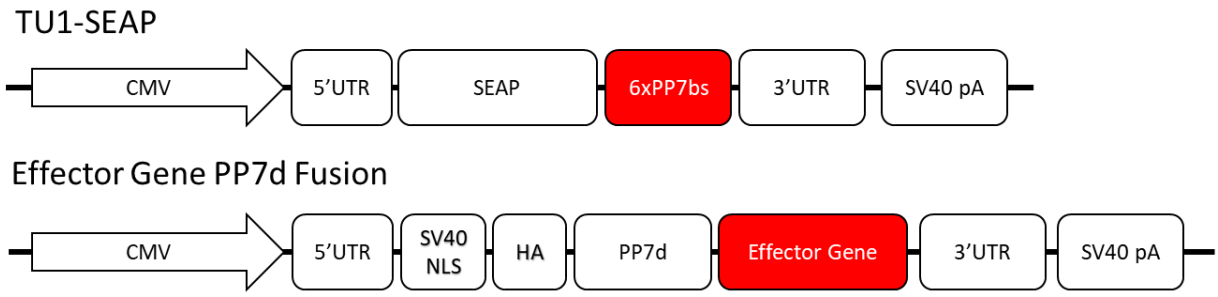


Figure 3.2 – A schematic of the TU1-SEAP-6xPP7bs reporter and effector gene PP7d fusion constructs which are used in this chapter.

A constant total volume of 200µl was maintained by addition of fresh, uncultured medium. The results of this experiment are shown in Figure 3.3a, with a strong linear relationship between absorbance and the proportion of SEAP-containing supernatant present in the assay. Next, $1*10^5$ HEK-293T cells in 24-well plates were transfected with increasing quantities of TU1-*SEAP* using PEI at a 1:5 DNA:PEI ratio, and their supernatants assayed for SEAP titre 72 hours post-transfection, by the in-house method in section 2.6.1. As shown in Figure 3.3b, a non-linear, second order exponential relationship was observed between SEAP titre and quantity of TU1-*SEAP* transfected. To further investigate this result, a constant quantity (20ng) of TU1-*SEAP* or FRT-TO-*GFP* (the pcDNA5 FRT TO vector expressing a GFP CDS) was cotransfected into $1*10^5$ HEK-293T cells in 24-well plates by PEI at a 1:5 DNA:PEI ratio, with increasing quantities of FRT-TO-*GFP*, and SEAP titre assayed 72 hours post-transfection, by the in-house method in section 2.6.1. As shown in Figure 3.3c, cells subject to cotransfection of increasing quantities of FRT-TO-*GFP* alongside TU1-*SEAP* displayed a linear increase in SEAP titre, with transfection of only GFP leading to no SEAP titre being measured. In combination with Figure 3.3a and Figure 3.3b, these data suggest that whilst total concentration of SEAP present is correlated linearly with assay absorbance, SEAP titre is stimulated by increased total DNA transfected, even if this DNA produces no SEAP assay signal when transfected alone. Therefore, all transfections should be performed with constant DNA quantity, with 'filler' DNA being used to make up the total weight if necessary. Finally, SEAP-containing supernatant harvested for the previous assay was measured at several timepoints after the start-point of the assay. Figure 3.3d shows a linear relationship between absorbance, and the timepoint at which it was measured. Therefore, supernatants must be

measured at the same timepoint post-assay start, for their fold-change in SEAP titre to be accurately comparable.

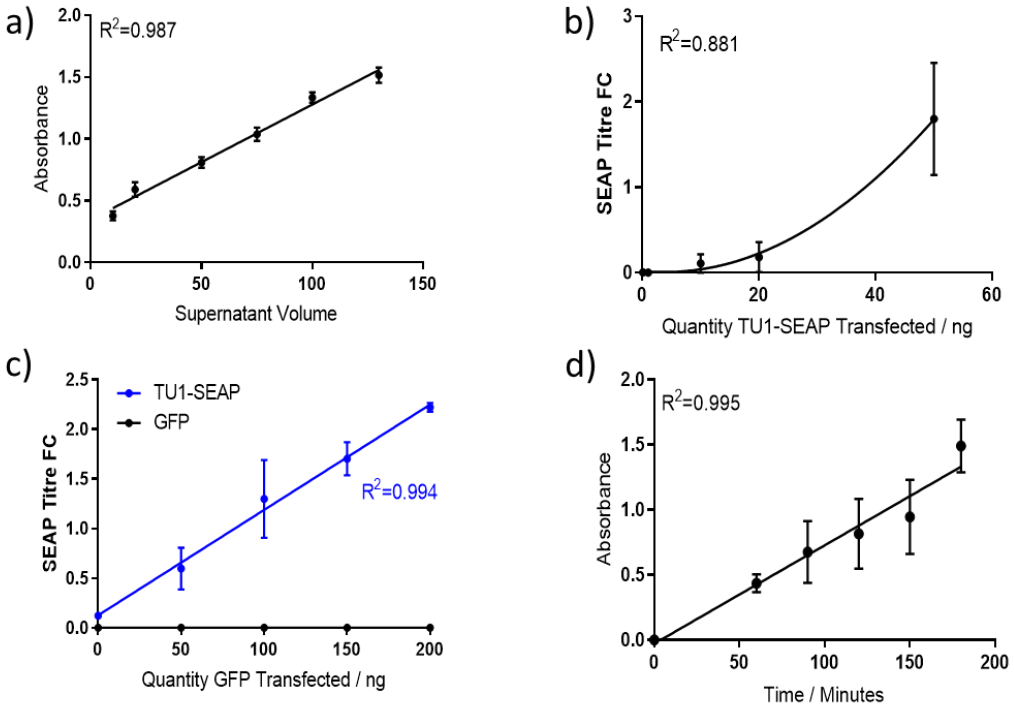


Figure 3.3 – SEAP assay characterisation using supernatant from HEK-293T cells transfected by PEI at a 1:5 DNA:PEI ratio with TU1-SEAP. A) Increasing quantities of SEAP-containing supernatant, made up to 200µl with fresh medium, were tested for SEAP titre, with a linear correlation observed between concentration of SEAP present and absorbance. B) Increasing quantities of TU1-SEAP were used to transfect HEK-293T cells, and supernatant tested 72h post-transfection showed an exponential relationship between quantity of SEAP transfected, and titre. C) A constant quantity of 20ng TU1-SEAP or FRT-TO-GFP were cotransfected with increasing quantities of FRT-TO-GFP. With TU1-SEAP present, a linear correlation between total transfected DNA quantity and SEAP titre was observed. With only FRT-TO-GFP present, no SEAP titre was detected. D) SEAP containing supernatant was assayed for SEAP titre, with absorbance taken at several timepoints. A linear correlation was observed between absorbance measured and time-post assay start. Results presented are of technical duplicates, from three biological replicates. FC: fold-change.

To tether the mRNA transcript of the *SEAP* reporter to the library of PP7d-fusion proteins, a TU1-*SEAP*-6xPP7bs reporter was created. Six PP7 stem-loops were inserted downstream of the SEAP 3'UTR, as described in section 2.13.1. and Table , by Gibson assembly, as in section 2.3.3., and correct construction verified by Sanger sequencing.

Name	**Sequence**
2xMS2bs	TCGAGGAATTCCGATCCTCTAGAGTCGACCTGCAGACATGGGTG ATCCTCATGTTTTCTAGAGTCGACCTGCAGACATGGGTGATCCTC ATGTTTTCTAGAGTCGACCTGCAGGCATGCAAGCTGCCCGGGGG ATCCACTAGTTCTAGCCGGAATTCCTCGA
2xPP7bs	GATCCGTAGAAATAAGGAGTTTATATGGAAACCCTTACTGCTGG GAGATCCGTAGAAATAAGGAGTTTATATGGAAACCCTTACTGCT GGGA

Table 3.2 – Sequences of the 2xPP7bs and 2xMS2bs stem loops. Coat protein binding sites sequence are well established in literature (Chao et al., 2008, 2012). Modules, containing 3 repeats of sequences of 2 stem-loops (totalling 6 binding sites), were isolated by PCR from an in-house laboratory plasmid, and placed downstream of the 3'UTR, directly upstream of the PolyA signal in the AstraZeneca in-house TU1 expression vector.

In order to test its' expression, 250ng TU1-*SEAP*-6xPP7bs was transfected by PEI at a 1:5 DNA:PEI ratio into $1*10^5$ HEK-293T cells in 24-well plates, and supernatants harvested 72h post-transfection for SEAP assay as in section 2.6.1., alongside an unmodified TU1-*SEAP* control. The fold change of titre from unmodified TU1-*SEAP* to TU1-*SEAP*-6xPP7bs was calculated, and is presented in Figure 3.4. The modified reporter retained moderately high SEAP expression, though titre was decreased by 24% compared to the unmodified control.

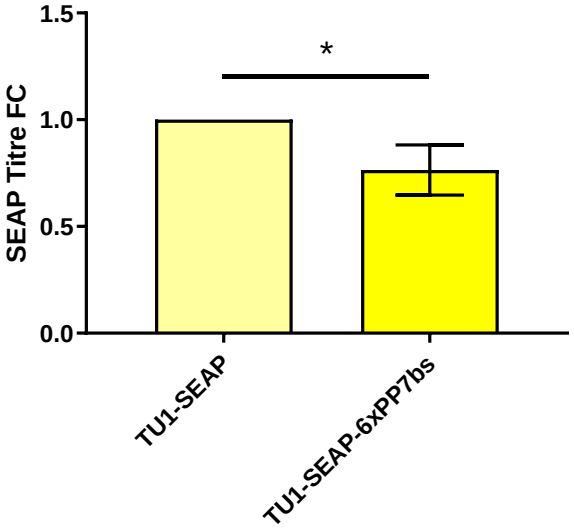


Figure 3.4 - Expression testing of TU1-SEAP-6xPP7bs reporter. SEAP titre from the modified reporter was 0.76-fold of the unmodified TU1-SEAP, in a statistically significant decrease (P=0.036). The sequence of 6xPP7bs can be found in section 2.13.1. and Table of this thesis. Results presented of technical duplicates, from three biological replicates. FC: fold-change.

3.2.3. Screening of SEAP-6xPP7bs with Overexpression and Tethering of the PP7d Fusion Library

The first round of screening intronless SEAP reporters tethered to RNA-binding proteins was performed using characterised proteins, known to influence RNA maturation, export, and translation – *NXF1-P15*, *ALYREF*, *CIRP*, *RBM3*, and *HuR*. In order to analyse the effect of tethering each of these proteins to a SEAP mRNA, 200ng of each fusion protein and 20ng of TU1-*SEAP* both without and with 6xPP7bs, measuring the effect of overexpression (e.g. P-*ALYREF* O) and tethering (e.g. P-*ALYREF* T) of each fusion protein, were transfected by PEI at a 1:5 DNA:PEI ratio into $1*10^5$ HEK-293T cells in 24-well plates. Supernatant was harvested 72h post-transfection, and assayed for SEAP titre as in section 2.6.1., relative to cotransfection of TU1-*SEAP*+/-6xPP7bs and FRT-TO-*GFP*. An additional control of TU1-*SEAP*+/-6xPP7bs cotransfected with the PP7d protein alone was also included. The resultant SEAP titre fold changes are shown in Figure 3.5. Both tethering and overexpression of *ALYREF*, *RBM3*, and *HuR* were associated with non-significant changes in titre. Overexpression of PP7d alone slightly increases SEAP titre, with tethering giving a significant 1.52-fold increase. PP7d-CIRP overexpression is associated with a significant 1.64-fold increase in titre, although this is not significantly different than with PP7d overexpression. Finally, both overexpression and tethering of NXF1-P15 we associated with a significant decrease in SEAP titre, of 0.52 and 0.48-fold respectively.

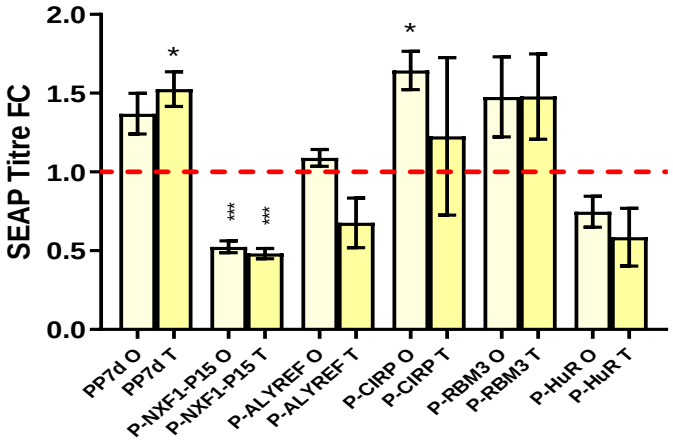


Figure 3.5 – SEAP titre analysis of characterised PP7d-RNA binding protein fusions, using both overexpression (O) and tethering (T), transfected by PEI at a 1:5 DNA:PEI ratio into HEK-293T cells. Fold change in titre is relative to TU1-SEAP cotransfection with FRT TO-GFP, shown as a red dotted line. Significant titre increases are observed with PP7d tethering and PP7d-CIRP overexpression, whereas significant titre decreases are given by both overexpression and tethering of PP7d-NXF1-P15. Results presented are of technical duplicates, from three biological replicates. FC: fold-change.

In the second round of screening, a panel of PP7d fusions with uncharacterised proteins shown to bind RNA (Castello *et al.*, 2012) was similarly created, and tested for effect on SEAP titre using the same method as the first. As shown in Figure 3.6, neither overexpression nor tethering of PP7d-*C1orf52*, PP7d-*C1orf131*, PP7d-*C11orf68*, PP7d-*C7orf50*, PP7d-*C14orf156*, or PP7d-*C14orf166* were associated with a significant change in SEAP titre, compared to the FRT-TO-*GFP* control. However, both overexpression and tethering of PP7d-*C1orf35* led to significant increases in titre, of 3.35-fold and 2.58-fold respectively. Tethering, but not overexpression of PP7d-*C14orf93* led to a significant increase of 2.57-fold in titre. In contrast, overexpression, but not tethering of PP7d-*KNOP1* led to a significant increase of 1.4-fold in titre. This increase, however, is smaller in magnitude than that given by PP7d alone, as shown in Figure 3.5.

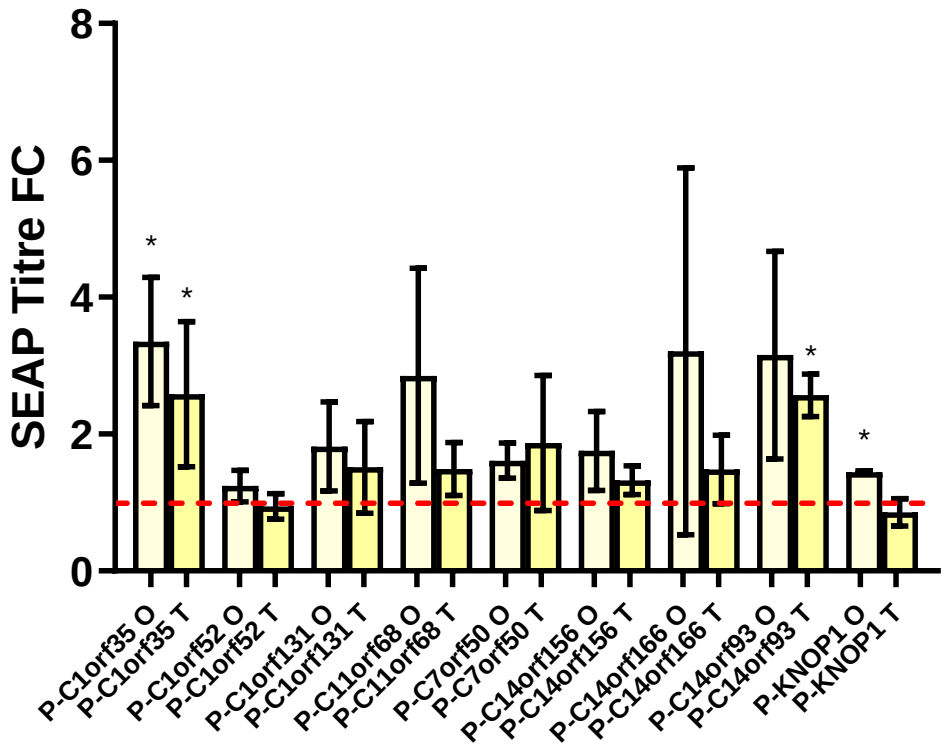


Figure 3.6 - SEAP titre screen of uncharacterised PP7d-RNA binding protein fusions, using both overexpression (O) and tethering (T), transfected by PEI at a 1:5 DNA:PEI ratio into HEK-293T cells. Fold change in titre is relative to TU1-SEAP cotransfection with FRT TO-GFP, shown as a red dotted line. Significant titre increases are given by PP7d-C1orf35 overexpression and tethering, PP7d-C14orf93 tethering, and PP7d-KNOP1 tethering. Results presented are of technical duplicates, from three biological replicates. FC: fold-change.

3.2.4. *Overexpression of PP7d-C1orf35 is associated with increased growth and transient SEAP titre*

The next aim was to replicate and expand upon identification of PP7d-*C1orf35* as a potential overexpression target to increase transient SEAP titre. Therefore, 200ng PP7d-*C1orf35* or PP7d were cotransfected with 20ng of both intronless and intron-containing TU1-*SEAP* by PEI at a 1:5 DNA:PEI ratio, both in absence and presence of PP7 binding sites, to demonstrate the effects of overexpression and tethering. These transfections were performed on $1*10^5$ HEK-293T cells in 24 well-plates, and SEAP titres measured 72h post-transfection, as in section 2.6.1. As shown in Figure 3.7a, the ability of PP7d-*C1orf35* overexpression to stimulate intronless SEAP titre was replicated, producing a 3.30-fold increase compared to an FRT-TO-*GFP* control, statistically significantly greater than the 1.37-fold increase associated with PP7d overexpression. The ability of PP7d-C1orf35 tethering to enhance intronless SEAP titre was observed, but the change non-significant, producing a 2.47-fold increase compared to the FRT-TO-*GFP* control, compared to the 1.54-fold increase associated with PP7d tethering. However, when tethered to intron-containing SEAP, as shown in Figure 3.7b, PP7d-*C1orf35* overexpression was associated with a significant decrease in SEAP titre, of 0.52-fold relative to the FRT-TO-*GFP* control, compared to a 2.64-fold increase in titre associated with PP7d tethering. When overexpressed, PP7d-C1orf35 non-significantly reduces intron-containing SEAP titre, to 1.67-fold, compared to the 2.33-fold associated with PP7d overexpression alone. Taken together, these data replicate the identification of PP7d-*C1orf35* overexpression as stimulatory to transient intronless SEAP titre, and may imply that presence of *C1orf35* at an mRNA influences splicing.

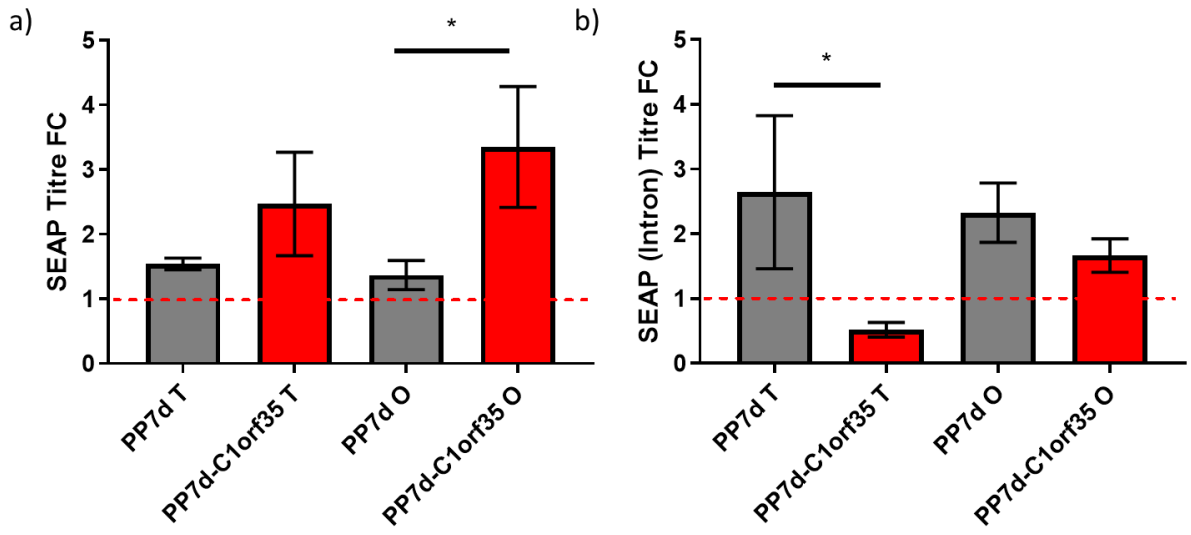


Figure 3.7 – Testing effect on SEAP titre of PP7d-C1orf35 after cotransfection by PEI at a 1:5 DNA:PEI ratio into HEK-293T cells. A) PP7d-C1orf35 tethering non-significantly increases TU1-SEAP titre compared to PP7d tethering, at 2.47-fold and 1.54-fold respectively (P=0.118). PP7d-C1orf35 overexpression increases intronless TU1-SEAP titre by 3.30-fold, relative to a control overexpressing FRT TO-GFP, significantly higher than PP7d alone, which increases titre by 1.37-fold (P=0.023). B) PP7d-C1orf35 tethering significantly decreases intron-containing SEAP titre by 0.52-fold, compared to a 2.64-fold increase conferred by PP7d tethering alone (P=0.036). PP7d-C1orf35 overexpression non-significantly decreases intron-containing SEAP titre by 1.67-fold, compared to a 2.33-fold increase associated with PP7d overexpression alone (P=0.094). Results presented are of technical duplicates, from three biological replicates. O: overexpression; T: tethering; FC: fold-change.

To further investigate and replicate the stimulatory properties of PP7d-*C1orf35* in a more industrially relevant CHO cell line, 2µg TU1-*SEAP* was cotransfected with 10µg of either PP7d-*C1orf35* or FRT-TO-*GFP* into a CHO transient host cell line by PEI at a 1:5 DNA:PEI ratio, using 10ml culture at 1*10^{6}cells/ml in cultiflasks. Cell growth was monitored every 24h, and SEAP titre measured 72h post-transfection as in section 2.6.1. Figure 3.8a shows that PP7d-*C1orf35* overexpression was associated with a 1.39-fold increase in SEAP titre compared to the FRT-TO-*GFP* control. Figure 3.8b implies cell growth as the cause of this increase, with VCD of PP7d-*C1orf35* overexpressing cultures growing to 1.38-fold higher than control after 72h. Combining these data reveals that no significant change in SEAP specific productivity occurs between PP7d-*C1orf35* overexpression and FRT-TO-*GFP* controls, as displayed in Figure 3.8c. Therefore, the stimulatory effect of PP7d-*C1orf35* on transient SEAP titre may be mediated through improved cell culture performance.

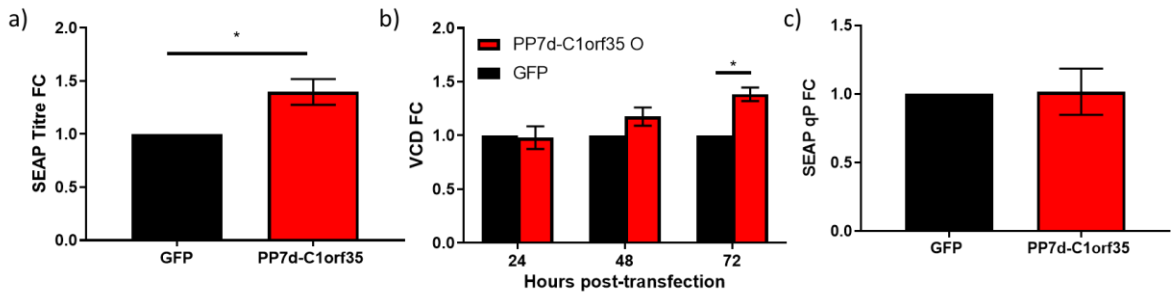


Figure 3.8 - PP7d-C1orf35 overexpression increases SEAP titre in CHO transient host cells when transfected with PEI at a 1:5 DNA:PEI ratio, through cell culture performance. A) Overexpression of PP7d-C1orf35 significantly increases transient SEAP titre by 1.39-fold, compared to an FRT-TO-GFP control (P=0.041). B) This is accompanied by a significant increase in cell growth, fold change in viable cell density compared to the FRT-TO-GFP control increasing from 0.98 at 24h, 1.18 at 48h, and 1.38-fold 72h post-transfection (P=0.019). C) There is no significant difference in SEAP specific productivity upon PP7d-C1orf35 cotransfection. Results presented are of technical duplicates, from three biological replicates. O: overexpression; FC: fold-change.

3.2.5. Overexpression of FLAG-C1orf35 in HEK-293T cells is associated with an increase in cell growth

Since PP7d-*C1orf35*'s stimulatory effect on SEAP titre occurred without being tethered, it was decided to investigate whether overexpression of a non-fusion *C1orf35* protein could reproduce this effect. Therefore, an expression vector bearing a FLAG-tagged *C1orf35* was created, by insertion of *C1orf35* into the p3X-FLAG expression vector, using Gibson assembly as in section 2.3.3., and correct construction checked by Sanger sequencing. A representative plasmid map can be found under 'p3X-FLAG-*C1orf35*' in Appendix A. 2µg of this FLAG-*C1orf35* vector was cotransfected with 200ng TU1-*SEAP* by PEI at a 1:5 DNA:PEI ratio into $1*10^6$ HEK-293T cells in 6cm dishes, and SEAP titre relative to a FLAG cotransfection control measured 72h post-transfection, as in section 2.6.1. Figure 3.9a displays that SEAP titre was increased 2.19-fold by FLAG-*C1orf35* overexpression, compared to the FLAG control. Cells from these transfections were also analysed for *SEAP* mRNA abundance by RT-qPCR, as in section 2.9.1. No significant difference in *SEAP* mRNA levels were detected between FLAG-*C1orf35* and FLAG coexpression, as displayed in Figure 3.9b. Finally, the protein concentration in the cell lysate 72h post-transfection was measured by Bradford assay, as a proxy measurement for cell growth, as in section 2.4.4.1. The data from this experiment are presented in Figure 3.9c, exhibiting a 1.47-fold increase in protein concentration upon FLAG-*C1orf35* cotransfection, compared to a FLAG control. These data corroborate earlier findings, presented in Figure 3.9, that *C1orf35* increases transient SEAP titre independently of tethering, and that this effect is mediated through cell culture performance, rather than specific productivity.

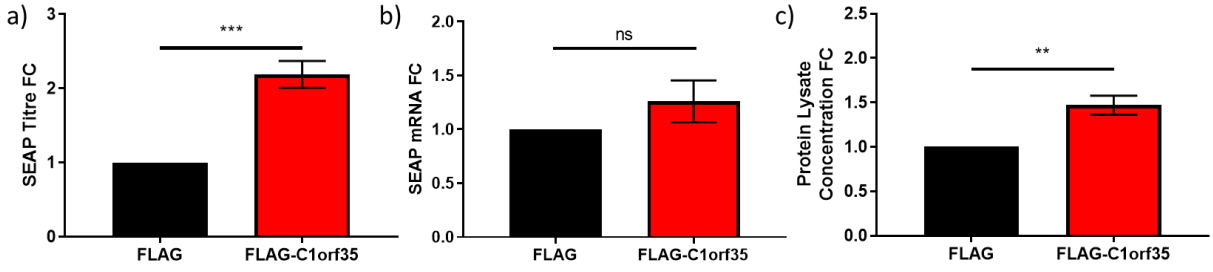


Figure 3.9 – FLAG-C1orf35 overexpression increases SEAP titre in HEK-293T cells when transfected with PEI at a 1:5 DNA:PEI ratio through cell culture performance. A) Overexpression of FLAG-C1orf35 significantly increases SEAP titre by 2.19-fold, compared to a FLAG-bearing control (P=0.0003). B) SEAP mRNA is not significantly increased in abundance by FLAG-C1orf35 overexpression. C) Protein lysate concentration 72h post-

transfection is significantly increased by 1.47-fold by FLAG-C1orf35 overexpression, compared to a FLAG-bearing control (P=0.004). Results presented are of technical duplicates, from three biological replicates. FC: fold-change.

3.2.6. Overexpression of C1orf35 is associated with mixed effects on titre in CHO cells

To ascertain whether *C1orf35* overexpression could also stimulate the cell culture performance and titre of a stable producing CHO cell line, *C1orf35* was codon optimised for CHO cell expression, and inserted into the in-house CHO expression vector (TU-*C1orf35*) by restriction/ligation cloning between the AgeI and SbfI sites as in section 2.3.6., with correct construction verified by Sanger sequencing. A representative plasmid map can be found under 'TU1' in Appendix A. CHO cells stably expressing an ETE mAb, were transfected with this vector against a TU1-*GFP* control. To examine whether any effect was dose-responsive, or whether there was an optimal dosage for titre, TU-*C1orf35* was titrated against TU1-*GFP* at five strengths (0%, 25%, 50%, 75%, 100%). A constant total load of 10μg DNA was transfected using PEI at a 1:5 DNA:PEI ratio, into 10ml CHO cell culture at $1*10^6$ cells/ml in cultiflasks. Growth was monitored at 24h, 48h, and 72h by ViCell as in section 2.4., and IgG titre was measured 72h post-transfection as in section 2.10.1. As shown in Figure 3.10a, some significant effects on growth were observed, with 50% and 75% dosage resulting in a temporary reduction in VCD at 24h (0.90 and 0.85-fold respectively), and 25% dosage leading to a 1.14-fold increase in VCD at 72h. However, no significant effects were measured across all titrations on viability 24h, 48h, or 72h post-transfection, or on titre 72h post-transfection, as displayed in Figure 3.10b and Figure 3.10c.

To investigate whether *C1orf35* was being correctly expressed by this cell line and vector, a FLAG-tag was added to the 5′ terminus of the *C1orf35* CDS by site-directed mutagenesis as in section 2.3.2., creating the TU1-FLAG-*C1orf35* vector, with correct construction verified by Sanger sequencing. 10μg of this vector was transiently transfected by PEI at a 1:5 DNA:PEI ratio into stable IgG-expressing CHO cells in 10ml culture at $1*10^6$ cells/ml in cultiflasks, and cell lysates analysed 72h post-transfection by western blot using an anti-FLAG primary antibody. This western blot is displayed in Figure 3.11. A band is clearly visible in lane 2, representing cells transiently transfected with TU1-FLAG-*C1orf35* at the predicted molecular

weight of the FLAG-*C1orf35* protein, 32kDa. No such band is found in the corresponding lane representing untransfected cells.

Collectively, Figure 3.10 and Figure 3.11 imply that whilst *C1orf35* is properly expressed in a stable IgG-producing CHO cell line, it is unable to significantly affect cell culture performance, or IgG titre.

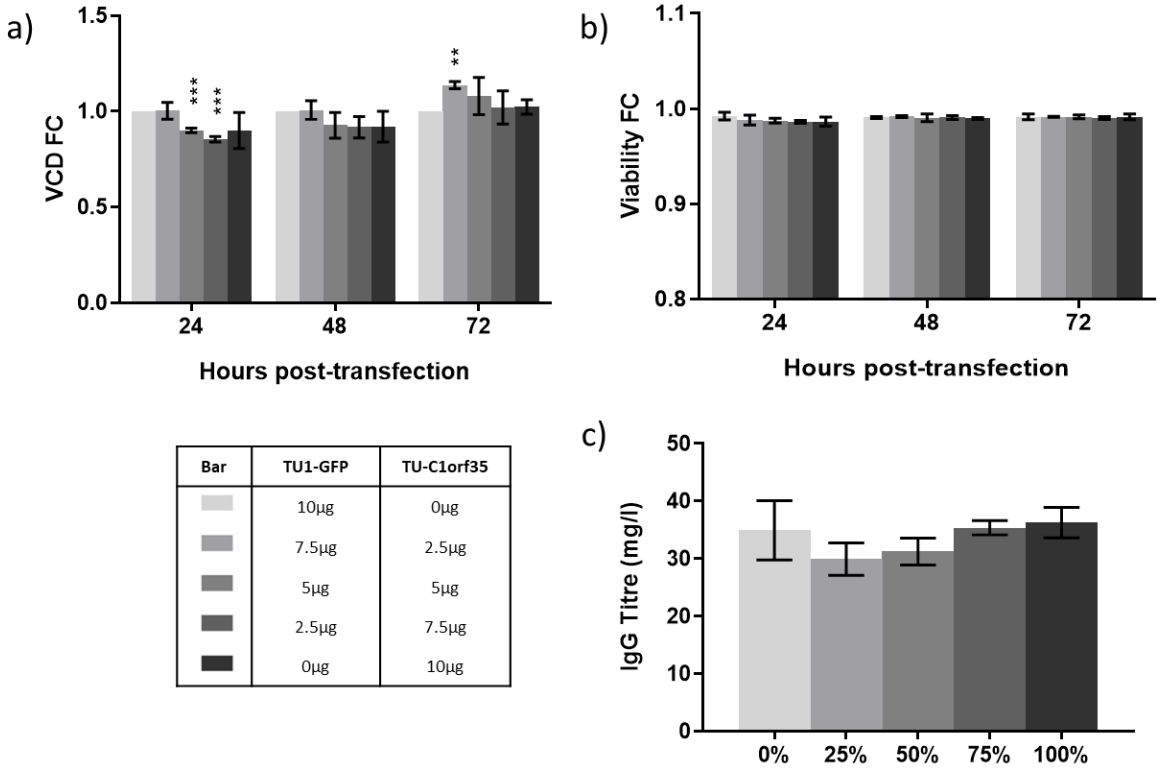


Figure 3.10 - TU-C1orf35 overexpression titration by PEI at a 1:5 DNA:PEI ratio has no significant effect on titre of a stable IgG expressing CHO cell line. A) 50% and 75% titration of TU-C1orf35 led to a temporary reduction of VCD 24h post-transfection of 0.90-fold (P=0.0009) and 0.85-fold (P=0.0005), respectively. 25% titration led to a slight increase in VCD 72h post-transfection of 1.14-fold (P=0.0018). B) No significant changes in cell viability at 24, 48, or 72 hours post expression were measured. C) No significant changes in IgG titres were measured 72h post-transfection. Results presented are of technical duplicates, from three biological replicates. VCD: viable cell density, FC: fold-change.

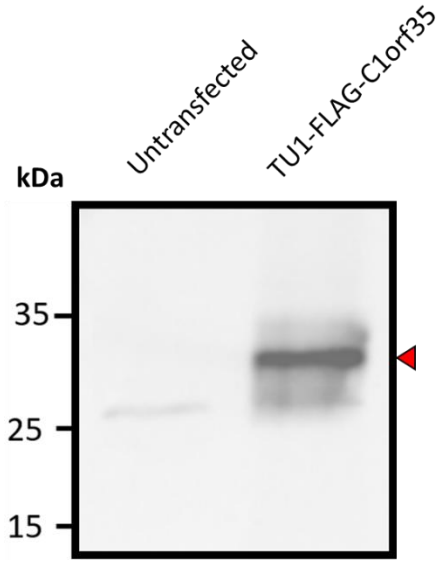


Figure 3.11 - FLAG-C1of35 is correctly expressed by the CHO transient host when transfected by PEI at a 1:5 DNA:PEI ratio. A western blot using an anti-FLAG primary antibody shows a band corresponding to the expected molecular weight of FLAG-C1orf35 (32kDa) for cells transfected with TU1-FLAG-C1orf35, and no band for untransfected cells. A representative blot from two biological replicates is presented here. kDa: kilodalton.

Next, it was examined whether the lack of significant effects imparted by TU-*C1orf35* seen in Figure 3.10 were due to the product being a non-fusion, as opposed to a PP7d-fusion protein. Similarly to the experiment presented in Figure 3.10, PP7d-*C1orf35* transient transfection was titrated against FRT-TO-*GFP* with PEI at a 1:5 DNA:PEI ratio into the stable IgG-producing CHO cell line, 10ml culture at $1*10^6$ cells/ml in cultiflasks. Growth was measured at 24h, 48h, and 72h post-transfection by ViCell as in section 2.4., and IgG titre measured 72h post-transfection by ELISA as in 2.10.1. The results of this experiment are displayed in Figure 3.12. Similarly to TU-*C1orf35* in Figure 3.10, PP7d-*C1orf35* overexpression titration led to some temporary effects on cell growth, with a 25% dosage leading to a 1.19-fold increase in VCD at 24h post-transfection. Also similarly, no significant changes in viability or titre were observed across the titration. These results suggest that neither *C1orf35* or PP7d-*C1orf35* confer a significant effect on cell culture performance or titre in a stable IgG-expressing CHO cell line.

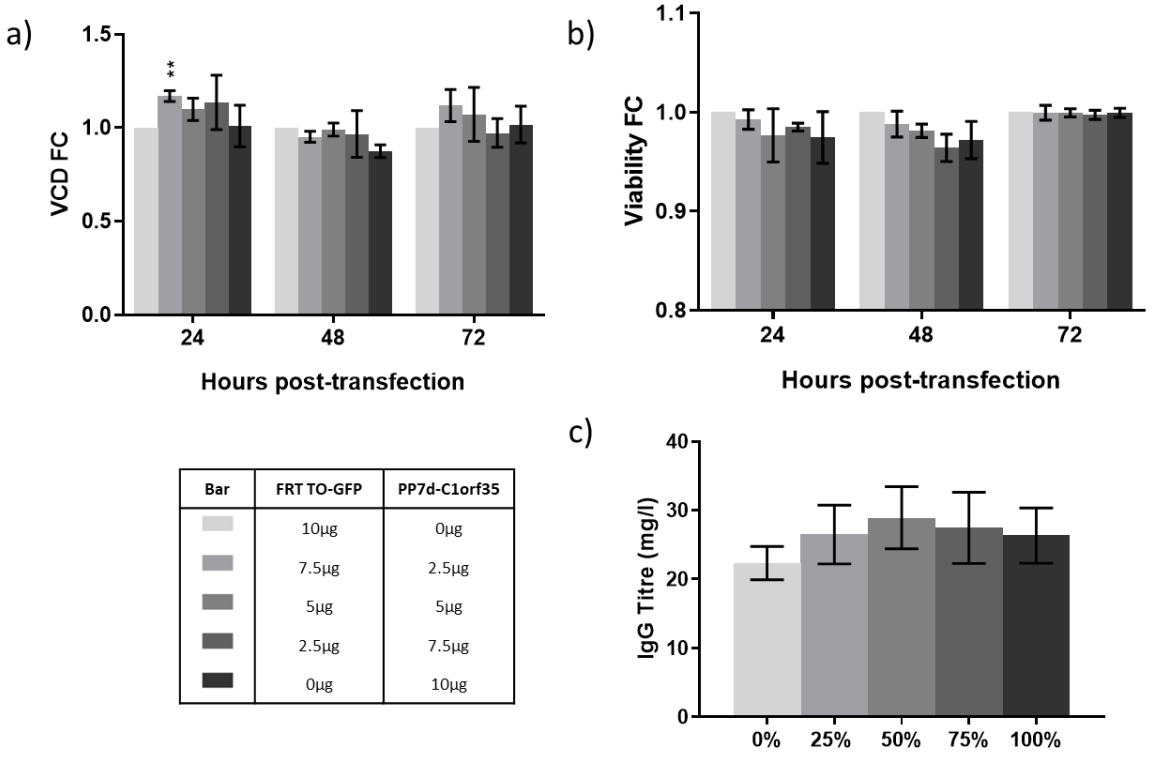


Figure 3.12 - PP7d-C1orf35 overexpression titration by PEI transfection at a 1:5 DNA:PEI ratio has no significant effect on titre of a stable IgG expressing cell-line. A) 25% dosage of PP7d-C1orf35 led to a temporary increase of VCD at 24h by 1.17-fold (P=0.0043). B) No significant changes in cell viability at 24, 48, or 72 hours post expression were measured. C) No significant changes in IgG titres were measured 72h post-transfection. Results presented are of technical duplicates, from three biological replicates. VCD: viable cell density; FC: fold-change.

To investigate whether *C1orf35* overexpression could stimulate cell culture performance and titre of a more industrially relevant transiently-expressed product, a constant load of a DTE mAb expression vector was cotransfected by PEI at a 1:5 DNA:PEI ratio into a transient CHO cell line alongside TU-*C1orf35*, titrated against TU1-*GFP*. As in previous experiments, 10ml cell culture at $1*10^6$ cell/ml in cultiflasks was used, growth measured at 24h, 48h, and 72h post-transfection by ViCell as in section 2.4., and IgG titre measured 72h post-transfection by ELISA as in section 2.10.1. As shown in Figure 3.13a and Figure 3.13b, no significant effect on cell growth or viability was seen with any titration of *C1orf35*. However, as displayed in Figure 3.13c, titration of TU-*C1orf35* was associated with a decline in IgG titre, from 1.00mg/l at 0% to 0.35mg/l at 100%.

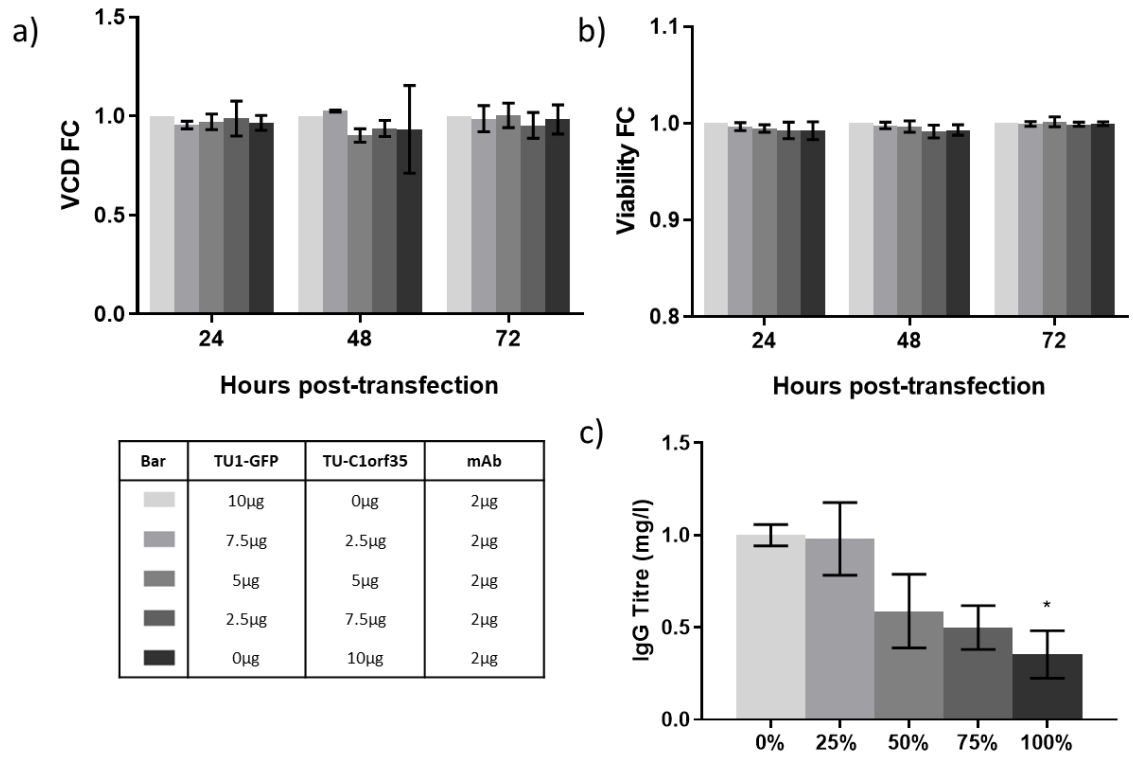


Figure 3.13 – TU-C1orf35 overexpression titration by PEI transfection at a 1:5 DNA:PEI ratio decreases transient titre of a DTE mAb. A) No overexpression titrations of TU-C1orf35 led to a significant change in VCD 24, 48, or 72h post-transfection. B) No overexpression titrations of TU1-C1orf35 led to a significant change in cell viability 24, 48, or 72h post-transfection. C) IgG titre of the DTE mAb decreases across the overexpression titration, with a significant difference of 1.00mg/l at 0% to 0.35mg/l at 100% (P=0.011). Results presented are of technical duplicates, from three biological replicates. VCD: viable cell density; FC: fold-change.

The results presented in this section indicate that the effect of *C1orf35* overexpression on cell culture recombinant protein production may be unpredictable, changing from positive to negative to non-significant, in a manner dependent on cell line, expression method, and protein product.

3.2.7. *Overexpression of PP7d-HuR in CHO cells is associated with increased specific productivity*

To assess the replicability of the assays described in section 3.2.3. between HEK-293T and CHO cells, the transient overexpression of four PP7d-fusion proteins (PP7d-*ALYREF*/*CIRP*/*HuR*/*C1orf35*) alongside TU1-*SEAP* was repeated in a CHO transient host cell line. To this end, 10ml of cells in cultiflask shaking culture were transfected by electroporation with 10μg PP7d-fusion protein and 2μg TU1-*SEAP*, and seeded at $0.2*10^6$ cells/ml. SEAP titre fold change was measured relative a control cotransfected with FRT-TO-*GFP*, 72h post-transfection, as in section 2.6.1. Figure 3.14 shows the SEAP titre fold-change associated with the overexpression of each fusion protein, with results from each cell line plotted against one another, taking the HEK-293T data from Figure and Figure 3.6. No significant correlation can be observed between titre fold changes in the two cell lines. PP7d-*C1orf35* retains a positive effect on SEAP expression in both cell lines, but this effect decreases from 3.35-fold in HEK-293T to 1.65-fold in CHO, whereas PP7d-*HuR* overexpression decreases titre to 0.75-fold in HEK-293T cells, but increases it by 4.85-fold in CHO cells.

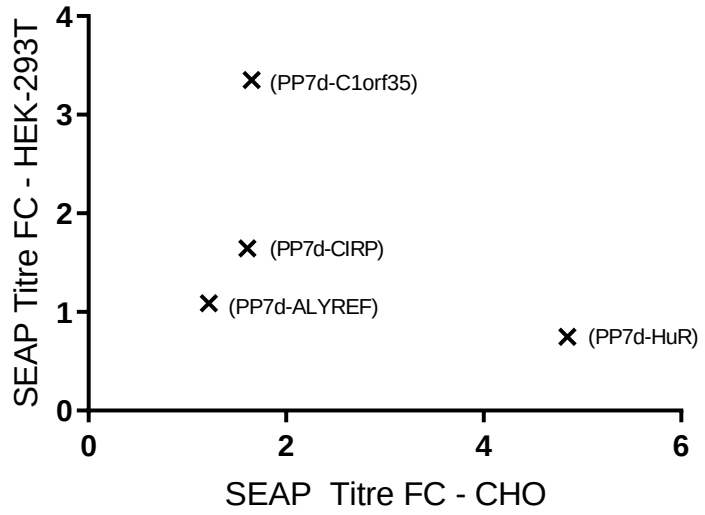


Figure 3.14 - Comparison of SEAP titre fold changes upon PP7d fusion protein coexpression following PEI transfection at a 1:5 DNA:PEI ratio between HEK-293T and CHO transient host cell lines. No significant correlation is observed in SEAP titre between the two cell lines. Results presented are of technical duplicates, from three biological replicates. FC: fold-change.

In order to validate and characterise the ability of PP7d-*HuR* overexpression to increase transient SEAP titre in CHO cells, the CHO transient host was transiently transfected using PEI at a 1:5 DNA:PEI ratio in 10ml shaking cultiflask culture at $1*10^6$ cells/ml with 10µg PP7d-*HuR* and 2µg TU1-*SEAP*. Growth was measured 24h, 48h, and 72h post-transfection by ViCell as in section 2.4., and SEAP titre measured 72h post-transfection, relative to a control cotransfected with FRT-TO-*GFP*, as in section 2.6.1. As shown in Figure 3.15a, cotransfection of PP7d-*HuR* was associated with a significant 2.95-fold increase in SEAP titre. This was achieved despite decreasing cell growth, as shown in Figure 3.15b, with viable cell density decreasing from 1.12-fold after 24h, to 0.67-fold after 48h, to a significant 0.58-fold decrease in VCD 72h-post transfection. Together, these results show that PP7d-*HuR* overexpression imparts a 4.78-fold increase in SEAP cell-specific productivity, as shown in Figure 3.15c.

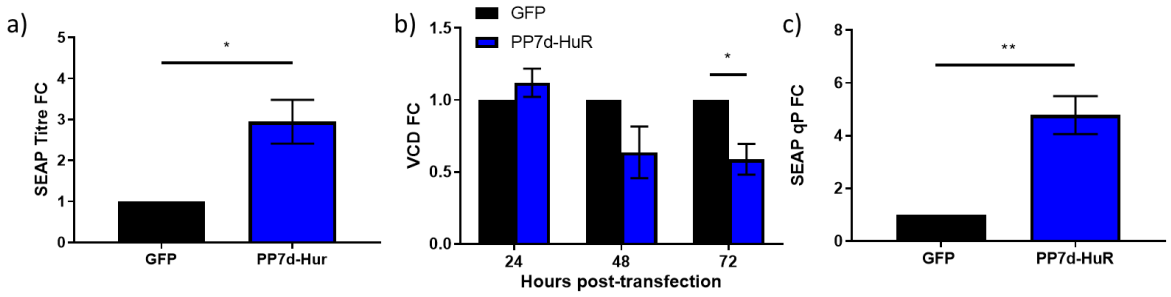


Figure 3.15 – PP7d-HuR overexpression by PEI transfection at a 1:5 DNA:PEI ratio increases transient SEAP titre in the CHO transient host through specific productivity. A) Overexpression of PP7d-HuR significantly increases transient SEAP titre by 2.95-fold, compared to an FRT-TO-GFP control (P=0.033). B) Cell division significantly decreases upon PP7d-HuR overexpression, fold change in VCD decreasing from 1.12 at 24h, 0.64 at 48h, and 0.58-fold 72h post-transfection (P=0.018). C) These data imply a significant 4.78-fold increase in SEAP specific productivity(P=0.0063). Results presented are of technical duplicates, from three biological replicates. FC: fold-change.

To further investigate the cause of the difference in SEAP titre brought about by PP7d-*HuR* overexpression in HEK-293T and CHO cells, $1*10^5$ HEK-293T cells in 24-well plates were transfected with 200ng PP7d-*HuR* and 20ng TU1-*SEAP* by PEI at a 1:5 DNA:PEI ratio, and their cell lysates analysed 72h post-transfection by RT-qPCR for relative abundance of *SEAP* mRNA, compared to a control cotransfected with FRT-TO-*GFP*, as in section 2.9.1. As shown in Figure 3.16, *SEAP* mRNA is significantly 2.73-fold more abundant in cells cotransfected with PP7d-*HuR* than with FRT-TO-*GFP*.

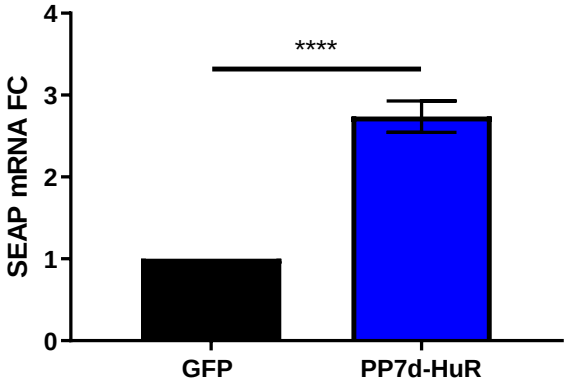


Figure 3.16 – PP7d-HuR overexpression by PEI transfection at a 1:5 DNA:PEI ratio increases mRNA abundance of transiently expressed SEAP in HEK-293T cells. Abundance of SEAP mRNA significantly increases by 2.73-fold, relative to an FRT-TO-GFP control (P<0.0001). Results presented are of technical triplicates, from three biological replicates. FC: fold-change.

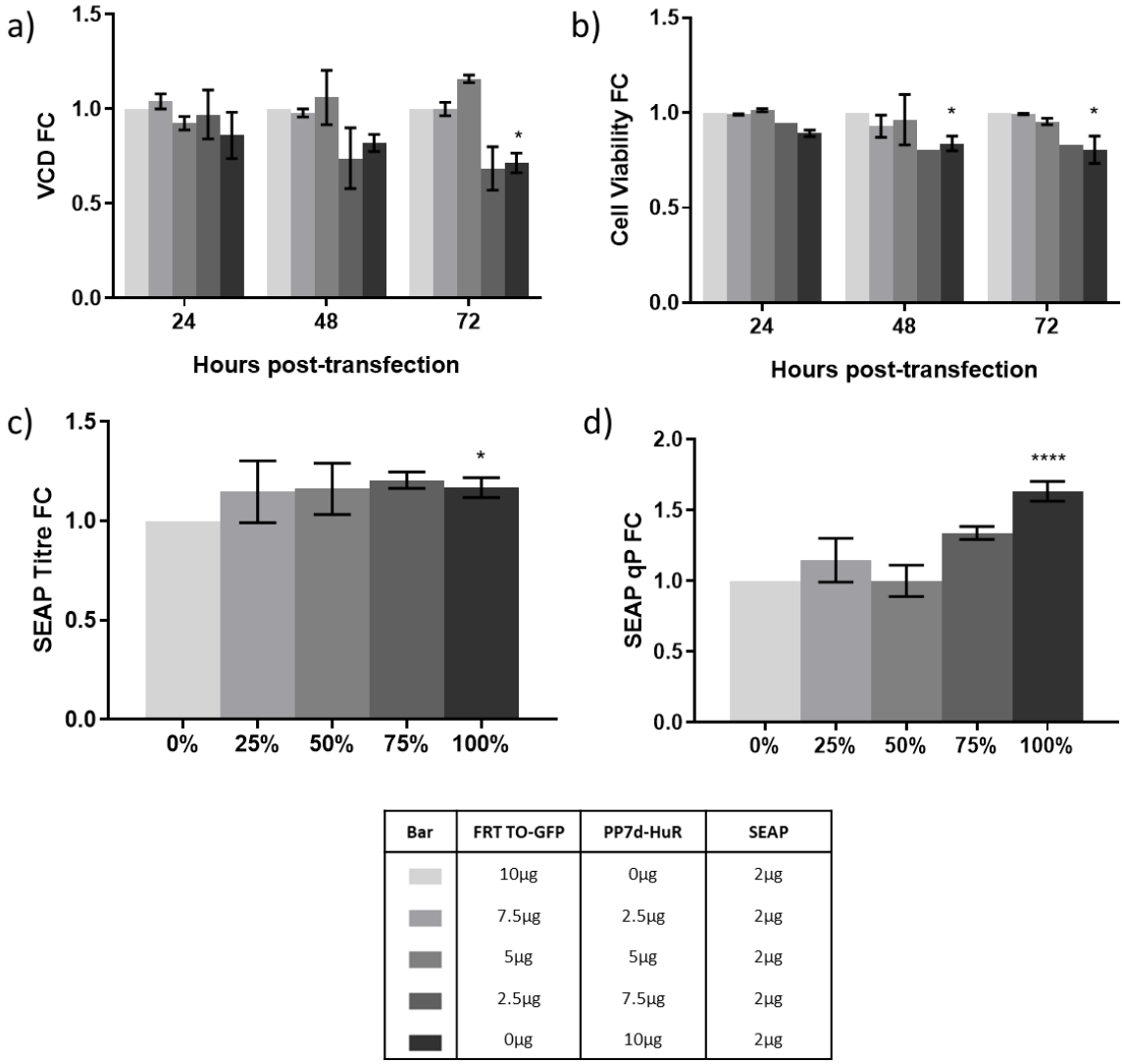
Taken together, the results from Figure 3.15 and Figure 3.16 imply that overexpression of PP7d-*HuR* can affect the cell-specific productivity of SEAP, through an effect granting a greater abundance of *SEAP* mRNA, but which also significantly decreases cell growth. The overall change in titre may therefore be derived from the difference in strength between each antagonistic effect.

3.2.8. Overexpression of PP7d-HuR is associated with an increase in transient SEAP specific productivity

Next, it was investigated whether this response to PP7d-*HuR* overexpression was dose-responsive, whether an optimal balance between cellular growth and specific productivity could be reached, and whether PP7d-*HuR* tethering could amplify this effect. To do this, the CHO transient host was transfected by PEI at a 1:5 DNA:PEI ratio with 2µg TU1-*SEAP*, both with and without 6xPP7bs in the 3'UTR, alongside a titration of PP7d-*HuR* against FRT-TO-*GFP*, from 0-10µg. These transfections were performed in 10ml cultiflask culture, at $1*10^6$ cells/ml. VCD of these transfections was measured 24h, 48h, and 72h post-transfection by ViCell as in section 2.4., and SEAP titre measured 72h post-transfection as in section 2.6.2. The overexpression results from SEAP without 6xPP7bs are displayed in Figure 3.17, and the tethering results from SEAP-6xPP7bs (SEAP-BS) are shown in Figure 3.18. In both cases, greater quantities of PP7d-*HuR* were associated with decreases in cell growth and viability, in a dose-responsive pattern. With SEAP lacking binding sites, the 100% titration led to a significant 0.72-fold decrease in viable cell density and a 0.81-fold decrease in cell viability 72h post-transfection, similar to SEAP-BS, where it conferred a 0.73-fold decrease in VCD and a 0.76-fold decrease in viability. In the case of SEAP, all titrations of *HuR* were associated with slight increases in SEAP titre, with 100% titration giving a significant 1.16-fold increase, as shown in Figure 3.17c. Figure 3.18c shows that no significant effect on titre was observed with SEAP-BS upon PP7d-*HuR* tethering titration. Small changes in titre despite decreases in cell growth imply a significant increase in specific productivity, and this is borne out by Figure 3.17d, where a dose-responsive increase in specific productivity of SEAP lacking binding sites is shown, up to a significant 1.63-fold increase at 100% titration. As shown in Figure 3.18d, this effect is more modest and less clearly dose-responsive when PP7d-*HuR* is tethered to SEAP-BS, with 75% and 100% titrations both significantly increasing specific productivity by 1.28-fold.

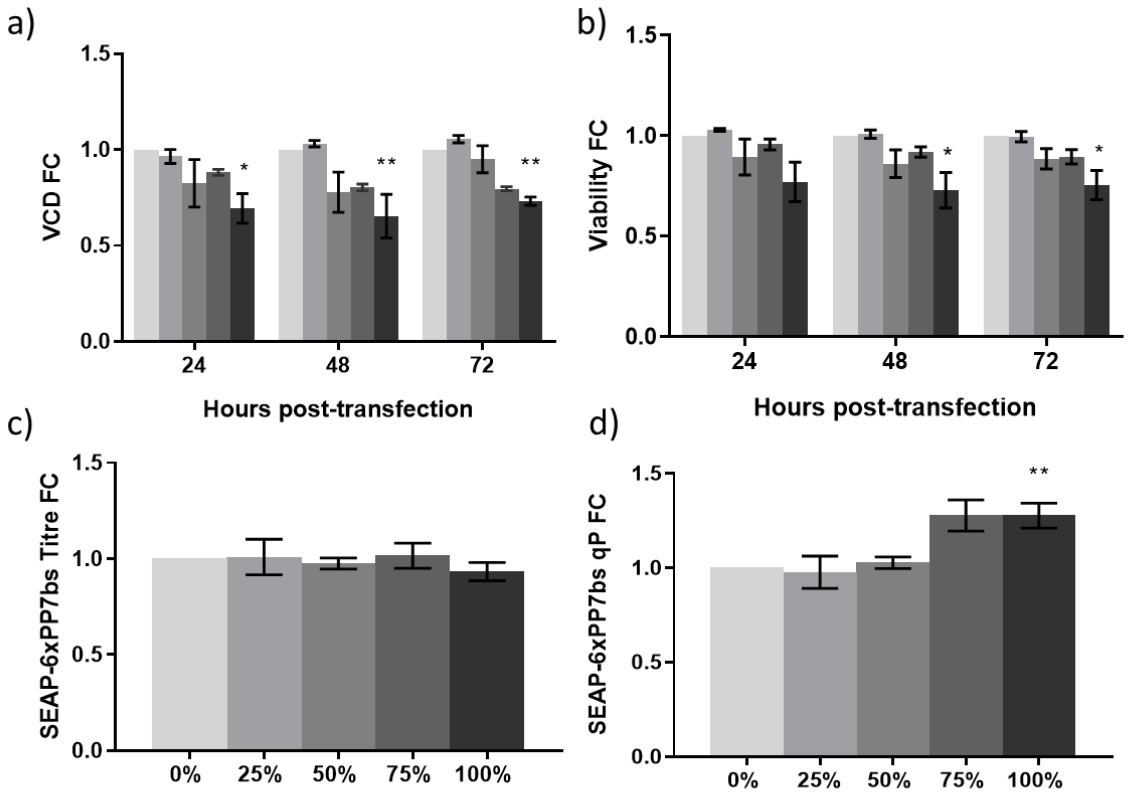
The data presented in Figure 3.17 and Figure 3.18 corroborate the hypothesis stated above, that PP7d-*HuR* overexpression affects transient SEAP titre by decreasing cell growth and increasing cell-specific productivity. Both of these effects were shown to be responsive to the total quantity of PP7d-*HuR* transfected, growth and specific productivity decreasing and

increasing respectively as PP7d-*HuR* was titrated into the transfections. For both SEAP and SEAP-BS, it showed that the optimal titration of PP7d-*HuR* in this case is 100%, giving both their highest SEAP titres and specific productivity. Finally, it showed that increases in titre occurred independently of mRNA-PP7d tethering. Moreover, tethering may actually inhibit the stimulation of specific productivity, whilst not mitigating the losses of cell growth associated with PP7d-*HuR* titration.



Bar	FRT TO-GFP	PP7d-HuR	SEAP
	10µg	0µg	2µg
	7.5µg	2.5µg	2µg
	5µg	5µg	2µg
	2.5µg	7.5µg	2µg
	0µg	10µg	2µg

Figure 3.17 – PP7d-HuR overexpression titration by PEI transfection at a 1:5 DNA:PEI ratio decreases cell growth and viability, and increases transient SEAP specific productivity in the CHO transient host. A) Overexpression titration of PP7-HuR decreases VCD, to a fold change of 0.71 from 0% to 100% at 72h (P=0.0044). B) Viability also decreases significantly, to 0.81-fold from 0% to 100% at 72h (P=0.014). C) SEAP titre increases significantly, to 1.17-fold between 0% to 100%. D) SEAP specific productivity increases significantly, to 1.63-fold at 72h (P<0.0001). Results presented are of technical duplicates, from three biological replicates. VCD: viable cell density; FC: fold-change; qp: specific productivity.



Bar	FRT TO-GFP	PP7d-HuR	SEAP-BS
	10µg	0µg	2µg
	7.5µg	2.5µg	2µg
	5µg	5µg	2µg
	2.5µg	7.5µg	2µg
	0µg	10µg	2µg

Figure 3.18 - PP7d-HuR tethering titration by PEI transfection at a 1:5 DNA:PEI ratio decreases cell growth and viability, and increases transient SEAP specific productivity in the CHO transient host. A) Overexpression titration of PP7-HuR decreases VCD, to a fold change of 0.73 from 0% to 100% at 72h (P=0.0072). B) Viability also decreases significantly, to 0.76-fold from 0% to 100% at 72h (P=0.046). C) No significant effect is measured on SEAP titre. D) SEAP specific productivity increases significantly, to 1.28-fold at 72h (P=0.0055). Results presented are of technical duplicates, from three biological replicates. VCD: viable cell density; FC: fold-change; qp: specific productivity.

3.2.9. Overexpression of HuR is associated with mixed effects on transient and stable mAb titre

To test whether *HuR* overexpression could replicate its stimulatory effect on SEAP specific productivity with a more industrially relevant protein, another cotransfection titration experiment was performed. As tethering was shown to have a negative impact on specific productivity compared to overexpression, a plasmid was constructed, by restriction/ligation cloning of CHO codon optimised *HuR* into the TU1 vector between the AgeI and SbfI sites as in section 2.3.6., and verification by Sanger sequencing, creating the TU1-*HuR* vector. A representative plasmid map can be found under 'TU1' in Appendix A. This construct was titrated up to 10μg against TU1-*GFP*, as in previous titration experiments. These titrations were cotransfected with 2μg of a plasmid bearing both the light and heavy chain of a DTE mAb using PEI at a 1:5 DNA:PEI ratio, into the transient CHO host, in 10ml cultiflask culture at $1*10^6$ cells/ml. Growth and viability were measured at 24h, 48h, 72h by ViCell as in section 2.4., and the titre of mAb measured 72h post-transfection by ELISA as in section 2.10.1. The dose-dependent decrease in VCD upon *HuR* titration is replicated in Figure 3.19a, with a significant decrease of 0.73-fold in VCD at 100% titration, 72h-post transfection. The effect on viability was not replicated, however. Figure 3.19b shows that viabilities across all TU1-*HuR* titrations were slightly, but not significantly decreased, and with no clear response to TU1-*HuR* dosage. In further contrast to results measured with SEAP, Figure 3.19c shows that titration of TU1-*HuR* was associated with a drastic decrease in mAb titre, in a dose-responsive pattern, and with a 10-fold reduction of 0.20mg/l at 0% to 0.021mg/l at 100% titration. Given the slightly decreased growth upon TU1-*HuR* titration, the effect on specific productivity was slightly less marked, but still clearly dose responsive, resulting in a reduction to 0.16-fold change between 0% and 100%. These data suggest that *HuR* overexpression may have a product-dependent effect on specific productivity of a recombinant protein.

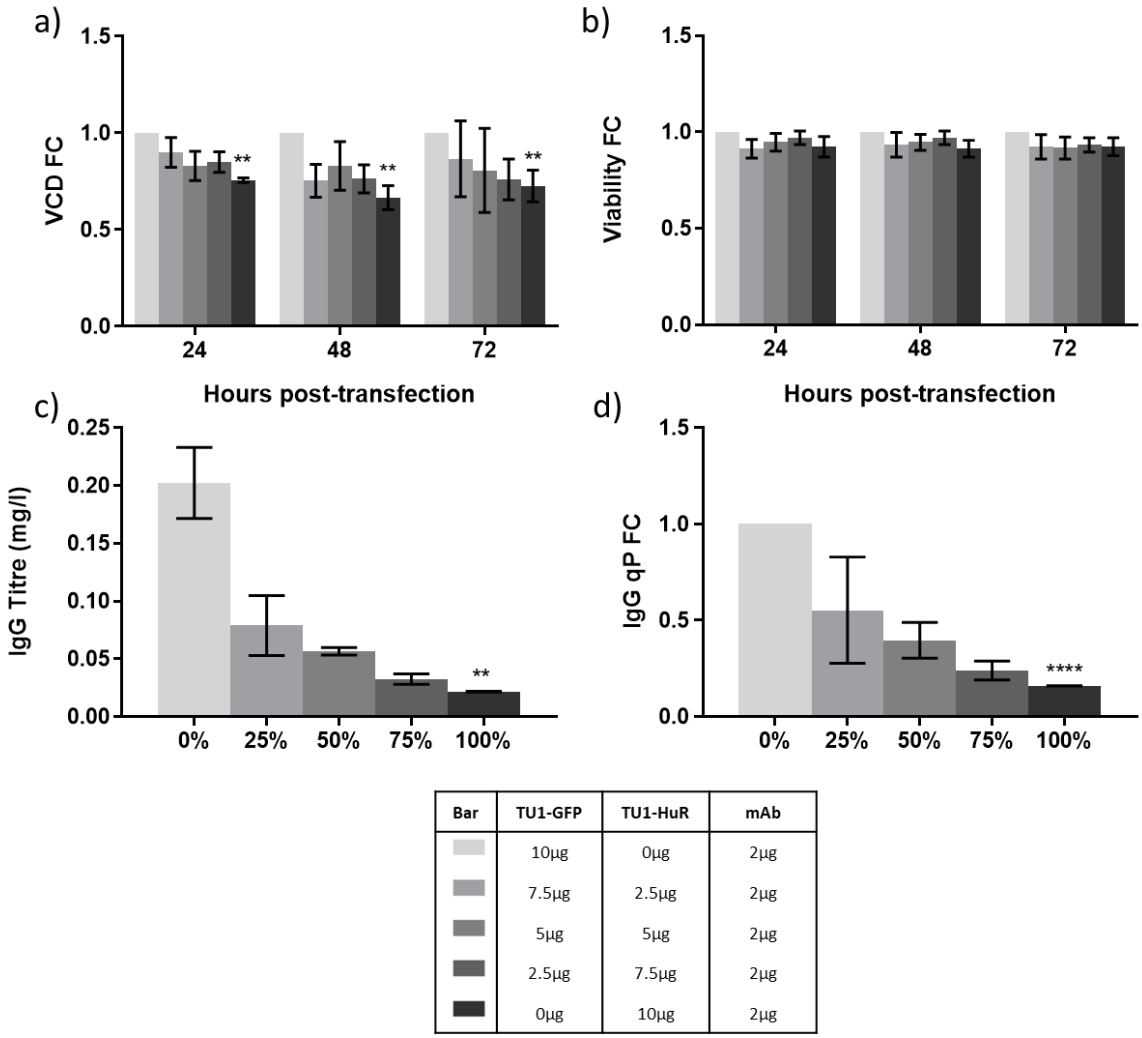


Figure 3.19 - TU1-HuR overexpression titration by PEI transfection at a 1:5 DNA:PEI ratio decreases cell growth, and transient expression of a DTE mAb in the CHO transient host. A) TU1-HuR overexpression titration decreases VCD, to 0.73-fold from 0% to 100% at 72h post-transfection (P=0.0019). B) No significant effect is measured on cell viability at 24h, 48h, or 72h post-transfection. C) IgG titre is significantly decreased across the overexpression titration, from 0.202mg/l at 0%, to 0.021mg/l at 100% (P=0.0042). D) IgG specific productivity is significantly decreased across the overexpression titration, to 0.16-fold from 0% to 100% (P<0.0001). Results presented are of technical duplicates, from three biological replicates. VCD: viable cell density; FC: fold-change; qp: specific productivity.

To investigate whether TU1-*HuR* overexpression could positively affect the production characteristics of a stable cell line, a similar titration of TU1-*HuR* was performed, into the CHO cell line stably expressing an ETE mAb. 10μg TU1-*HuR* was titrated against TU1-*GFP*, in 10ml cultiflask culture at $1*10^6$ cells/ml using PEI at a 1:5 DNA:PEI ratio. Growth was measured after 24h, 48h, 72h by ViCell as in section 2.4., and mAb titre measured 72h post-transfection by ELISA as in section 2.10.1. As shown in Figure 3.20a, titration of TU1-*HuR* initially led to a dose-responsive decrease in VCD, to a significant decrease of o.81-fold at 100% after 24h. However, the VCD of all titrations subsequently recovered compared to controls, and no significant differences are measured at either 48h or 72h. Figure 3.20b shows that no significant effects on cell viability were detected across the titrations, at any timepoint. Similarly, as shown in Figure 3.20c, no significant differences were measured in mAb titres across all of the transfections. These data suggest that TU1-*HuR* overexpression may not significantly affect the production characteristics of CHO cell lines stably expressing a recombinant protein.

Taken collectively, the data presented in this section suggest that in some circumstances, *HuR* overexpression can increase recombinant protein titres, by decreasing cell growth, and increasing cell-specific productivity. However, both the presence and strength of these effects appear dependent on the cell line, culture platform, product protein, and expression methods.

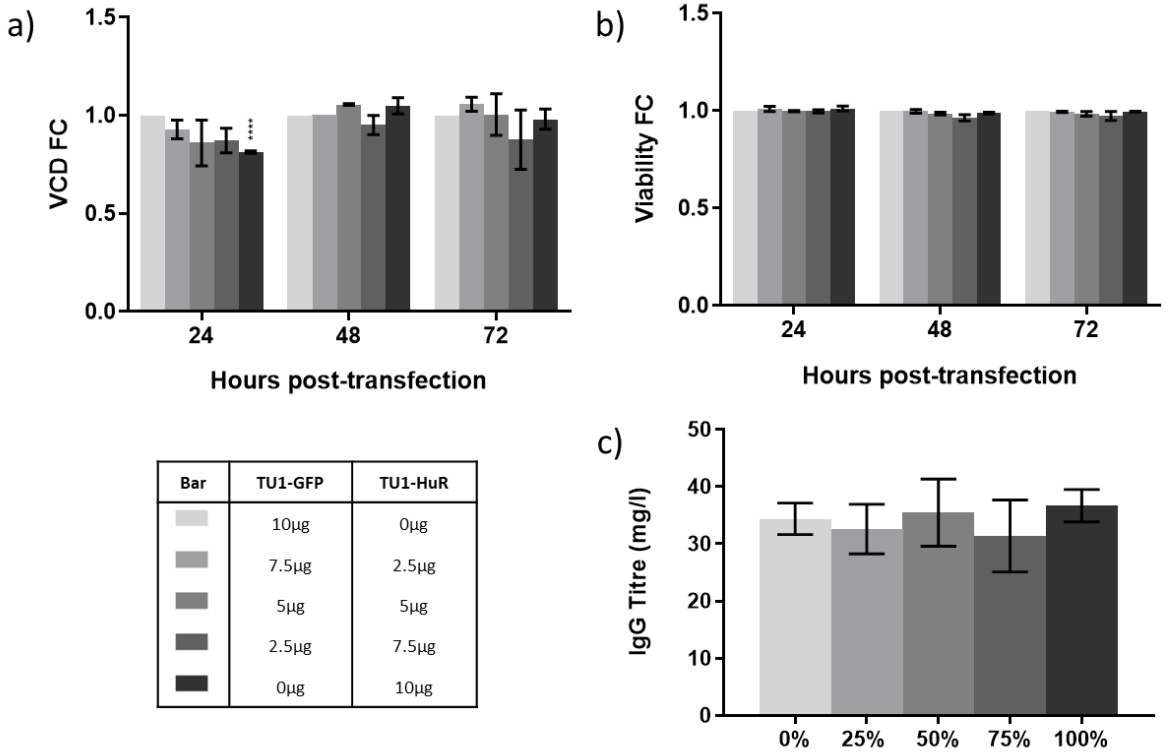


Figure 3.20 – TU1-HuR overexpression titration by PEI transfection at a 1:5 DNA:PEI ratio has no significant effect on a stable IgG expressing CHO cell line. A) Overexpression titration of TU1-HuR briefly decreases cell growth, with a significant 0.81-fold change between 0% and 100% at 24h (P<0.0001). Following this, no significant changes in VCD at 48h or 72h post-transfection were measured. B) No significant changes in cell viability at 24, 48, or 72 hours post transfection were measured. C) No significant changes in IgG titres were measured 72h post-transfection. Results presented are of technical duplicates, from three biological replicates. VCD: viable cell density; FC: fold-change.

3.3. Discussion

The investigations published in this chapter are mostly predicated on transient SEAP expression as a model for biopharmaceutical production. Whilst this is a commonly used model in literature (Brown *et al.,* 2019; Johari *et al.,* 2015; Peng *et al.,* 2011; Tastanova *et al.,* 2016), a set of rules for its' transfection and assay first had to be established. These rules were established by the results in Figure 3.3. Figure 3.3a demonstrates that, unlike in other assays such as ELISA, 405nm absorbance measured at the end of the assay is linearly correlated with concentration of SEAP present. This line of best fit does not intersect with the graph's point of origin, as a signal is still produced in blank samples lacking any SEAP. Therefore, after substituting the absorbance value of a blank sample, fold change in titre across an experiment can be inferred directly from the fold change in absorbance at 405nm, taken at the same time point. Figure 3.3b seems to contradict this finding, however, displaying an exponential increase in absorbance as quantity of SEAP transfected increases. This is explained by Figure 3.3c, in which it is shown that increased total DNA load, with a constant quantity of *SEAP* DNA, is stimulatory to SEAP titre. This is likely as a consequence of the increased transfection efficiency observed with HEK-293T cells and PEI, when overall DNA load is increased (Ehrhardt *et al.,* 2006). Therefore, the exponential relationship described in Figure 3.3b can be ascribed to the product of two linear factors: increased transfection efficiency due to DNA load, and increased SEAP titre due to increasing quantities of *SEAP* DNA. All directly-comparable transfections in future studies should therefore be performed with an equal total DNA load, and were in this research. The slight reduction in SEAP expression upon introduction of 6xPP7bs displayed in Figure 3.4 is not unexpected, as past studies have shown that inclusion of PP7/MS2 stem loops in mRNA can led to improper 3'UTR formation, and aberrant localisation in p-bodies, though this only becomes severe upon glucose starvation (Heinrich *et al.,* 2017).

Screening of PP7d fusion proteins with *SEAP* +/- PP7 binding sites gave the unexpected result that PP7d overexpression alone enhanced SEAP titre, compared to the GFP cotransfection control. This effect may be mediated by PP7d binding to mRNA having beneficial effects, for example in mRNA stability. However, this is unlikely, as the effect is replicated in the absence of PP7 binding sites, and it has been shown that the PP7 coat protein binding is very specific,

having decreased affinity by several orders of magnitude even for similar, but non-cognate RNA stem loops (Chao *et al.*, 2008; Lim *et al.*, 2001). All of the constructs used in this experiment utilise CMV promoters, which can interfere with one another through competition for transcription factors, in a manner dependent on transcript length (Huliák *et al.*, 2012). However, PP7d and GFP have very similar transcript lengths (828bp vs 720bp), making this explanation of the difference observed unlikely. Under conditions of ER stress, it has been shown that transiently expressed GFP can be secreted from the cell (Lee *et al.*, 2016a; Tanudji *et al.*, 2002). Therefore, cytoplasmic GFP could compete with SEAP for secretion, limiting expression in a mechanism neutralised by the NLS present in the PP7d construct.

Another surprising result from this screen was that overexpression of PP7d-*NXF1*-P2A-*P15* led to a significant decrease in SEAP titre, as shown in Figure 3.5. Numerous intronless mRNAs have been shown to enhance their export by sequestering *NXF1* (Braun *et al.*, 2001; Tian *et al.*, 2013; Yang *et al.*, 2014). Moreover, inclusion of an *NXF1*-sequestering motif has been shown to increase titre across SEAP, EPO, and Luciferase reporters, an increase which is further stimulated by co-expression of *NXF1* (Aihara *et al.*, 2011). An explanation for this may be that the PP7d fusion interferes with *NXF1*'s export capacity. An export-noncompetent *NXF1* that still bound to mRNPs would compete with WT *NXF1* present in the cell, diminishing exported mRNA and resultant titre. Alternatively, if the PP7d fusion rendered the SV40 NLS inert, the PP7d-*NXF1* construct could have sequestered competent mRNA in the cytoplasm, preventing its translation. However, this explanation would be contradictory to previous evidence suggesting that NXF1-P15 only binds mRNA after handover from TREX components (Viphakone *et al.*, 2012). This hypothesis of fusion proteins failing to re-enter the nucleus could nevertheless explain the general non-significant effect of overexpression and tethering of the effector gene fusion proteins. SV40 NLS's have been successfully utilized in combination with PP7d-fusion vectors in previous literature. However, in this case, an in frame polylinker was used to separate the two features, which may have prevented them from interfering with one another (Gesnel et al., 2009).

C1orf35 was identified from the initial screen presented in Figure 3.6 as stimulating transient SEAP titre. Further experiments confirmed this stimulatory effect, and protein lysate concentration measurements from HEK-293T cells, combined with VCD measurements from

transient host CHO cells implied that this effect was mediated through enhanced cell growth. This effect, however, was not replicated when co-expressing a DTE mAb, as shown in Figure 3.13. In this case, where expression of the DTE mAb may cause UPR induction and apoptosis (Chromikova *et al.*, 2015; Reinhart *et al.*, 2014), the stimulatory effect of *C1orf35* on growth may be negated. Whilst it is an uncharacterised protein, some work has been performed with *C1orf35*, with one study showing it's overexpression can increase cell growth and proliferation in NIH3T3 cells, by accelerating the transition from G1 to S phase in the cell cycle (Hu *et al.*, 2015). In Figure 3.7b, tethering of PP7d-*C1orf35* to an intron-containing SEAP vector unexpectedly reduced SEAP titre, and in Figure 3.13, titration of TU-*C1orf35* reduced titre of an intron-containing DTE mAb, implying a role for the protein in splicing. This observation is corroborated by a protein:protein interaction which has been detected between *C1orf35* and *Prp19* (Li *et al.*, 2015). A potential mechanism to link these two phenotypes could surface from *C1orf35* influencing alternative splicing, which has been shown to affect cell proliferation and survival (Prinos *et al.*, 2011). The picture is further complicated, however, by the detection of a nucleolar localisation sequence in *C1orf35* (Scott *et al.*, 2010), potentially implying a role in ribosome biogenesis. With all this considered, *C1orf35* may be an interesting gene for future research. However, its' failure to significantly reproduce this stimulation of cell culture performance in more industrially-relevant systems (a stable mAb-producing cell line, and transient expression of a DTE mAb), disqualify it as a predictable molecular biology tool, for positively controlling biopharmaceutical titre.

In contrast to *C1orf35*, neither PP7d-*HuR* overexpression nor tethering to SEAP in HEK-293T cells led to a significant change in titre, compared to GFP controls, as shown in Figure 3.5. However, when the experiment was replicated in transient-host CHO cells, to assess whether one cell line could be used as a model for the other, its overexpression produced a significant increase in SEAP titre, despite a negative effect on growth, as shown in Figure . Combined with Figure 3.16, in which it is shown at PP7d-*HuR* overexpression almost triples *SEAP* mRNA abundance, this discrepancy can be explained. Overexpression of *HuR* simultaneously increases specific productivity of recombinant proteins, and decreases cell growth. Since titre is a product of these two factors, the strength and significance of each antagonistic effect (dependent on the cell line, culture conditions, expression mode, etc.) relative to one another will dictate the overall change in titre.

HuR is well documented to increase the stability and translation efficiency of the mRNAs to which it binds. Binding of the nuclear-cytoplasmic shuttling protein *HuR* to *β-globin* mRNA via binding regions known as AREs enhanced the mRNAs stability (Fan and Steitz, 1998). Overexpression of the DNA damage response p53 protein upon UV irradiation has been shown to be partially due to *HuR* binding to the 3' UTR of p53 mRNA, leading to stabilisation and translational enhancement (Mazan-Mamczarz *et al.*, 2003). Knockdown of *HuR* led to reduced expression of genes containing 3' UTRs able to bind to *HuR*, as identified by RNA-protein crosslinking (Lebedeva *et al.*, 2011). Finally, it has been demonstrated that tethering of *HuR* to GFP mRNA by an MS2 system leads to greater cytoplasmic membrane localisation of GFP (Berkovits and Mayr, 2015). As shown in Figure 3.17 and Figure 3.18, this effect appears to be independent of, and perhaps even diminished by, tethering. This mechanism of increase in specific productivity through enhanced mRNA stability and translation is also compatible with the results presented Figure 3.19. Stimulating translation and lengthening mRNA half-life will effectively increase specific productivity in systems where the cell can fold, modify, and secrete the extra protein load it receives. This is the case with SEAP, where studies have shown that vector features can increase its' expression ~4-fold, before cell engineering must be performed to relieve bottlenecks downstream (Brown *et al.*, 2019). However, DTE mAbs most often form bottlenecks downstream, at the stages of post-translational modification and secretion, forming aggregates and inducing the UPR (Chromikova *et al.*, 2015; Reinhart *et al.*, 2014). Increasing the protein load in these systems would only further aggravate this UPR response, promoting ER-associated degradation, arresting growth, and inducing apoptosis. Like all synthetic expression tools, *HuR* cannot alleviate a bottleneck downstream of its point of action.

Regardless of their mechanism of action, neither of the candidate proteins investigated in this chapter were able to impart any permanent, significant impact on either the culture performance or specific productivity of the clonal stable mAb-producing CHO cell line. These experiments are shown in Figure 3.12 and Figure 3.20. Some titrations of TU-*C1orf35* or PP7d-*C1orf35* briefly increased VCD, and TU1-*HuR* briefly decreased VCD, before both reverted back to control levels. Neither construct at any titration had a significant effect on mAb titre compared to controls. This may be due to the problem with using effector genes for cell engineering explored in Chapter 1 of this thesis. CHO cells exhibit extremely high genetic

instability, meaning that non-clonal pools of cells – such as the CHO transient host – carry high genetic heterogeneity. This makes them a good generic background for the generalised testing of effector genes. However, solutions identified against this general background may not be effective in cells lines – such as the stable mAb producer – which have been exposed to rigorous clonal selection and expansion. Such cell lines have used the genetic heterogeneity of their ancestral pools as a resource, to find an optimal, single, genetic solution for the problem of the selection they underwent. For these reasons, the strategy of cellular engineering by effector genes investigated in this chapter does not represent a predictable synthetic biology tool in a model predicated on testing in heterogenous transient systems, before transfer to clonal, stable expression systems.

Complicating the dissection of these results is that fact that multiple variables were changed across experiments, for example, between Figure 3.17 and Figure 3.19, both the reporter (from SEAP to mAb) and effector gene (from unoptimized PP7d-HuR to codon-optimized TU1-*HuR*) were changed. This makes it difficult to disentangle which factor is responsible for the reversal of the effect on product titre observed. In this example, a positive control, cotransfecting optimized TU1-*HuR* with the SEAP reporter, could be performed to elucidate this question. However, in light of the limited time and resources of a PhD project, and the observation that as the experimental systems more closely resembled industrial conditions, interventions became more detrimental or ineffective toward reporter expression, it was decided to change methodology, toward direct manipulation of biopharmaceutical vectors, as will be shown in chapters 4, 5, and 6 of this thesis.

3.4. Conclusion

In this chapter, I have presented the screening of multiple candidate genes, using RNA:protein tethering technology, and an optimised in-house SEAP assay to measure their ability to stimulate and control biopharmaceutical expression. I identified two novel candidate proteins for cell engineering. *C1orf35* showed a putative ability to enhance cell culture performance, as PP7d-*C1orf35* overexpression was associated with a 1.38-fold increase in VCD 72 hours post-transfection. *HuR* showed putative and novel increases in specific productivity, with its' overexpression associated with an increase in SEAP mRNA of 2.73-fold in HEK-293T cells, and a 1.63-fold increase in SEAP specific productivity 72 hours post-transfection.

If effective, predictable, and titratable, these could have been powerful tools for biopharmaceutical expression. For instance, a reliable stimulator of cell growth and proliferation could have ameliorated the often-time-consuming step of culture scale-up, whilst an enhancer of cell-specific productivity could have aided the production of complex, next generation molecules, such as bispecifics. However, both tools were limited, unable to consistently replicate their effects across different reporters due to their mechanism of action. *C1orf35* failed to significantly alter cell growth when overexpressed in a stable IgG-expressing cell line and was associated with a 0.35-fold decrease in transient titre of a DET mAb. Likewise, *HuR* overexpression produced no significant change in a stable IgG-expressing cell line, and was associated with a 0.16-fold decrease in transient DTE mAb specific productivity. Due to these undesirable outcomes, combined with the seemingly intractable incompatibility of effector-gene based cell engineering between transient and stable production systems, neither of these candidate genes can be judged to be effective, predictable, synthetic molecular biology tools for biopharmaceutical expression. In future work, presented in the following chapters of this thesis, an alternative approach will be taken, utilising the direct modification of biopharmaceutical vector sequences.

3.5. References

Aihara, Y., Fujiwara, N., Yamazaki, T., Kambe, T., Nagao, M., Hirose, Y., and Masuda, S. (2011). Enhancing recombinant protein production in human cell lines with a constitutive transport element and mRNA export proteins. J. Biotechnol. 153, 86–91.

Berkovits, B.D., and Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature 522, 363–367.

Boyne, J.R., Colgan, K.J., and Whitehouse, A. (2008). Recruitment of the complete hTREX complex is required for Kaposi's sarcoma-associated herpesvirus intronless mRNA nuclear export and virus replication. PLoS Pathog. 4.

Braun, I.C., Herold, A., Rode, M., Conti, E., and Izaurralde, E. (2001). Overexpression of TAP/p15 Heterodimers Bypasses Nuclear Retention and Stimulates Nuclear mRNA Export. J. Biol. Chem. 276, 20536–20543.

Brown, A.J., Gibson, S.J., Hatton, D., Arnall, C.L., and James, D.C. (2019). Whole synthetic pathway engineering of recombinant protein production. Biotechnol. Bioeng. 116, 375–387.

Callis, J., Fromm, M., and Walbot, V. (1987). Introns increase gene expression in cultured maize cells. Genes Dev. 1, 1183–1200.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell 149, 1393–1406.

Chanarat, S., and Sträßer, K. (2013). Splicing and beyond: The many faces of the Prp19 complex. Biochim. Biophys. Acta - Mol. Cell Res. 1833, 2126–2134.

Chao, J. a, Patskovsky, Y., Almo, S.C., and Singer, R.H. (2008). Structural basis for the coevolution of a viral RNA-protein complex. Nat. Struct. Mol. Biol. 15, 103–105.

Chi, B., Wang, K., Du, Y., Gui, B., Chang, X., Wang, L., Fan, J., Chen, S., Wu, X., Li, G., et al. (2014). A Sub-Element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18. Nucleic Acids Res. 42, 7305–7318.

Chromikova, V., Mader, A., Steinfellner, W., and Kunert, R. (2015). Evaluating the bottlenecks of recombinant IgM production in mammalian cells. Cytotechnology 67, 343–356.

Ehrhardt, C., Schmolke, M., Matzke, A., Knoblauch, A., Will, C., Wixler, V., and Ludwig, S. (2006). Polyethylenimine, a cost-effective transfection reagent. Signal Transduct. 6, 179–184.

Fan, X.C., and Steitz, J.A. (1998). Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. EMBO J. 17, 3448–3460.

Gesnel, M., Del Gatto-Konczak, F., & Breathnach, R. (2009). Combined Use of MS2 and PP7 Coat Fusions Shows that TIA-1 Dominates hnRNP A1 for K-SAM Exon Splicing Control. Journal Of Biomedicine And Biotechnology, 2009, 1-6. doi: 10.1155/2009/104853

Heath, C.G., Viphakone, N., and Wilson, S.A. (2016). The role of TREX in gene expression and disease. Biochem. J. 473, 2911–2935.

Heinrich, S., Sidler, C.L., Azzalin, C.M., and Weis, K. (2017). Stem–loop RNA labeling can affect nuclear and cytoplasmic mRNA processing. RNA 23, 134–141.

Hocine, S., Raymond, P., Zenklusen, D., Chao, J. a, and Singer, R.H. (2012). Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. Nat. Methods 10, 1–5.

Hu, W.-X., Luo, S.-Q., Zhong, Y., Bu, X.-F., and Zhou, Y. (2015). Abstract LB-070: The studies of tumor-associated gene C1orf35 in pathogenesis of human multiple myeloma. p.

Huliák, I., Sike, A., Zencir, S., and Boros, I.M. (2012). The objectivity of reporters: Interference between physically unlinked promoters affects reporter gene expression in transient transfection experiments. DNA Cell Biol. 31, 1580–1584.

Johari, Y.B., Estes, S.D., Alves, C.S., Sinacore, M.S., and James, D.C. (2015). Integrated cell and process engineering for improved transient production of a "difficult-to-express" fusion protein by CHO cells. Biotechnol. Bioeng. 112, 2527–2542.

Juillard, F., Hiriart, E., Sergeant, N., Vingtdeux-Didier, V., Drobecq, H., Sergeant, A., Manet, E., and Gruffat, H. (2009). Epstein-Barr virus protein EB2 contains an N-terminal transferable nuclear export signal that promotes nucleocytoplasmic export by directly binding TAP/NXF1. J. Virol. 83, 12759–12768.

Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. Mol. Cell 43, 340–352.

Lee, J.G., Takahama, S., Zhang, G., Tomarev, S.I., and Ye, Y. (2016). Unconventional secretion of misfolded proteins promotes adaptation to proteasome dysfunction in mammalian cells. Nat. Cell Biol. 18, 765–776.

Lei, H., Dias, A.P., and Reed, R. (2011). Export and stability of naturally intronless mRNAs require specific coding region sequences and the TREX mRNA export complex. Proc. Natl. Acad. Sci. 108, 17985–17990.

Lei, H., Zhai, B., Yin, S., Gygi, S., and Reed, R. (2013). Evidence that a consensus element found in naturally intronless mRNAs promotes mRNA export. Nucleic Acids Res. 41, 2517–2525.

Li, X., Wang, W., Wang, J., Malovannaya, A., Xi, Y., Li, W., Guerra, R., Hawke, D.H., Qin, J., and Chen, J. (2015). Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. Mol. Syst. Biol. 11, 775.

Lim, F., Downey, T.P., and Peabody, D.S. (2001). Translational Repression and Specific RNA Binding by the Coat Protein of the Pseudomonas Phage PP7. J. Biol. Chem. 276, 22507–22513.

Lu, S., and Cullen, B.R. (2003). Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. RNA 618–630.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev. 19, 1512–1517.

Mazan-Mamczarz, K., Galbán, S., López de Silanes, I., Martindale, J.L., Atasoy, U., Keene, J.D., and Gorospe, M. (2003). RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation. Proc. Natl. Acad. Sci. U. S. A. 100, 8354–8359.

Okamura, M., Inose, H., and Masuda, S. (2014). RNA export through the NPC in eukaryotes. Genes (Basel). 6, 124–149.

Peng, R.W., Abellan, E., and Fussenegger, M. (2011). Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells. Biotechnol. Bioeng. 108, 611–620.

Prinos, P., Garneau, D., Lucier, J.F., Gendron, D., Couture, S., Boivin, M., Brosseau, J.P., Lapointe, E., Thibault, P., Durand, M., et al. (2011). Alternative splicing of SYK regulates mitosis and cell survival. Nat. Struct. Mol. Biol. 18, 673–679.

Rathus, C., Bower, R., and Birch, R.G. (1993). Effects of promoter, intron and enhancer elements on transient gene expression in sugar-cane and carrot protoplasts. Plant Mol. Biol. 23, 613–618.

Reinhart, D., Sommeregger, W., Debreczeny, M., Gludovacz, E., and Kunert, R. (2014). In search of expression bottlenecks in recombinant CHO cell lines - A case study. Appl. Microbiol. Biotechnol. 98, 5959–5965.

Scott, M.S., Boisvert, F.M., McDowall, M.D., Lamond, A.I., and Barton, G.J. (2010). Characterization and prediction of protein nucleolar localization sequences. Nucleic Acids Res. 38, 7388–7399.

Tan, H.K., Lee, M.M., Yap, M.G.S., and Wang, D.I.C. (2008). Overexpression of cold-inducible RNA-binding protein increases interferon-gamma production in Chinese-hamster ovary cells. Biotechnol. Appl. Biochem. 49, 247–257.

Tanudji, M., Hevi, S., and Chuck, S.L. (2002). Improperly folded green fluorescent protein is secreted via a non-classical pathway. J. Cell Sci. 115, 3849–3857.

Tastanova, A., Schulz, A., Folcher, M., Tolstrup, A., Puklowski, A., Kaufmann, H., and Fussenegger, M. (2016). Overexpression of YY1 increases the protein production in mammalian cells. J. Biotechnol. 219, 72–85.

Tian, X., Devi-Rao, G., Golovanov, A.P., and Sandri-Goldin, R.M. (2013). The interaction of the cellular export adaptor protein Aly/REF with ICP27 contributes to the efficiency of herpes simplex virus 1 mRNA export. J. Virol. 87, 7210–7217.

Tintaru, A.M., Hautbergue, G.M., Hounslow, A.M., Hung, M.-L., Lian, L.-Y., Craven, C.J., and Wilson, S. a (2007). Structural and functional analysis of RNA and TAP binding to SF2/ASF. EMBO Rep. 8, 756–762.

Tunnicliffe, R.B., Hautbergue, G.M., Wilson, S.A., Kalra, P., and Golovanov, A.P. (2014). Competitive and Cooperative Interactions Mediate RNA Transfer from Herpesvirus Saimiri ORF57 to the Mammalian Export Adaptor ALYREF. PLoS Pathog. 10.

Viphakone, N., Hautbergue, G.M., Walsh, M., Chang, C.-T., Holland, A., Folco, E.G., Reed, R., and Wilson, S.A. (2012). TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nat. Commun. 3, 1006.

Viphakone, N., Sudbery, I., Griffith, L., Heath, C.G., Sims, D., and Wilson, S.A. (2019). Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. Mol. Cell 1–14.

Wu, B., Chao, J.A., and Singer, R.H. (2012). Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. Biophys. J. 102, 2936–2944.

Wu, B., Chen, J., and Singer, R.H. (2014). Background free imaging of single mRNAs in live cells using split fluorescent proteins. Sci. Rep. 4, 3615.

Xia, Z., Zheng, X., Zheng, H., Liu, X., Yang, Z., and Wang, X. (2012). Cold-inducible RNA-binding protein (CIRP) regulates target mRNA stabilization in the mouse testis. FEBS Lett. 586, 3299–3308.

Yang, C.C., Huang, E.Y., Li, H.C., Su, P.Y., and Shih, C. (2014). Nuclear export of human hepatitis B virus core protein and pregenomic RNA depends on the cellular NXF1-p15 machinery. PLoS One 9.

Zhu, X., Bührer, C., and Wellmann, S. (2016). Cold-inducible proteins CIRP and RBM3, a unique couple with activities far beyond the cold. Cell. Mol. Life Sci. 1–21.

4. The 3'UTR and mRNA Stability

4.1. Introduction

Suboptimal mRNA stability can drastically affect production of recombinant protein, in both stable and transient systems (Hung *et al.,* 2010; Mason *et al.,* 2012). Traditionally, the most relevant remedy for mRNA stability in biopharmaceutical production has been maximising codon optimality (Hung *et al.,* 2010; Presnyak *et al.,* 2015). However this approach is limited, for instance in cases where overly-efficient translation elongation may result in sub-optimal protein folding, or where mRNA instability is insensitive to codon optimisation, such as with poly-proline tracts causing ribosome stalling (Rodnina, 2016).

Multiple mechanisms exist to control mRNA stability, which could be developed as molecular biology tools. Many RNA-binding proteins effect mRNA stability, in particular those that bind to the 3'UTR. AU-rich elements (AREs) in the 3'UTR for instance regulate binding of a proteins such as *TTP*, which encourages rapid degradation of mRNA through binding of *CNOT1*, a scaffolding protein for the CCR4/NOT1 complex (Fukao and Fujiwara, 2017). On the other hand, binding of AREs by *HuR* (overexpression of which was explored earlier in this thesis) has been shown to extend mRNA half-life and translation (Berkovits and Mayr, 2015; Fan and Steitz, 1998; Wu *et al.,* 2015). In the 5'UTR, binding of 5' terminal oligo-pyrimidine motifs (5'TOP) by *LARP1* can extend half-life of mRNA by sequestering the 40S ribosome subunit (Gentilella *et al.,* 2017). Related to this is the hypothesis that efficiency of translation initiation, as opposed to elongation, is primarily responsible for the stability of most mRNAs, as competition for binding provided by the PIC inhibits decapping and deadenylation machinery, and that slowing of elongation can in some cases increase mRNA half-life (Chan *et al.,* 2018). Presence of miRNA binding sites in the 3'UTR can also influence mRNA stability and protein output (Radhakrishnan and Green, 2016). Proliferating cells have been shown to transcribe shortened 3'UTRs, in order to avoid this regulation (Sandberg *et al.,* 2008b), and design of synthetic 3'UTRs to avoid miRNA recognition sites could be a powerful tool. The sequence and structural features of the 3'UTR also significantly affect mRNA stability, for instance observed in the positive correlation between 3'UTR length and mRNA stability in Zebrafish (Mishima and Tomari, 2016). 3'UTR secondary structural features can influence

mRNA half-life through factors such as optimising the distance from the PolyA signal to cleavage site by formation of stem-loops (Wu and Bartel, 2017), or by specific stem-loop binding by proteins such as *HRNPA2B1* (Goodarzi *et al.,* 2012).

4.1.1. 3'UTR Vector Elements for Modulating mRNA Stability

In light of the results presented in Chapter 3 of this thesis, and the concluded difficulty of translating cellular engineering strategies from transient to stable expression systems, it was decided to pursue a strategy of vector engineering for control of mRNA stability. As the 3'UTR primarily coordinates mRNA stability at the vector level, it was selected for engineering. In this chapter, two potential vector tools will be tested for control of biopharmaceutical titre through mRNA stability: 3'triple helices, and stability elements.

3′ Triple Helices

Triple helices denote a wide family of RNA secondary structures, with a diverse series of functions. For instance, the SAM-II triple helix functions as a riboswitch, coordinating gene expression in response to specific S-adenosylmethionine binding, and the TER triple helix acts as an essential cofactor for telomerase activity (Conrad, 2014). The category of triple helices which may be candidates for this molecular biology tool are Elements for Nuclear Expression (ENEs). They have been discovered in many organisms, localised near the 3′ terminus of multiple transcripts, such as in noncoding RNA from DNA and RNA viruses (Tycowski *et al.,* 2012), mammalian lncRNAs (Brown *et al.,* 2012), and in transposable element RNAs across plants and fungi (Tycowski *et al.,* 2016). Their generalised function is to protect RNA from deadenylation and decay, and their enrichment in intronless RNAs suggest that they may have evolved to overcome expression deficits associated with a lack of splicing (Conrad, 2014; Tycowski *et al.,* 2016). Structural studies have shown that both viral and mammalian ENEs wrap around and 'lock down' a PolyA sequence or A-rich tract, sterically inhibiting deadenylation to achieve this effect (Brown *et al.,* 2014b; Mitton-Fry *et al.,* 2010). The best studied viral ENE, found in the Kaposi's sarcoma–associated herpesvirus PAN RNA, has been shown to extend the half-life of both the PAN RNA (Mitton-Fry *et al.,* 2010), and an intronless *β-globin* transcript (Brown *et al.,* 2014b; Tycowski *et al.,* 2012). The two best characterised mammalian ENEs, found in the MALAT1 and MENβ lncRNAs, have been shown not only to replicate this effect, but to increase translational efficiency when inserted into mRNA, and to

produce comparable expression levels to a PolyA control, whilst lacking a PolyA downstream (Brown *et al.,* 2012, 2014b; Wilusz *et al.,* 2012).

Stability Elements

As opposed to the prescriptive method of target identification taken above, assigning a characterised family of features to a function, a descriptive method can be taken, by deploying an MPRA platform, as described in section 1.2.3.7. of this thesis. The stability elements tested in this chapter were identified in this manner, in Oikonomou *et al.,* 2014. The 3'UTRs of multiple vertebrate genomes were analysed for conserved 34nt sequences, producing a library of 16,332 sequences. These sequences were synthesized on a microarray, amplified, inserted downstream of a fluorescent reporter, and transfected together into a pool of *Flp*-293T cells. Cells were sorted by fluorescent intensity by FACS, and the bins subjected to deep sequencing. 34nt sequences enriched in the highest-expressing bins were denoted activators of expression, and when inserted downstream of an mCherry reporter, demonstrated increased protein expression and mCherry transcript abundance, as measured by flow cytometry and RT-qPCR respectively (Oikonomou *et al.,* 2014).

4.2. Results

4.2.1. Generating the SEAP-3' Triple Helices library

In order to test the hypothesis that 3' triple helices (3'TH) could control mRNA dynamics, and thereby titre of a recombinant protein, a library of SEAP-3' triple helix reporters was created, utilising the MALAT1 and MENβ sequences (Brown *et al.,* 2012), and PAN ENE (Brown *et al.,* 2014b). To identify endogenous CHO sequences for MALAT1 and MENβ, published consensus sequences (Wilusz *et al.,* 2012) were aligned against the CHO-K1 genome assembly (2014), and the sequence covering the length of the alignment used. For the non-endogenous PAN ENE, the sequence was taken directly from literature (Mitton-Fry *et al.,* 2010). These sequences were then inserted downstream of the 3'UTR, in the TU1-*SEAP* reporter vector. As they have been tested both with (Brown *et al.,* 2012) and without (Wilusz *et al.,* 2012) PolyA tails, both variants of the MALAT1 and MENβ constructs were created, using the SV40 PolyA sequence found in unmodified TU1-*SEAP*. The PolyA tails forms an essential part of the PAN ENE structure (Mitton-Fry *et al.,* 2010), so no PolyA-lacking variant was created. Molecular

cloning was performed by restriction/ligation with SbfI/NotI and Not1/KpnI as in section 2.3.6., correct construction verified by Sanger sequencing, the sequences used can be found in section 2.13.2 and Table , and a representative plasmid map found under 'TU1-SEAP-Triple Helices' in Appendix A. Schematics of the TU1-SEAP-Triple Helices reporters used in this chapter are shown in Figure 4.1.

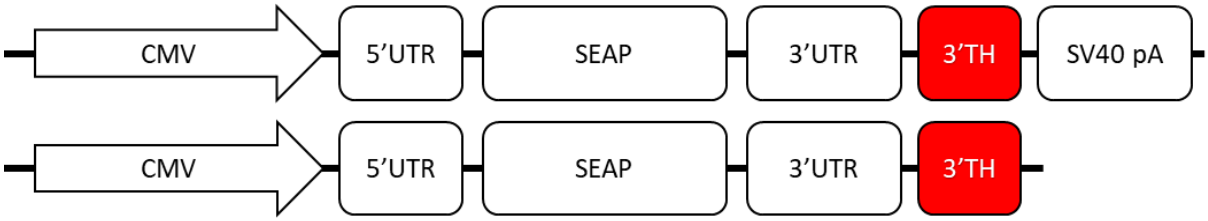


Figure 4.1 – Schematics of the TU1-SEAP-Triple Helices constructs used in this chapter. 3'TH: 3' Triple Helix.

Name	Sequence
MALAT1	CAGTAGGGCTGTAAAGGTTTTTCTTTTCCTGAGAAAACAAACTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTAAAAAAAAAGCAAAAGACACTGGTGGCCGGCACTCCTGGTCTCCAGGACGGGGTTCAAATCCCTGCGGTGTCT
MENβ	GTAGGGCTGTAAAGGTTTTTCTTTTCCTGAGAAAACAAACTTTTGTTTTCTCAGGTTTTGCTTTTTGGCCTTTCCCTAGCTTAAAAAAAAAGCAAAAGACACTGGTGGCCGGCACTCCTGGTCTCCAGGACGGGGTTCAAATCCCTGCGGTGTCT
PAN	TGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAGGTTTTTATCCCAGTGTATATT

Table 4.1 – The sequences of the MALAT1, MENβ, and PAN triple helices used in this study. CHO MALAT1 and MENβ triple helices sequences were identified by aligning established human triple helices sequences (Brown et al., 2012, 2014b), with the CHO-K1 genome, taking the region of homology as the CHO triple helices. The PAN triple helix sequence was taken from literature (Mitton-Fry et al., 2010; Tycowski et al., 2012). Triple helices were placed directly downstream of the 3'UTR, in the AstraZeneca in-house TU1 vector, inserted between the SbfI/NotI restriction sites, both with and without a PolyA sequence downstream, between the NotI/ KpnI restriction sites, by restriction/ligation cloning.

4.2.2. The addition of 3′ Triple Helices is not associated with an increase in transient SEAP titre

To analyse this library, transient host CHO cells at a density of $1*10^6$ cells/ml in 96 deep-well plates were transfected with each TU1-*SEAP*-3′TH reporter using PEI at a 1:5 DNA:PEI ratio. At 72 hours post-transfection, fold change in cell growth was measured by Prestoblue assay as in section 2.4.4.2., and fold change in SEAP titre by Sensolyte assay as in section 2.6.2., compared to a control transfected with unmodified TU1-*SEAP*. As seen in the results of this screen presented in Figure 4.2, transfection of both constructs containing an endogenous triple helix, in the absence of PolyA, drastically reduced SEAP titre compared to TU1-*SEAP*, to 0.10-fold for SEAP-MALAT1, and o.11-fold for SEAP-MENβ, albeit with slightly reduced growth resulting in a marginally higher fold-change in SEAP specific productivity for each. Addition of a PolyA downstream increased titre of both to an extent, up to 0.30-fold for SEAP-MALAT1-pA and 0.27-fold for SEAP-MENβ-pA, once again with reduced growth slightly increasing the specific productivity relative to TU1-*SEAP*. In contrast, SEAP-PAN-pA displayed a 1.24-fold and 1.37-fold increase in SEAP titre and specific productivity, respectively. These fold changes, however, were non-significant. These results show that the endogenous MALAT1 and MENβ triple helices are not adequate replacements or supplements for PolyA sequences in recombinant protein expression. They also imply that the PAN ENE may be capable of increasing SEAP titre, though statistical significance could not be discerned in a high-throughput system which displays high variability.

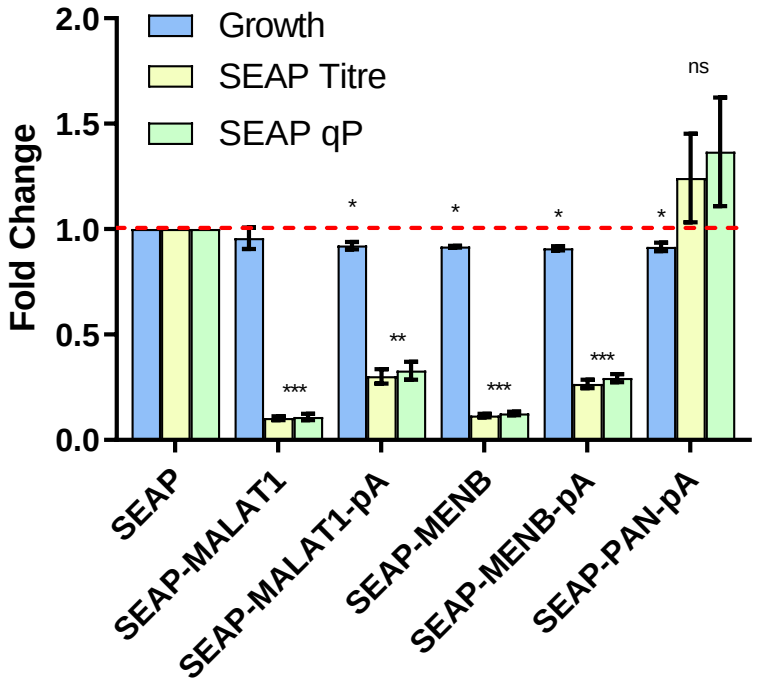


Figure 4.2 – Comparative analysis of TU1-SEAP-3' Triple Helix library by PEI transfection at a 1:5 DNA:PEI ratio in the CHO transient host in 96 deep-well plates. Four of the constructs led to a significant (P<0.05) decrease in growth: SEAP-MALAT1-pA (0.92-fold), SEAP-MENβ (0.92-fold), SEAP-MENβ-pA (0.91-fold), and SEAP-PAN-pA (0.92-fold). SEAP-MALAT1 and SEAP-MENβ led to a significant (P<0.0001) decrease in SEAP titre and qP, of 0.11-fold and 0.12-fold, respectively. SEAP-MALAT1-pA and SEAP-MENβ-pA led to a significant (P<0.0005) decrease in SEAP titre and qP, of 0.32-fold and 0.29-fold, respectively. SEAP-PAN-pA gives a non-significant1.24-fold increase in SEAP titre, and 1.37-fold increase in SEAP qP. All triple helix sequences referenced can be found in section 2.13.2. and Table of this thesis. Results presented are of technical duplicates, from four biological replicates. qP: specific productivity.

To confirm whether the PAN ENE increases SEAP titre in a lower-throughput, less variable system, TU1-*SEAP*-PAN-pA and the TU1-*SEAP* control were transfected into transient host CHO cells by PEI at a 1:5 DNA:PEI ratio, in 10ml culture at $1*10^6$ cells/ml in cultiflasks. Cell density and viability were measured 72h and 120h post-transfection by ViCell as in section 2.4., and these values used to calculate the IVCD of each culture. Fold change in SEAP titre was measured 72h and 120h post-transfection by Sensolyte assay, as in section 2.6.2. The results from this experiment are shown in Figure 4.3. No significant changes were measured in either cell culture performance, shown by IVCD in Figure 4.3a and by cell viability in Figure 4.3b, or in SEAP titre, shown 72h post-transfection in Figure 4.3c, and 120h post-transfection in Figure 4.3d. These data, combined with those presented in Figure 4.3, suggest that the PAN-ENE triple helix does not significantly affect transient production of SEAP, and that any non-significant increases in titre and specific productivity originally detected were products of the high variability associated with high throughput screening systems.

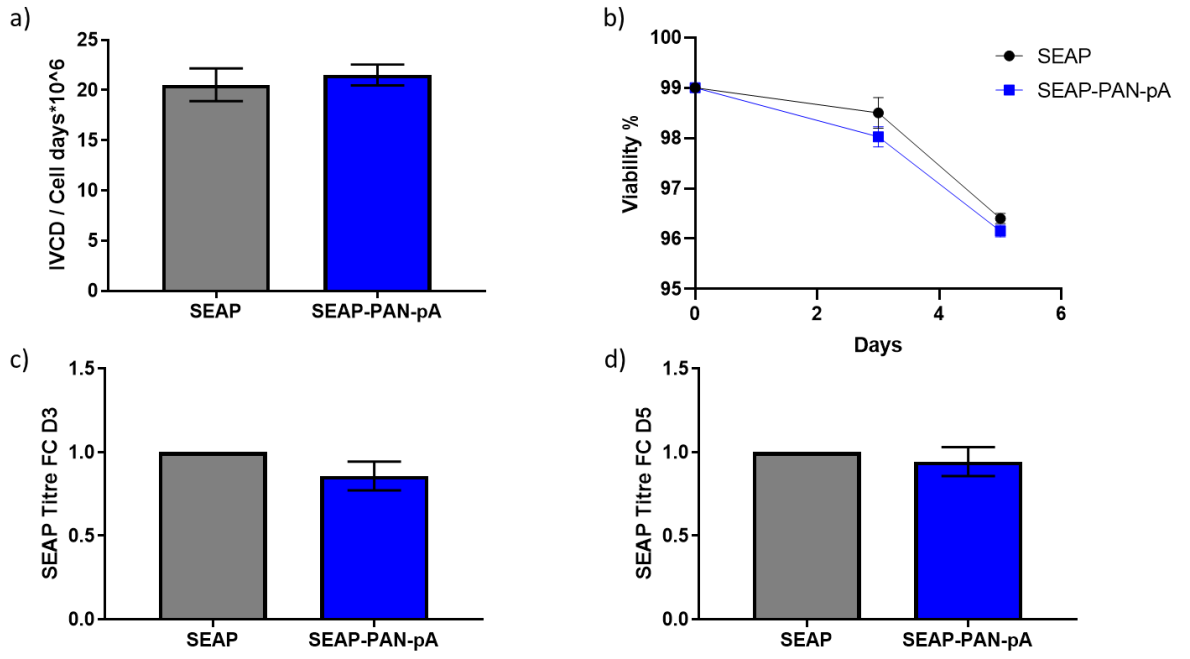


Figure 4.3 – The PAN ENE has no significant effect on transient SEAP titre, or cell culture performance when transfected by PEI at a 1:5 DNA:PEI ratio into the CHO transient host in cultiflasks. A) No significant difference is found between the IVCD of cultures transfected with TU1-SEAP and TU1-SEAP-PAN-pA, with values of $20.5*10^6$ and $21.5*10^6$ cell days respectively. B) No significant difference in viability is found, with values of 98.5% and 98.0% respectively at day 3, and 96.4% and 96.2% respectively at day 5. C) Titre of SEAP-PAN-pA is non-significantly reduced by 0.86-fold compared to SEAP, 72h post-transfection. D) Titre of SEAP-PAN-pA is non-significantly reduced by 0.94-fold compared to SEAP, 120h post-transfection. The PAN triple helix sequences referenced can be found in section 2.13.2. and Table of this thesis. Results presented are of technical duplicates, from four biological replicates. IVCD: integral of viable cell density; FC: fold-change.

4.2.3. Generating the SEAP-3′ Stability Element Library

To build a library of short stability elements, to test for their ability as part of a recombinant mRNA to control mRNA half-life, and thereby protein expression, the 34nt conserved sequences identified in Oikonomou *et al.*, 2014 were utilised. The total library of elements discovered in this study were ranked according to three metrics: their q-value, describing the significance of their enrichment in high-abundance compared to low-abundance mRNA bins, the ratio of their enrichment in the highest, compared to the lowest abundance mRNA bins, and their total enrichment in the highest abundance mRNA bin. The seven highest of each category were chosen. In addition, a randomly shuffled element, and the element most significantly enriched in low-expressing mRNA bins were included, making a total library of 24 constructs, including a non-modified control. All of these 34nt elements were inserted directly downstream of the *SEAP* CDS in the TU1-*SEAP* vector by site-directed mutagenesis as in section 2.3.2., and correct insertion verified by Sanger sequencing. Stability element sequences can be found in section 2.13.3. and Table 1, and a plasmid map showing their location in Appendix A. A schematic of the TU1-SEAP-stability element constructs can be found in Figure 4.4.

Figure 4.4 – A schematic of the TU1-SEAP-stabiltiy element constructs used in this chapter. SE: stability element.

Name	Oikonomou *et al.*, 2014 SequenceID	Sequence
SE1	C3U-seq15901	TTTGTTTTAGATGGAATAGCACAAGGAGAAAAAT
SE2	C3U-seq10578	GTTTTTTGAGGAATCTCAAGATGTGATATATTGG
SE3	C3U-seq4185	ATGTCTCCAGTTACAACTCCGCAGTGGATGTGAA
SE4	C3U-seq8401	GCAATTTAGCATGTTGGAACGTCTAGGGAGAAGG
SE5	C3U-seq12424	TCTCATTCCAGTAAGGCAGTTAGACACTTGAGTT
SE6	C3U-seq5465	CAGTTGAGATGAAGCACGTCGTTAGAACGTTGTT
SE7	C3U-seq212	AAAATGTAAAAATGTAACTATAGCATATGAATTG
SE8	C3U-seq1204	AAGTGGAGGTCTGGTTTGTAACTTTCCTTGTACT
SE9	C3U-seq8665	GCCTTAGGAGACTGGAAGTTTAAAAATGTACAAG
SE10	C3U-seq3598	ATAGCTGTACAAATATAAGAATAAAATGTTGAAA
SE11	C3U-seq9065	GGAGAAAGCTTCTCTATTTTGGATGCATTTCAGA
SE12	C3U-seq477	AAAGTTGCAAGATAAACAGCTGTAATTCGGACAA
SE13	C3U-seq10512	GTTTTAAGTAACTTTTTATAGCAAGATGATACAA
SE14	C3U-seq16091	TTTTGACTATTTTTATATATAAAGAAGAACTCAA
SE15	C3U-seq14491	TTATTGTGGATAACAAAGATATCTTTTCTTTAGA
SE16	C3U-seq13367	TGGCAGGTATTCCCATGATTCACAGAGTTACATT
SE17	C3U-seq248	AAACAAAAGCCTGGCTGAGTTGATGTTTTACATT
SE18	C3U-seq5093	CACAGTATTCGTGAATAAGTTGATTCTGTCCCCC
SE19	C3U-seq5184	CACTGAAGAGGTGGAAAAATAATCGTGTCAATCT
SE20	C3U-seq2244	ACTATAAATGCTTTGCAAAAATGGTTTCACGTTT
SE21	C3U-seq13645	TGTAGATCATAGGATAGCTGACTTTGACAGTCAC
SE22	C3U-seq1502	AATGGAACACAGACAGTGTAGAAGAATTCCTGAG
SE23	N/A	AGTAAACTGACGTTGTCCAACGTGCATATGGATT

Table 1.2 – The stability element sequences used in this study. 34nt stability elements were extracted from the data of Oikonomou et al., 2014, and placed directly downstream of the reporter CDS-bordering SbfI restriction site in the AstraZeneca in-house TU1 vector, upstream of the 3'UTR, by site-directed mutagenesis. SE: stability element.

4.2.4. Screening of the SEAP-3′ Stability Element library

This library was transfected into transient host CHO cells at a density of $1*10^6$ cells/ml using PEI at a 1:5 DNA:PEI ratio, in 96 deep-well plates. Fold-change in cell growth was measured by Prestoblue assay as in section 2.4., and SEAP titre by Sensolyte assay as in section 2.6.2., 72h post-transfection. As shown by the results presented in Figure 4.5a, a number of stability elements were associated with a slight increase in SEAP titre, three of which did so significantly: SE8, 10, and 20, to a maximum increase of 1.74-fold with SE8. Figure 4.5b shows that expression of 14 stability elements was associated with a slight decrease in cell growth after 72h. However, these decreases in cell growth display a regular pattern of periodicity every 12 samples, which will be discussed later in this chapter. Finally, in Figure 4.5c these data show a significant increase in SEAP specific productivity associated with 6 stability elements: SE2,3,7,8,10, and 20, to a maximum increase of 1.90-fold, again with SE8.

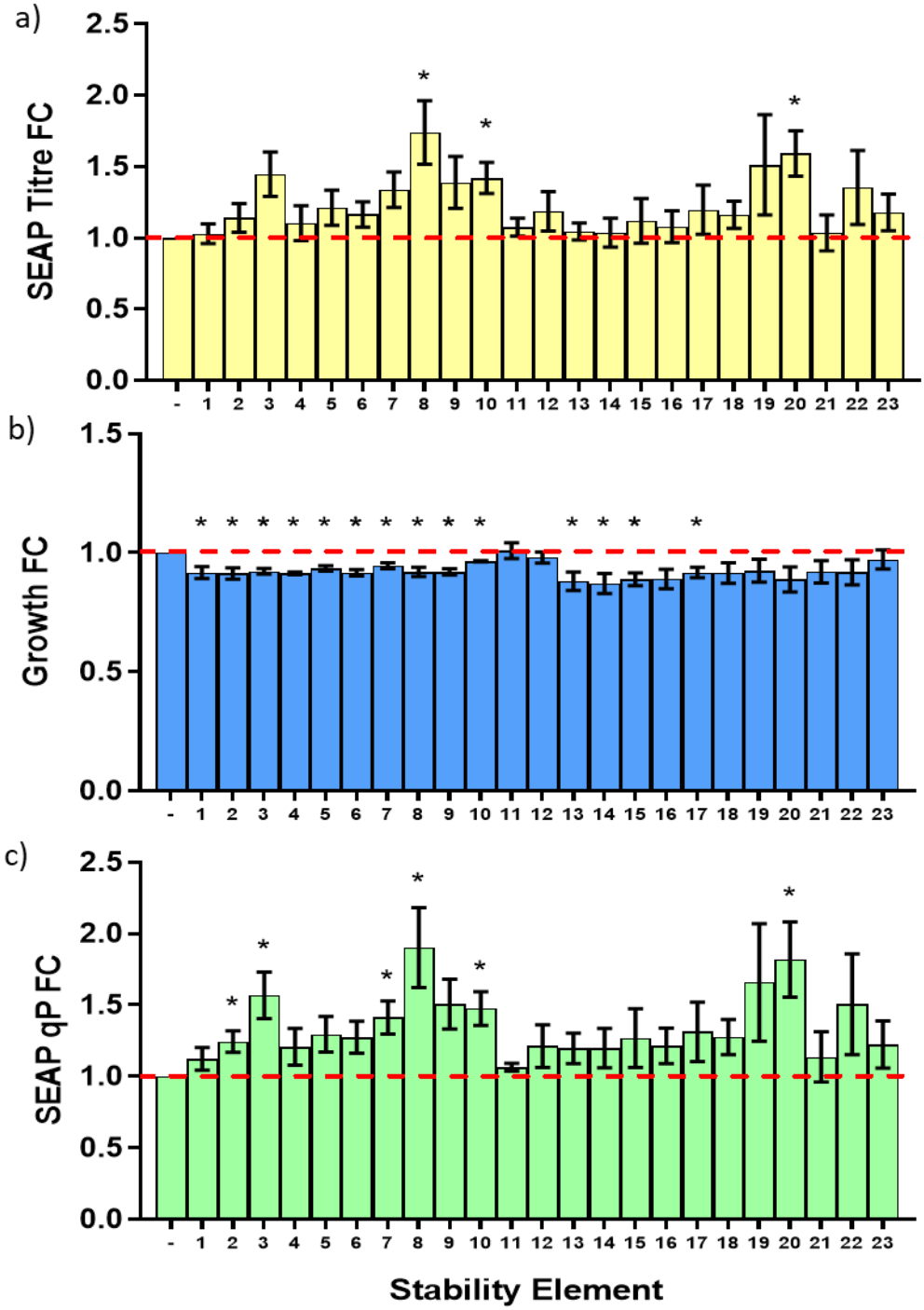


Figure 4.5 – Transient screening of the TU1-SEAP stability element library, by transfection with PEI at a 1:5 DNA:PEI ratio into the CHO transient host in 96 deep-well plates. A) Three stability elements were associated with a significant (P<0.05) increase in SEAP titre: SE8, 10, and 20 imparting 1.74-fold, 1.42-fold, and 1.59-fold increases in titre respectively. B) Expression of 14 stability elements (SE1-10,13-15,17) was associated with a slight, but significant (P<0.05) decrease in cell growth, to an average of 0.92-fold. C) Expression of 6 stability elements gave rise to a significant (P<0.05) increase in SEAP specific productivity: SE2 by 1.24-fol, SE3 by 1.57-fold, SE7 by 1.41-fold, SE8 by 1.90-fold, SE10 by 1.48-fold, and SE20 by 1.82-fold. The sequence of all stability elements referenced can be found in section 2.13.3. and Table 1 of this thesis. Results presented are of technical duplicates, from three biological replicates. FC: fold-change; qP: specific productivity.

To validate the ability of candidate stability elements to increase SEAP titre, the four stability elements associated with the largest significant increases in specific productivity in Figure were selected. TU1-*SEAP* plasmids bearing these stability elements were transfected with PEI at a 1:5 DNA:PEI ratio into the CHO transient host at $1*10^6$ cells/ml in 10ml culture in cultiflasks. VCD, cell viability, and SEAP titre of each culture was measured 72h and 120h post-transfection, by ViCell as in section 2.4. and Sensolyte assay as in section 2.6.2., respectively. As displayed in Figure 4.6a and Figure 4.6b, no significant changes in cell culture performance, reflected in either culture's IVCD or viability, were measured between cells transfected with different stability elements. As shown in Figure 4.6c and Figure d, cells transfected with SEAP-SE20 exhibited 1.62-fold higher titre than the SEAP control 72h post-transfection, and a 1.28-fold increase 120h post-transfection. SEAP-SE10 was also associated with a significant 0.72-fold decrease in SEAP titre 120h post-transfection. Figure e shows that cells transfected with SEAP-SE20 exhibited a 1.45-fold increase in SEAP specific productivity compared to the SEAP control, across the 120h culture length. SEAP-SE3 and SEAP-SE10 in contrast, were associated with a reduction in SEAP specific productivity across their cultures by 0.81-fold and 0.74-fold, respectively. Finally, as shown in Figure f, the titre fold change associated with each stability element, compared to the unmodified SEAP control, decreased from 72h to 120h in a consistent fashion: SE20 by 21%, SE8 by 23%, SE3 by 18%, and SE10 by 17%. These data validate the ability of SE20 to increase transient titre of SEAP, whilst raising questions about the mechanistic effect which could reduce the titre fold change of each element as culture progresses.

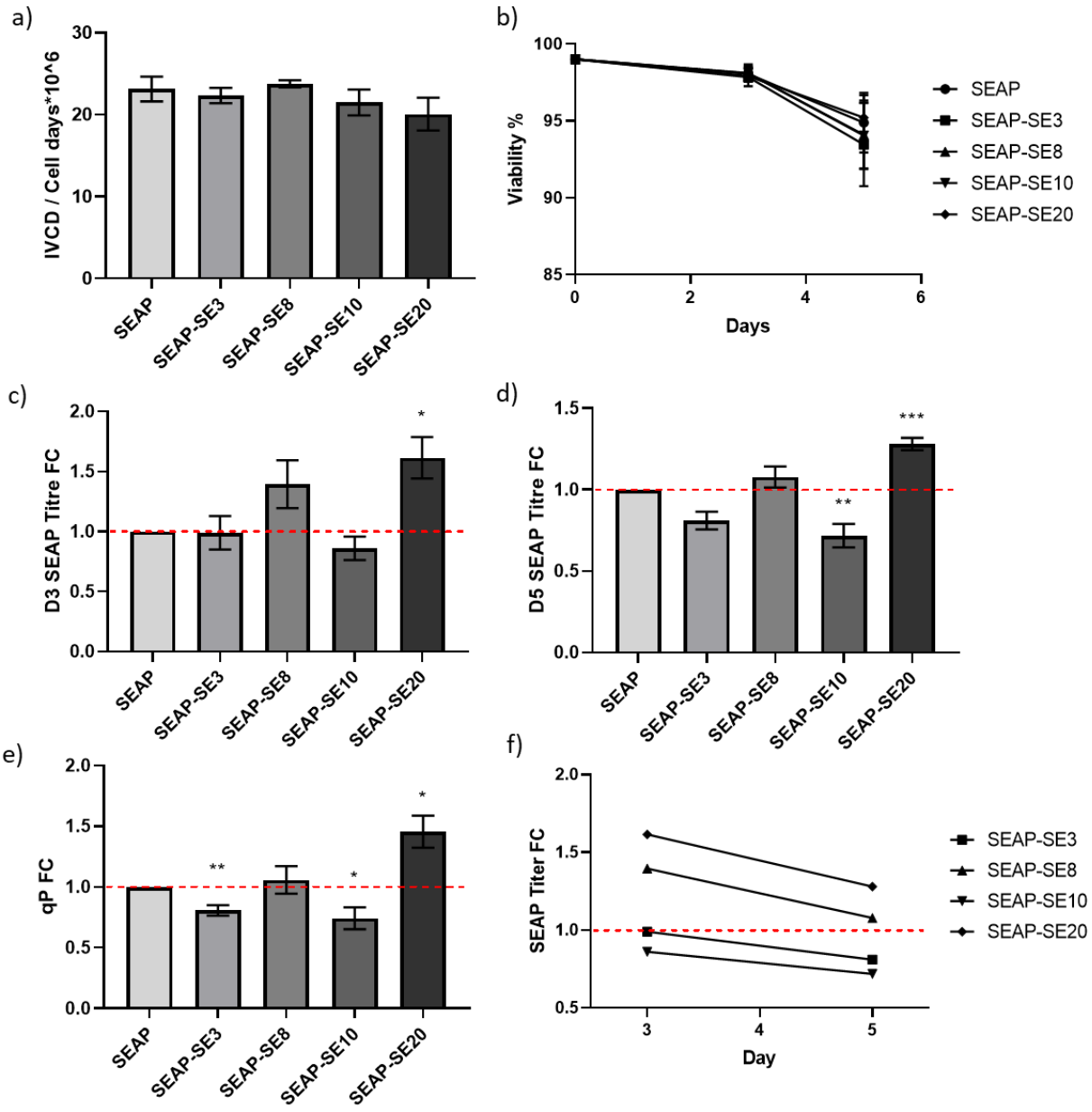


*Figure 4.6 – Transient analysis of candidate SEAP-SE constructs by transfection with PEI at a 1:5 DNA:PEI ratio into the CHO transient host in cultiflasks. A) No significant difference was measured in IVCD across the candidate library, with values ranging from 20.1*10⁶-23.1*10⁶ cell days, across 120h culture. B) No significant difference was measured in viability across the candidate library, with values ranging from 97.9-98.1% 72h post-transfection, and 93.5-95.2% 120h post-transfection. C) Expression of SE20 is associated with a significant (P=0.012) 1.62-fold increase in titre 72h post transfection, with SE8 associated with a non-significant (P=0.095) 1.40-fold increase. D) Expression of SE20 is associated with a significant (P=0.0004) 1.28-fold increase in SEAP titre 120h post-transfection, whereas SE10 significantly (P=0.0076) decreases titre by 0.72-fold. E) Expression of SE20 is associated with a significant (P=0.014) 1.45-fold increase in SEAP specific productivity, whereas SE3 is associated with a 0.81-fold decrease (P=0.0043), and SE10 a 0.74-fold decrease (P=0.029). F) The titre fold change of every SEAP-SE construct decreases relative to unmodified SEAP titre from 72h to 120h post-transfection. The sequence of all stability elements referenced can be found in section 2.13.3. and Table 1 of this thesis. Results presented are of technical duplicates from three biological replicates. SE: stability element; FC: fold-change; qP: specific productivity.*

VCD measurements during this experiment taken on culture days 0, 3, and 5, are shown in Figure . These data indicate that cells were in log phase from day 0 to 3, roughly doubling in density every day, before transitioning to stationary phase between day 3 and day 5, with density increasing little, and even decreasing in the case of cells transfected with SEAP-SE8.

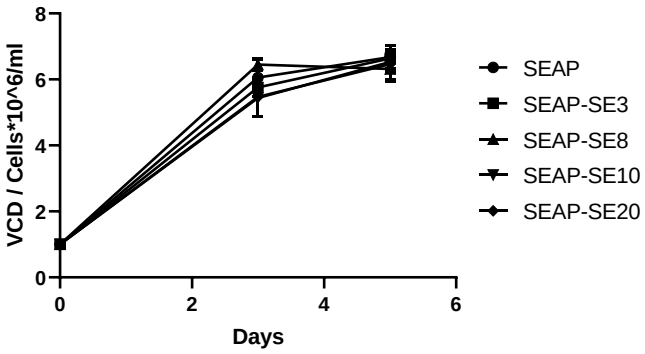


Figure 4.7 - Viable cell densities of SEAP-3'SE cultiflask screen. Cell cultures appear to transition from growth to stationary phase around day 3 post-transfection. Results presented are of technical duplicates from three biological replicates. VCD: viable cell density.

4.2.5. The addition of SE20 is associated with an increase in SEAP mRNA half-life

To investigate the mechanistic basis of the stimulation of SEAP titre by SE20, transient host CHO cells at $1*10^6$ cells/ml in 10ml culture in cultiflasks were transfected with TU1-*SEAP* and TU1-*SEAP*-SE20 using PEI at a 1:5 DNA:PEI ratio. 48h post transfection, pellets were spun down of each transfection for analysis by RT-qPCR as in section 2.9.2., with the relative abundance of *SEAP* mRNA measured compared to *fkbp1a*, as recommended in Brown *et al.*, 2018. The results of this experiment are shown in Figure a. No significant difference in abundance is measured between *SEAP* and *SEAP*-SE20. Immediately following the extraction of these samples, cell cultures were treated with 20μM Actinomycin D, in order to inhibit their transcription, as in section 2.12. Following this treatment, cell pellets were similarly sampled, at timepoints of 1h, 2h, 4h, 8h, and 24h post-treatment. These samples were analysed by RT-qPCR for relative abundance of *SEAP* mRNA, compared to *fkbp1a*. The results of this experiment are displayed in Figure b, plotting the relative abundance of each *SEAP* mRNA against the timepoint at which it was measured. Each set of points is fitted with a one-phase exponential decay curve, of which the curve for *SEAP* exhibits a noticeably faster drop-off in abundance. The half-life of each mRNA, as derived from this fitted curve, is shown in Figure c, with *SEAP*-SE20 exhibiting a half-life of 4.04h, compared to 0.68h for *SEAP*.

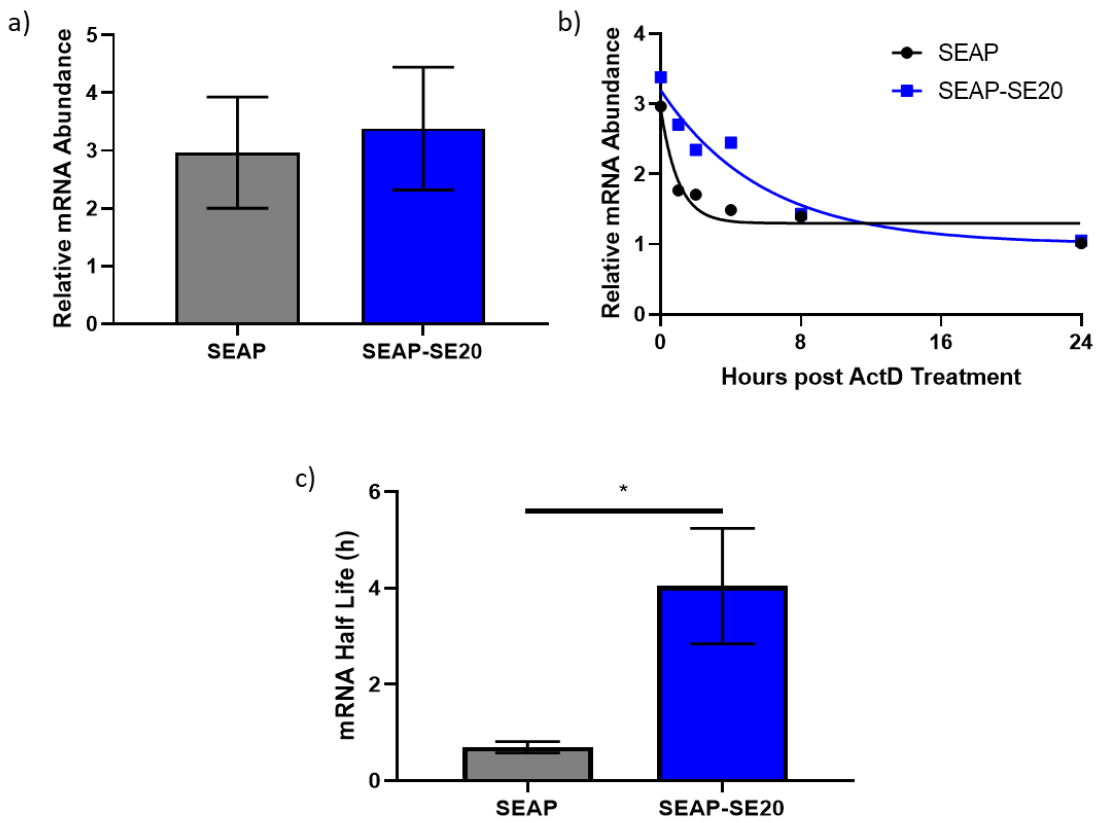


Figure 4.8 – The addition of SE20 is associated with extended half-life of SEAP mRNA when transfected by PEI at a 1:5 DNA:PEI ratio into the CHO transient host in cultiflasks. A) No significant difference is measured in the abundance of SEAP and SEAP-SE20 mRNA pre-ActD treatment, with relative abundances of 2.96 and 3.38, respectively. βB) Abundance of each SEAP transcript was measured 1h, 2h, 4h, 8h, and 24h post-ActD treatment, and fitted by a one-phase exponential decay curve. C) SEAP-SE20 mRNA exhibits a significantly (P=0.032) longer half-life than SEAP, at 4.04h compared to 0.68h. The sequence of SE20 can be found in section 2.13.3. and Table 1 of this thesis. Results presented are of technical triplicates, from three biological replicates.SE: stability element.

4.3. Discussion

Comparative analysis of mammalian 3' triple helices for their effect on SEAP expression resulted in a severe, though not total, reduction in titre. Whether or not this result is unexpected is difficult to disentangle. Significant upregulation of mRNA levels upon addition of MALAT1/MENβ before the PolyA signal of intronless *β-globin* has been demonstrated (Brown *et al.,* 2012), but this data may be misleading, as *β-globin* is almost uniquely extremely affected by intron removal, reducing expression ~35-fold (Lu and Cullen, 2003). This effect has not been shown with a WT *β-globin* transcript and may not be replicated in a transcript less sensitive to intronless expression, such as SEAP.

The lack of annotated CHO triple helices may also have led to incomplete or improper sequence being used, as alignment of the *H.sapiens* triple helices against the CHO genome may miss non-conserved regions essential for CHO function. Triple helices do show a measure of sequence variation between species (Tycowski *et al.,* 2016), which is slightly diminished between mammalian genomes (Wilusz *et al.,* 2012), but may be exacerbated by the genetic instability of immortalised CHO cell lines. Evidence for their correct assembly may be seen in the only partial rescue of expression by downstream insertion of the SV40 PolyA in both cases. It can be inferred from expression levels of SEAP-PAN-pA, and by the stability element screen, that insertion of sequence upstream of the PolyA in the 3'UTR is not deleterious to expression unless the sequence in question carries an actively deleterious effect. Therefore, the MALAT1 and MENβ sequences used in this study are unlikely to be entirely non-functional, though a partial inhibition of function remains possible.

Analysis of SEAP-PAN-pA gave an initially promising result, delivering a higher, though non-significantly so, titre and specific productivity than SEAP controls, when screened in 96 deep-well plates. However, the hypothesis that PAN-pA increases SEAP titre was conclusively proven incorrect by screening in scaled-up cultiflask culture, where no significant changes in cell culture performance or SEAP titre were observed compared to controls. This false positive can therefore be attributed to the higher error associated with high-throughput experimental platforms, where small pipetting errors in cell culture and transfection mixes become proportionally more influential, and consistent conditions may not be upheld across a multi-well plate (to be discussed further below). Therefore, any ambiguous

results and conclusions gained from high-throughput testing, should be verified in alternative culture platforms. Collectively, these data suggest that 3' triple helices are not an appropriate tool for predictably enhancing biopharmaceutical expression by improving mRNA stability.

Screening of 3' stability elements in 96 deep-well plates led to a significant increase in SEAP titre with three constructs. Furthermore, some slight, but significant reductions in cell growth imply 6 constructs granted a significant increase in SEAP specific productivity. However, these reductions in cell growth appear to follow a regular periodicity every 12 constructs. This is likely the result of different conditions at the edge of plates, arising either from the cell culture platform (edge wells exhibiting higher cell growth), or the growth assay (edge wells giving a higher Prestoblue signal). If the effect can be attributed to the former mechanism, the implied increases in specific productivity would be accurate, but if it is the latter, they may be misleading. As stated above, in such ambiguous cases, results should be verified in an alternative culture platform.

When verified in a scaled-up culture platform, one stability element (SE20) granted a significant increase in day 3 titre, day 5 titre, and specific productivity across the culture length, validating its' ability to increase SEAP expression. No significant changes in IVCD or viability were measured, corroborating the hypothesis that differences in growth measured in the 96 deep-well plate screen were artefacts of an edge effect. Another observation from this screen was that the titre of all SEAP constructs bearing a stability element decreased compared to the SEAP control from day 3 to day 5. As the elements used in this study were identified in a descriptive, rather than prescriptive manner, identifying sequences enriched in highly-expressed mRNA without describing their mechanism or synthetically designing their function, it is difficult to infer a mechanistic explanation for this decrease. However, the data presented in Figure , showing that this decrease in titre is accompanied by a halt in cell growth may imply that these stability elements are more effective when cell cultures are in their growth, as opposed to stationary phase.

Whilst a mechanistic basis of SE20's stimulation of SEAP titre cannot be directly drawn from literature; some inferences can be made from the results presented in Figure 4.7. These data show that addition of SE20 to the 3'UTR of a *SEAP* is associated with a significant increase in mRNA half-life. This result is plausibly consistent with literature, wherein it is shown that

3'UTR sequences have a significant influence over mRNA half-life, via mechanisms such as occlusion of RNA-binding protein landing sites and miRNA target sites, and creation of optimal distances between the stop codon at PolyA signal (Berkovits and Mayr, 2015; Mishima and Tomari, 2016; Oikonomou *et al.,* 2014; Sandberg *et al.,* 2008a; Wu and Bartel, 2017).

Despite these feasible explanations, there are no readily detectable sequence features which explain why SE20 performs this function, as opposed to any other stability element. The stability elements used in this thesis tend to be A/T-rich, (64.7% A/T over the entire library with a standard deviation of 8.7%), but SE20 is not remarkable in this metric, with 70.6% A/T content. Moreover, no significant correlation is found between stability element A/T content and effect on SEAP titre or specific productivity, as measured in Figure . A/T-rich tracts in the 3'UTR conferring effects on mRNA stability may imply a role for ARE-binding proteins. However, SE20 does not contain any of the motifs used to detect ARE-protein binding (Fallmann et al., 2016). No significant *C.griseus* miRNA sites of complementarity are found in SE20 by search in miRBase (Griffiths-Jones et al., 2008). Formation of secondary structures is another possible explanation; however, no significant secondary structure is found in SE20, with a minimum free energy of -2.70 kcal/mol predicted by the RNAfold web server, compared with -3.40 kcal/mol for the shuffled stability element control. Taking another descriptive approach, all stability elements which sufficiently activated SEAP expression to be taken forward to the validation screen (SE3, 8, 10, 20) were analysed for conserved motifs using the DMINDA 2.0 web server (Yang et al., 2017), and a conserved TTGNAA motif was found in all four sequences. However, this motif is also found in SE4 (TTGGAA) and SE12 (TTGCAA), both of which failed to significantly increase SEAP titre. Finally, no regions of alignment are found between SE20 and the TU1-*SEAP* vector from which it was being expressed.

Despite an increase in mRNA half-life, no significant difference in relative abundance was measured in *SEAP*-SE20, compared to the *SEAP* control. As the rate of cytoplasmic turnover of *SEAP*-SE20 is lower than *SEAP* alone, its rate of generation must therefore be diminished as well. Whilst it is unlikely that transcription would be significantly affected by an addition of 34bp to a transcript, the processing and export of mRNA can be affected by 3'UTR structures (Heinrich *et al.,* 2017), in which case the titre stimulation granted by SE20 could be

attributed to higher translation efficiency; and a higher rate of translation could itself provide an explanation for the extension of half-life (Bicknell and Ricci, 2017). Alternatively, a long-lived pool of recombinant mRNA could trigger a feedback mechanism to reduce transcription, in which case increases in titre could be attributed toward a more energetically efficient solution for the cell: investing less in transcription, and more in translation.

More work to understand these mechanisms would need to be performed in order to design a stability element fully synthetically, render its' effect titratable, or even answer questions such as whether different elements may be needed for different vectors or recombinant proteins. Nonetheless, the evidence presented suggests that stability elements, specifically SE20, could represent a predictable molecular biology tool for the stimulation of recombinant protein production. Increasing the attractiveness of this novel tool is the validated mechanism of extension of mRNA half-life, and the ease of incorporated a short vector sequence (34bp) into a vector. More generally, these data suggest the potential power in the application of sequences discovered via MPRA platforms to biopharmaceutical expression.

4.4. Conclusion

In this chapter, I have presented two different proposed solutions to increase mRNA stability in recombinant protein production, both based on insertion of sequence elements into the 3'UTR of recombinant transcripts. The first proposal of 3' triple helices failed to meet the standards of such a molecular biology tool; the mammalian structures having a detrimental effect on titre, and no effect being measured from the viral element. The second solution, short mRNA stability elements, present a novel method to increase recombinant protein production. SE20 was shown to increase the specific productivity of transiently expressed SEAP by 1.45-fold over a 120h culture, and it's mechanism validated, as it was associated with an extension of SEAP mRNA half-life from 0.68-4.04h. Whilst further investigation and design will be required to render this tool truly synthetic, user-defined, and titratable, the data presented in this chapter show it as a predictable enhancer of recombinant protein titre. Furthermore, the relative simplicity of incorporating a 34bp element into a vector render it an attractive solution for multiple rounds of testing, when transiently optimising a vector for stable production. Combined with results of a multi-element screen to be presented in Chapter

6 of this thesis, stability elements, specifically SE20, appear to represent an effective tool for the control of mRNA stability, and thereby protein titre, in biopharmaceutical production.

4.5. References

Berkovits, B.D., and Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature 522, 363–367.

Bicknell, A.A., and Ricci, E.P. (2017). When mRNA translation meets decay. Biochem. Soc. Trans. 45, 339–351.

Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2012). Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENβ noncoding RNAs. Proc. Natl. Acad. Sci. U. S. A. 109, 19202–19207.

Brown, J.A., Bulkley, D., Wang, J., Valenstein, M.L., Yario, T.A., Steitz, T.A., and Steitz, J.A. (2014). Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. Nat. Struct. Mol. Biol. 21, 633–640.

Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. Elife 7, 1–32.

Conrad, N.K. (2014). The emerging role of triple helices in RNA biology. Wiley Interdiscip. Rev. RNA 5, 15–29.

Fallmann, J., Sedlyarov, V., Tanzer, A., Kovarik, P., and Hofacker, I.L. (2016). AREsite2: An enhanced database for the comprehensive investigation of AU/GU/U-rich elements. Nucleic Acids Res. 44, D90–D95.

Fan, X.C., and Steitz, J.A. (1998). Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. EMBO J. 17, 3448–3460.

Fukao, A., and Fujiwara, T. (2017). The coupled and uncoupled mechanisms by which trans-acting factors regulate mRNA stability and translation. J. Biochem. 161, 309–314.

Gentilella, A., Morón-Duran, F.D., Fuentes, P., Zweig-Rocha, G., Riaño-Canalias, F., Pelletier, J., Ruiz, M., Turón, G., Castaño, J., Tauler, A., et al. (2017). Autogenous Control of 5′TOP mRNA Stability by 40S Ribosomes. Mol. Cell 67, 55-70.e4.

Goodarzi, H., Najafabadi, H.S., Oikonomou, P., Greco, T.M., Fish, L., Salavati, R., Cristea, I.M., and Tavazoie, S. (2012). Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature 485, 264–268.

Griffiths-Jones, S., Saini, H.K., Van Dongen, S., and Enright, A.J. (2008). miRBase: Tools for microRNA genomics. Nucleic Acids Res. 36, 154–158.

Heinrich, S., Sidler, C.L., Azzalin, C.M., and Weis, K. (2017). Stem–loop RNA labeling can affect nuclear and cytoplasmic mRNA processing. RNA 23, 134–141.

Hung, F., Deng, L., Ravnikar, P., Condon, R., Li, B., Do, L., Saha, D., Tsao, Y.S., Merchant, A., Liu, Z., et al. (2010). mRNA stability and antibody production in CHO cells: Improvement through gene optimization. Biotechnol. J. 5, 393–401.

Lu, S., and Cullen, B.R. (2003). Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. RNA 618–630.

Mason, M., Sweeney, B., Cain, K., Stephens, P., and Sharfstein, S.T. (2012). Identifying bottlenecks in transient and stable production of recombinant monoclonal-antibody sequence variants in chinese hamster ovary cells. Biotechnol. Prog. 28, 846–855.

Mishima, Y., and Tomari, Y. (2016). Codon Usage and 3′ UTR Length Determine Maternal mRNA Stability in Zebrafish. Mol. Cell 61, 874–885.

Mitton-Fry, R.M., DeGregorio, S.J., Wang, J., Steitz, T.A., and Steitz, J.A. (2010). Poly(A) Tail Recognition by a Viral RNA Element Through Assembly of a Triple Helix. Science (80-.). 330, 1244–1247.

Oikonomou, P., Goodarzi, H., and Tavazoie, S. (2014). Systematic identification of regulatory elements in conserved 3′ UTRs of human transcripts. Cell Rep. 7, 281–292.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. Cell 160, 1111–1124.

Radhakrishnan, A., and Green, R. (2016). Connections Underlying Translation and mRNA Stability. J. Mol. Biol. 428, 3558–3564.

Rodnina, M. V. (2016). The ribosome in action: Tuning of translational efficiency and protein folding. Protein Sci. 1390–1406.

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008a). Proliferating Cells Express mRNAs with Shortened 3′ Untranslated Regions and Fewer MicroRNA Target Sites. Science (80-.). 320, 1643–1648.

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P. a, and Burge, C.B. (2008b). Proliferating cells express mRNAs with shortened 3′ UTRs and fewer microRNA target sites. Science (80-.). 320, 1643–1647.

Tycowski, K.T., Shu, M. Di, Borah, S., Shi, M., and Steitz, J.A. (2012). Conservation of a triple-helix-forming RNA stability element in noncoding and genomic rnas of diverse viruses. Cell Rep. 2, 26–32.

Tycowski, K.T., Shu, M. Di, and Steitz, J.A. (2016). Myriad Triple-Helix-Forming Structures in the Transposable Element RNAs of Plants and Fungi. Cell Rep. 15, 1266–1276.

Wilusz, J.E., Jnbaptiste, C.K., Lu, L.Y., Marzluff, W.F., Kuhn, C., Joshua-tor, L., and Sharp, P. a (2012). A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. Genes \& Dev. 26, 2392–2407.

Wu, X., and Bartel, D.P. (2017). Widespread Influence of 3′-End Structures on Mammalian mRNA Processing and Stability. Cell 169, 905-917.e11.

Wu, X., Lan, L., Wilson, D.M., Marquez, R.T., Tsao, W.C., Gao, P., Roy, A., Turner, B.A., McDonald, P., Tunge, J.A., et al. (2015). Identification and Validation of Novel Small Molecule Disruptors of HuR-mRNA Interaction. ACS Chem. Biol. 10, 1476–1484.

Yang, J., Chen, X., McDermaid, A., and Ma, Q. (2017). DMINDA 2.0: integrated and systematic views of regulatory DNA motif identification and analyses. Bioinformatics 33, 2586–2588.

5. 5′TOP Motifs and Translation Initiation

5.1. Introduction

5.1.1. 5′TOP Motifs for the Modulation of Titre by Translation Initiation Control

Besides the ubiquitous Kozak sequence, no synthetic, predictable and titratable molecular biology tools exist for controlling initiation, which is most often the rate-limiting step in translation (Aitken and Lorsch, 2012), and at which stage bottlenecks in expression are often observed (Rajendra *et al.*, 2015a). In this chapter, the use of 5′ Terminal Oligo Pyrimidine (5′TOP) motifs - short RNA sequences exhibiting translational control - as a molecular biology tool for controlling biopharmaceutical titre through translation initiation will be investigated and developed.

5′TOP motifs are found in mRNA, with each part of this name describing an obligatory feature of the motif. 5′ Terminal refers to the fact that TOP motifs must be placed precisely at the transcription start site, and therefore at the 5′ terminus of an mRNA (Eliseeva *et al.*, 2013; Philippe *et al.*, 2018), typically by the $TC_{+1}T$ core promoter motif, of which the C_{+1} is the obligatory start of the TOP motif (Parry *et al.*, 2010; Vo ngoc *et al.*, 2017). Following this C_{+1} is the Oligo Pyrimidine sequence, typically comprising 5-15 bases, Cs and Ts in DNA, Cs and Us in the resultant mRNA (Meyuhas and Kahan, 2015). Endogenous 5′TOP motifs are found in mRNA transcripts coding for proteins required for growth and proliferation, enabling regulation of their expression, and therefore cell growth and proliferation, in response to a number of factors, such as nutrient availability (Miloslavski *et al.*, 2014; Roux and Topisirovic, 2018; Yao *et al.*, 2017), growth factors (Markou *et al.*, 2010; Tuxworth *et al.*, 2008), and even viral infection (Hopkins *et al.*, 2015). 5′TOP motifs can be found in transcripts for various proteins, such as elongation factors, but are mostly found in ribosomal mRNAs (Fonseca *et al.*, 2018; Yamashita *et al.*, 2008).

5.1.2. MTORC1 Regulation

5′TOP motifs are subject to a subset of the more global translational regulation performed by *mTOR* (mammalian target of rapamycin), more specifically within MTORC1 (*mTOR* complex

1), with the protein factors *Raptor* and *mLST8* (Aylett *et al.,* 2016). MTORC1 is a central regulatory node, which integrates various growth signals, and coordinates cellular response, by phosphorylation of factors to upregulate anabolic processes, such as nucleotide and lipid synthesis, and downregulate catabolic processes such as lysosome biogenesis and autophagy (Ben-Sahra and Manning, 2017).

The sensing of these various stimuli by MTORC1 is primarily coordinated through a logical 'AND-gate', wherein full activation of the complex is achieved by recruitment to the lysosomal membrane by both growth factor-sensing *Rheb*, and the amino-acid sensing *Rag* factors (Buel and Blenis, 2016; Saxton and Sabatini, 2017). *Rheb* is a GTPase, which can only bind and activate MTORC1 in its *Rheb*-GTP form. The GTPase activating protein (GAP) *TSC2* (part of the TSC complex) promotes conversion of this form to *Rheb*-GDP, thus inhibiting MTORC1. The TSC complex is in turn inhibited via phosphorylation by factors such as *AKT*, *ERK*, and *RSK*, all of which are activated by kinase cascades started by binding of growth factors, oxygen and energy supply, and presence of glucose (Ben-Sahra and Manning, 2017; Nandagopal and Roux, 2015; Roux and Topisirovic, 2018). The four Rag proteins (*RagA*/*B*/*C*/*D*) form the obligate heterodimers *RagA*/*B* and *RagC*/*D*, which bind GTP and GDP respectively, and it is in these forms that they recruit and activate MTORC1. *GATOR1* is a GAP which acts on the *Rag*s to inhibit MTORC1 activation, and is in turn inhibited by the presence of cytosolic Arginine and Leucine, acting through *CASTOR1* and *Sestrin1*/*2* respectively (Yao *et al.,* 2017). *Rag*-MTORC1 association is upregulated by *Ragulator*, a guanine exchange factor (GEF) for the *Rag*s, which is in turn upregulated by cytosolic glutamine and lysosomal arginine, through *vATPase* and the arginine-sensing *SLC38A9* (Ben-Sahra and Manning, 2017; Buel and Blenis, 2016). This nutrient-sensing ability is more critical than growth factor-sensing, meaning that whilst both are required for full MTORC1 activation, nutrients alone are sufficient for a basal activation level, whereas growth factors alone fail to activate the complex (Valvezan and Manning, 2019).

Once activated, *mTOR* primarily controls global translation by activation of the *S6K*s, and inhibition of the *4E-BP*s, both by phosphorylation. The *S6K*s (*S6K1* and *S6K2* in mammals) regulate the phosphorylation state of multiple translation-associated proteins, such as *Rps6* and *eEF2*-kinase (*eEF2K*), to generally upregulate translation (Iadevaia *et al.,* 2014; Meyuhas,

2015). When unphosphorylated, *4E-BP*s bind and sequester *eIF4E*, partially inhibiting translation initiation, through disruption of PIC formation (Musa *et al.,* 2016). As discussed in section 1.4.3. of this thesis, not every factor is strictly obligatory for translation initiation, rather, initiation factors form a web of interconnected redundancies, from which particular transcripts may be more or less dependent on particular factors (Mayberry *et al.,* 2009). Thus, transcripts that contain 5'UTRs dependent on the presence of *eIF4E* are considered '*mTOR*-sensitive'. Identification of *mTOR*-sensitive transcripts is difficult, with issues of assay sensitivity and transcript abundance initially leading to the belief that only 5'TOP motifs were *mTOR*-sensitive (Masvidal *et al.,* 2017; Thoreen *et al.,* 2012), and a spurious supposed correlation between 5'UTR length and *mTOR* sensitivity (Leppek *et al.,* 2018; Thoreen *et al.,* 2012). Current analysis places *mTOR*-sensitive transcripts into two general categories: short transcripts, enriched in mitochondrial function, sensitive to *eIF4E*, but insensitive to *eIF4A*, and long transcripts, enriched in growth factors, sensitive to both *eIF4E* and *eIF4A* (Gandin *et al.,* 2016; Masvidal *et al.,* 2017).

5.1.3. 5'TOP Motifs and *LARP1*

In addition to these global mechanisms of translational control, transcripts containing 5'TOP motifs are subject to specific control by another MTORC1-regulated protein, *LARP1*, rendering them MTORC1-hypersensitive (Ben-Sahra and Manning, 2017; Hong *et al.,* 2017). Proteomic studies initially discovered that *LARP1* associates with 5'TOP transcripts in an MTORC1-dependent manner, and controls their translation (Tcherkezian *et al.,* 2014). Subsequently, X-ray crystallography has shown *LARP1* specifically binding polypyrimidine sequences via a repurposed and unique HEAT domain (Lahr *et al.,* 2015), and specifically binding both 5'cap analogues and 5'caps followed by a cytidine (Lahr *et al.,* 2017), explaining the need for 5'TOP motifs to be 5' terminal. This binding domain was shown to be sufficient to specifically regulate 5'TOP transcripts (Philippe *et al.,* 2018), and this binding is relieved upon *LARP1* phosphorylation by MTORC1 and *S6K*. Based on these observations, a model was developed in which *LARP1* inhibits translation initiation by outcompeting *eIF4F* for 5'cap binding, until MTORC1 activation, when *LARP1* is phosphorylated, the binding is released, and the inhibition relieved (Hong *et al.,* 2017). *LARP1* has also been shown to bind to the 3'UTR, via an association with *PABP* (Aoki *et al.,* 2013). When *mTOR* is inactive, and therefore *LARP1* is bound to 5'TOP sequences, this facilitates a circularisation of the mRNA, which is

then recruited to 40S ribosomes. This protects 5'TOP mRNA from degradation, in order to maintain a constantly available pool of growth-associated mRNAs, ready to be translated upon MTORC1 activation (Gentilella *et al.,* 2017). When MTORC1 is activated, *LARP1* dissociates from the 5'TOP motif, but remains bound to *PABP*, where it recruits active MTORC1 to the mRNA, further promoting translation (Hong *et al.,* 2017; Philippe *et al.,* 2018).

5.1.4. MTORC1 in CHO Cells

A number of studies have investigated the potential use of MTORC1 manipulation in CHO cells for recombinant protein expression. Dreesen and Fussenegger found that transient overexpression of *mTOR* increased the yield of both SEAP and an FC-fusion protein, both expressed transiently and stably. Stable *mTOR* overexpression was then added to a stable mAb-producing cell line, showing increased cell growth, viability, and specific productivity (Dreesen and Fussenegger, 2011). Studies by Dadehbeigi *et al.* and Courtes *et al.* both investigated the treatment of cell cultures with a low concentration of Rapamycin (an inhibitor of *mTOR*), using 20ng/ml and 100nM, respectively. Both studies found that by slightly compromising maximal growth by this method, culture could be extended, and overall titre of recombinant proteins improved by 50% and 16%, respectively. Conversely, both studies found that the more significant increase in maximum cell count, culture length and titre associated with the common practice of regularly feeding a batch culture (a fed-batch), may be partially creditable to the additional MTORC1 activation given by this nutrient supply (Courtes *et al.,* 2014; Dadehbeigi and Dickson, 2015). Josse *et al.* investigated MTORC1's effects on mAb yield from a number of differentially-expressing cell lines, by looking at relative abundance and phosphorylation of *eIF4E* and *4E-BP*. Though they were unable to produce any changes in titre by manipulating *4E-BP* phosphorylation, a correlation was observed between mAb titre and *eIF4E/4E-BP* stoichiometry (Josse *et al.,* 2016). Whilst some approaches to modulate global cell-line performance by manipulation of MTORC1 have been examined, no investigations have been made into the direct use of 5'TOP sequences in biopharmaceutical production.

5.2. Results

5.2.1. Constructing the 5'TOP library

In order to test their effect on biopharmaceutical expression when placed inside a recombinant mRNA transcript, a library of 5'TOP motifs was constructed. To rationally select motifs exhibiting high expression levels, an Omics dataset was utilised, firstly containing transcriptomic data gathered by RNA-seq from a CHO cell line stably expressing an ETE mAb (Geoghegan et al., 2018). Secondly, an MS-based proteomic dataset gathered inhouse by AstraZeneca, utilising pulsed SILAC, as in Schwanhäusser et al. 2011, from the same stably-expressing CHO cell line. For this dataset, cells were pulse-labelled with 'heavy' SILAC media, containing ^{13}C and ^{15}N-labelled L-arginine and L-lysine. Proteins were extracted, separated by SDS-polyacrylamide gel electrophoresis, and analysed by LC-MS/MS. The sum of peptide peak intensities were divided by the number of theoretically observable tryptic peptides to determine absolute protein amounts. The mean of three technical replicates was used. As a CHO transcription start site (TSS) database was not available, genes were extracted from this dataset, by cross-reference against a study identifying 5'TOP transcripts from *H.sapiens* and *M.musculus* TSS datasets, taking only genes positively identified in both organisms (Yamashita *et al.*, 2008). These putative CHO 5'TOP genes were then ranked by the following criteria in the Omics dataset:

- mRNA abundance, measured in FPKM.
- Specific translational activity, measured by the number of amino acids incorporated into each respective protein, per cell, per hour.
- Specific translation efficiency, given by controlling the specific translational activity of each gene by its' mRNA abundance.

Once 5'TOP genes were ranked in this way, the following iterative process was performed to compile the library of motifs:

1. Find the highest-ranked gene of each list in the CHO-K1 genome (2014).
2. Identify a sequence that satisfies the following criteria:
    a. Upstream of the start codon.
    b. Nearest to the start of the annotated gene.

c. Begins with a Cytidine.
d. Comprises an exclusive tract of 5+ pyrimidines.

3. If no such sequence can be identified, restart this process with the highest ranked gene of the next list.

Using this process, a library of 46 5'TOP sequences was compiled (TOP3-48), the gene of origin and sequence of which are shown in Table 2. An expression plasmid was constructed for these sequences, removing the CMV promoter and 5'UTR from the inhouse TU1-*SEAP* vector, and replacing them with an Ef1α promoter and 5'UTR, to create the TU1-Ef1α vector. This molecular cloning was performed by restriction/ligation between EcoRI/PpuMI sites as in section 2.3.6, and correct construction verified by Sanger sequencing. The Ef1α promoter contains a TCT motif at its transcription start site, followed by a 5'TOP sequence, which functions as the first control for this library (TOP1). Alternative 5'TOP sequences were therefore integrated by direct switching of TOP1 by the sequence in question. The second control was established by removal of the 5'TOP sequence from the vector (TOP2). Every 5'TOP motif used in this thesis can be found in section 2.13.4. and Table 5.1, and a representative map of the 'TU1-Ef1α' plasmid found in Appendix A. A schematic of the TU1-TOP-SEAP library used in this section are shown in Figure .

Figure 5.1 – A schematic of the TU1-TOP-SEAP library used in this thesis.

Source	Name	Sequence
EF1a	TOP1	CTTTTTC
NON-TOP	TOP2	
ACTB	TOP3	CTCTTTCTTC
RPLP2	TOP4	CTTCCTTTC
RPL11	TOP5	CCTCCT
RPS14	TOP6	CCTCCCCT
GAPDH	TOP7	CTCTCT
RPL13A	TOP8	CCTCCTCCTTTCCC
RPL14	TOP9	CCTTCTTCCTTCTC
RPS10	TOP10	CTTCTC
RPS25	TOP11	CTCCTCCC
RPS23	TOP12	CCTTCCT
RPS24	TOP13	CCTTCC
RPS7	TOP14	CCTCTTTCT
RPL31	TO*P15*	CTTCCCTTCCC
RPL18A	TOP16	CTTCCTTTT
RPS3A	TOP17	CTCCCC
RPS3A#2	TOP18	CCCTTTT
HSP90AB1	TOP19	CTTCTC
TUBB	TOP20	CCTTCCCTCCT
RPS27A	TOP21	CCTCTCTTCTC
VIM	TOP22	CCTCT
RPS8	TOP23	CCTTCCC
EEF1G	TOP24	CCTTTT
EEF2	TOP25	CTCTTCTCCT

Source	Name	Sequence
HSPD1	TOP26	CCCTCCC
IPO5	TOP27	CTCCCTCCTCCTTCTCTCTCTC
RPL7	TOP28	CTTCCTCTCTCT
RPL10	TOP29	CTTTTCCTCC
EIF2S3	TOP30	CCTTCCTCTCT
HNRNPF	TOP31	CCTCTTCCTCCTC
RPL3	TOP32	CCTCT
RPL30	TOP33	CTTCCTTTCT
RPS2	TOP34	CCTTCCCC
RPL23A	TOP35	CCTTTT
EEF2	TOP36	CTCTTCTCC
ENO2	TOP37	CTTTCTCCTTCCTCC
DDX21	TOP38	CTTTCTTCCTCTCTCTTTT
IPO7	TOP39	CTTCTCTTTCCTTTC
GARNL3	TOP40	CCTTTTTTTTTTTTCTC
LAMB1	TOP41	CCCCTTCCT
DDX39B	TOP42	CTCTTCT
GNAI3	TOP43	CCCCTCTCCC
CS	TOP44	CCCTTCCT
MATR3	TOP45	CCTCCTT
EIF2S3	TOP46	CCTTCCTCTCT
EIF3A	TOP47	CTTTCC
KPNA3	TOP48	CTCTTT
-	synTOP	CTTTCT
-	TOP36-1	CTCTTCTC

Table 2.1 –The 5'TOP motif sequences used in this study. 5'TOP motifs were substituted with the natural sequence from the Ef1α transcription start site in both the Ef1α and modified CMV-TCT promoters, by site-directed mutagenesis. The motif was removed from both of these promoters as a NON-TOP control.

5.2.2. Transient screening of the 5'TOP-SEAP library

The library of 5'TOP sequences described above in the TU1-Ef1α promoter were placed upstream of the *SEAP* CDS, creating the 5'TOP-*SEAP* library. This molecular cloning was performed by site-directed mutagenesis as in section 2.3.2., and sequence verified by Sanger sequencing. This library was transfected by electroporation into transient host CHO cells in 96 deep-well plate shaking culture, seeding at $0.2*10^6$ cells/ml. At 72 hours post-transfection, fold-change in growth and SEAP titre were measured by Prestoblue and Sensolyte assays, respectively, as in sections 2.4.4.2 and 2.6.2. Figure a shows that cells transfected with three 5'TOP sequences exhibited significantly increased titre compared to NON-TOP controls, with five significantly reduced. Surprisingly, removal of the natural Ef1α TOP (TOP1) was associated with a significant increase in titre. This picture is complicated by the data presented in Figure b, with significant variability observed in growth across the library. Ten 5'TOP sequences were associated with significantly lower growth than NON-TOP controls. However, significant deviation was observed across the screen, with an average SEM of 0.18 fold-change, greater than 0.30 fold-change in ten cases. As shown in Figure c, these data show a significant increase in SEAP specific productivity associated with 10 5'TOP sequences, compared to NON-TOP controls, up to a maximum of 2.32-fold with TOP42. These data are influenced by the highly variable results in Figure b, for example TOP26 moving from a non-significant 0.87-fold change in titre to a significant 2.04-fold increase in specific productivity, and TOP37 moving from a significant 0.68-fold reduction in titre to a significant 1.62-fold increase in specific productivity.

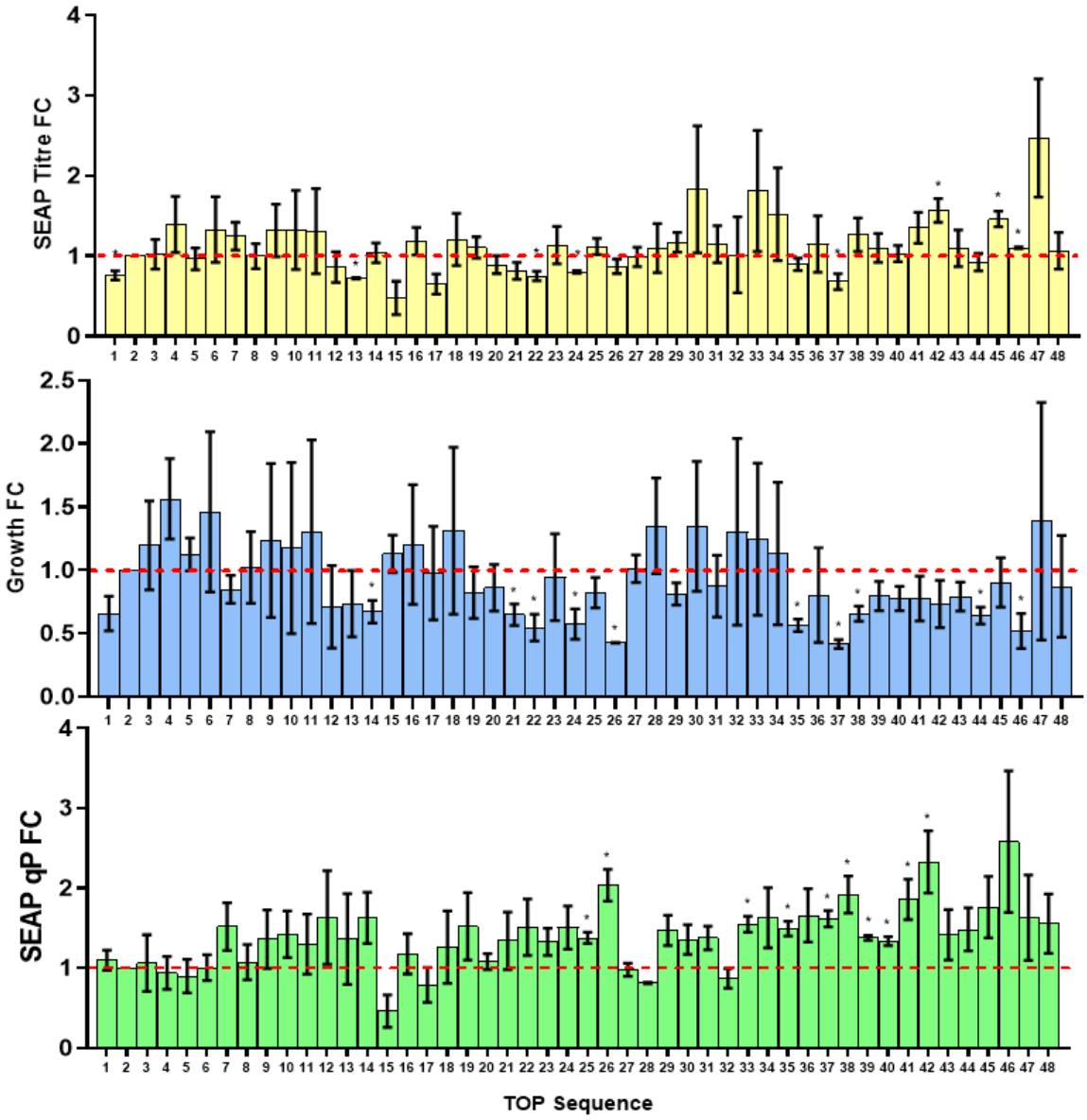


Figure 5.2 – Screening by electroporation of the 5'TOP-SEAP library in the CHO transient host, in 96 deep-well plates. A) Three sequences gave a significantly (P<0.05) higher SEAP titre than NON-TOP controls, TOP42 (1.57-fold), TOP45 (1.46-fold), and TOP46 (1.10-fold), whereas five sequences led to a significantly lower titre: TOP1, 13, 22, 24, and 37, to a minimum of 0.68-fold with TOP37. B) Ten sequences were associated with significantly (P<0.05) lower growth than NON-TOP controls: TOP14, 21, 22, 24, 26, 35, 37, 38, 44, and 46, to a minimum of 0.42-fold with TOP37. C) Ten sequences gave a significantly higher SEAP specific productivity than NON-TOP controls: TOP25, 26, 33, 35, 37, 38, 39, 40, 41, and 42, to a maximum of 2.33-fold with TOP42. All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1 of this thesis. Results presented are of technical duplicates from four biological replicates. FC: fold change; qP: specific productivity.

To investigate whether these disparities between titre and specific productivity were the genuine impact of 5'TOP motifs, or an artefact of a highly-variable transfection method, the 5'TOP-*SEAP* library was screened again with an alternative transfection method. The library was transfected by PEI at a 1:5 DNA:PEI ratio into transient host CHO cells at $1*10^6$ cells/ml in 96 deep-well plate shaking culture. Again, 72 hours post-transfection, fold change in growth was measured by Prestoblue assay and fold change in SEAP titre was measured by Sensolyte assay, as in sections 2.4.4.2., and 2.6.2. A wide range of SEAP titre can be observed in Figure a, from a 0.49-fold reduction associated with TOP8, to a 3.34-fold increase with TOP36. In a more expected result, cells transfected with the natural Ef1α TOP outperformed NON-TOP controls in titre, by 1.87-fold. The fold changes in growth presented in Figure b show substantially lower variation than those presented in Figure b, displaying a greater than 10-fold reduction in their average standard error of 0.014 fold-change, and ranging only between a 0.87-1.04 fold-change in growth. Similarly to data presented in Figure 4.5, the growth data displays a pattern of periodicity every 12 samples, in a manner concluded to be an edge effect of a multiwell plate, on either cell growth or Prestoblue assay. However, these variations make little difference, rendering only three fold-changes in specific productivity (Figure c) significant when they were not so in titre. Fold changes in specific productivity correlate significantly better with titre in Figure (R^2=0.87) compared to Figure 5.2 (R^2=0.09). As with titre, a significant range in specific productivity is observed, ranging from 0.50-fold in cells transfected with TOP8 to 3.96-fold with TOP36.

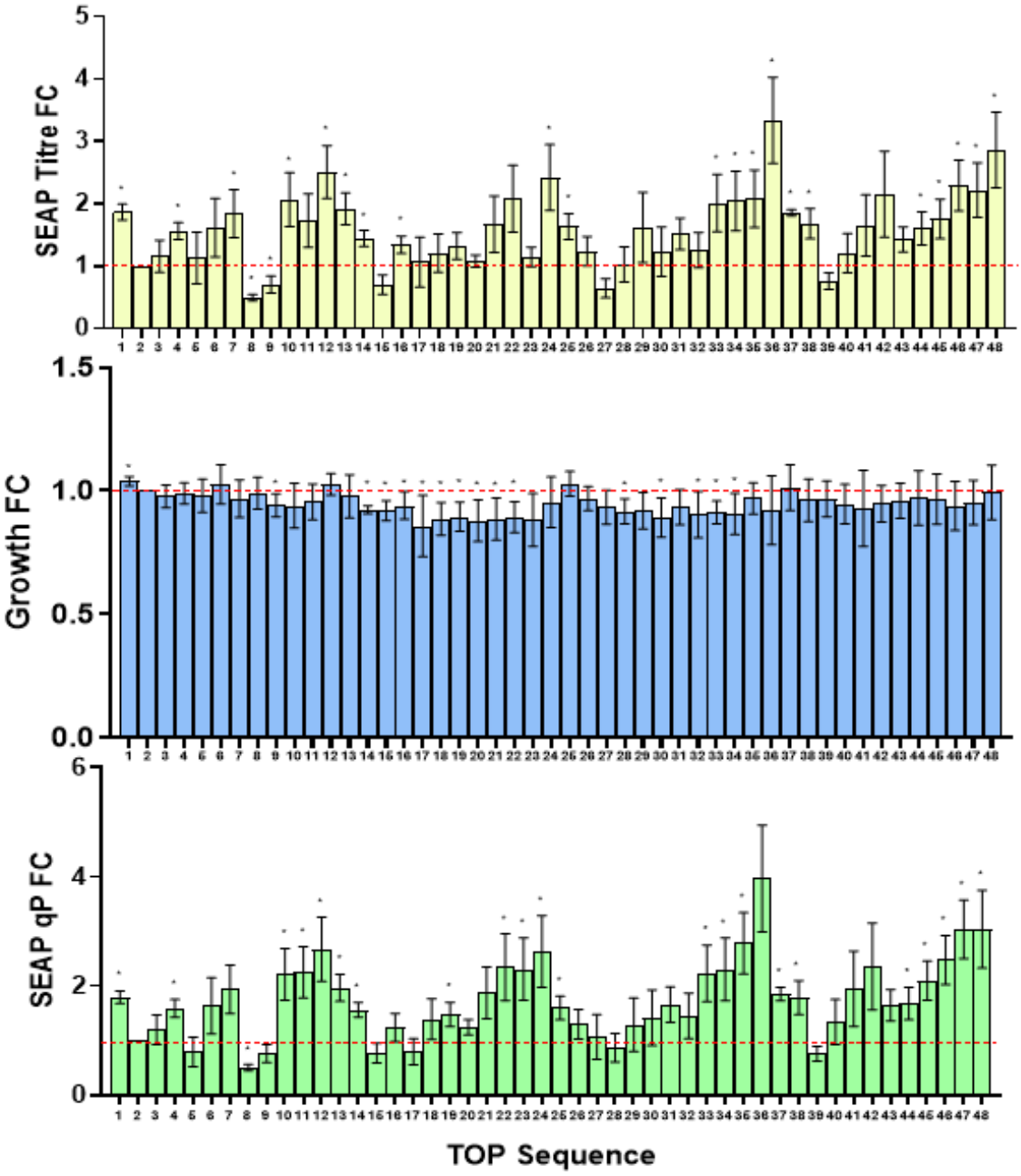


Figure 5.3 – Transfection by PEI at a 1:5 DNA:PEI ratio of the 5'TOP-SEAP library in the CHO transient host, in 96 deep-well plates. A) 21 sequences led to a significant (P<0.05) increase in SEAP titre, up to a maximum of 3.34-fold with TOP36, whereas two sequences gave a significant reduction in titre, TOP8 to 0.49-fold, and TOP9 to 0.71-fold. B) 16 sequences were associated with small, but significant (P<0.05) changes in cell growth, ranging from a 0.87-fold reduction with TOP16 to a 1.04-fold increase with TOP1. C) 23 sequences gave a significant (P<0.05) increase in SEAP specific productivity, to a maximum of 3.96-fold with TOP36, whereas only TOP8 at 0.50-fold gave a significant reduction. All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1 of this thesis. Results presented are of technical duplicates from four biological replicates. FC: fold change; qP: specific productivity.

To compare the expression activation of 5'TOP motifs between electroporation and PEI transfection techniques, the fold change imparted by the total library upon addition of a 5'TOP motif are displayed in Figure 5.4, with titre in Figure 5.4a and specific productivity in Figure 5.4b. As a group, 5'TOP-*SEAP* reporters were associated with significantly (P<0.0001) higher titre fold change compared to NON-TOP controls when transfected by PEI than by electroporation, with average fold-changes of 1.59-fold and 1.15-fold, respectively. A similar result is found for specific productivity, with a significant difference between the average 1.75-fold increase with PEI and 1.39-fold increase with electroporation. Collectively from this data, it can be concluded that 5'TOP motifs are capable of facilitating significant changes in recombinant protein expression characteristics, and that the PEI, as opposed to electroporation, transfection method gives both data with less variability, and significantly activates the expression of the 5'TOP sequences.

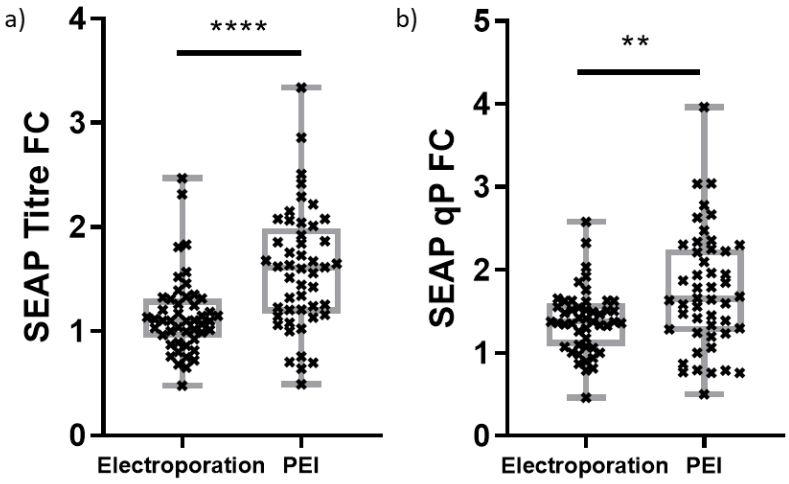


Figure 5.4 – Transfection by PEI activates expression of the 5'TOP library significantly more than electroporation. A) Fold change in titre of the library is significantly (P<0.0001) higher with PEI (1.59) than electroporation (1.15). B) Fold change in specific productivity is significantly (P=0.0030) higher with PEI (1.75) than electroporation (1.39). Results presented are of technical duplicates from four biological replicates. FC: fold change; qP: specific productivity.

Taking the data from Figure 5.3, a negative correlation was observed, between length of the 5'TOP motif in nucleotides and the expression characteristics of SEAP. This correlation with titre is displayed in Figure 5.5a, and with specific productivity in Figure 5.5b. The correlations may suggest that shorter 5'TOP sequences give greater expression activation than their longer counterparts.

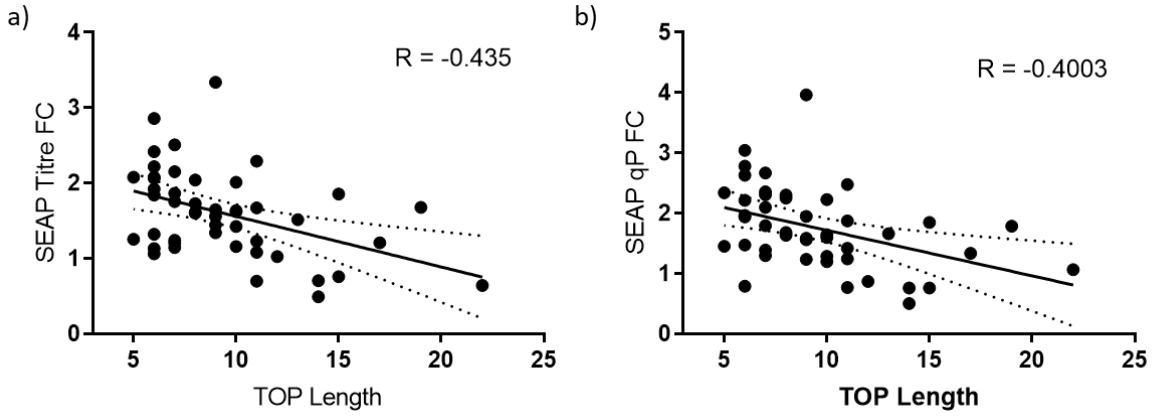


Figure 5.5 – Fold change in 5'TOP-SEAP expression compared to TOP2-SEAP upon PEI transfection into the CHO transient host in 96 deep-well plates is negatively correlated with 5'TOP length, reflected in both A) Titre, and B) Specific productivity. Results presented are of technical duplicates from four biological replicates. FC: fold change; qP: specific productivity.

5.2.3. Expression of 5'TOP-SEAP is not specifically enhanced by addition of cell culture feed or amino acids

To test whether expression of 5'TOP motifs could be specifically activated or repressed, TOP1-*SEAP* and TOP2-*SEAP* were transfected by PEI at a 1:5 DNA:PEI ratio, into the transient host CHO cell line at $1*10^6$ cells/ml, into 96 deep-well plates in shaking culture. At 24 hours post-transfection, these cells were treated with one of three titrated supplements: an amino acid mix of Leucine, Glutamine, and Arginine (the three amino acids most associated with MTORC1 activation (Yao *et al.*, 2017)) from 0-20mM, a commercial feed mix ranging from 0-20%, or rapamycin, ranging from 0-200ng/ml, as in section 2.11. At 72h post-transfection, fold change in cell growth was measured by Prestoblue assay as in section 2.4.4.2., and SEAP titre measured by Sensolyte assay as in section 2.6.2. To assess the activation of 5'TOP expression, and the effect this has on growth, the ratio of TOP1-SEAP/TOP2-SEAP titre and growth were calculated for each treatment and are shown in Figure 5.6. Local maxima and minima on these plots would indicate specific activation or repression of 5'TOP expression, through their acute MTORC1-mediated regulation. Figures 5.6a and 5.6b show no local maxima in TOP1/TOP2 ratio across either cell growth or SEAP titre. A slight downward trend is in fact observed, with a titre ratio of 1.50 with 0mM amino acids, compared with a ratio of 1.18 upon treatment with 20mM. A similar pattern is seen with feed supplementation, Figures 5.6c and 5.6d, where no significant local maxima are observed, and a negative correlation appears between feed titration and titre ratio, ranging from 1.51-fold at 0% to 1.14-fold at 20%. In contrast, a local minimum in TOP1/TOP2 titre ratio is seen in Figure 5.6e, moving from 1.29 at 12.5ng/ml rapamycin, to 0.58 at 25ng/ml, and back up to 1.06 at 50ng/ml, both of which are statistically significant changes. No such local minima are observed in growth ratio, presented in Figure 5.6f. These data imply that expression of 5'TOP-SEAP cannot be specifically activated by treatment with amino acids or feed solution, but may be specifically repressed by treatment with the correct concentration of rapamycin, at concentrations lower than is necessary for the global downregulation of translation given by rapamycin to take effect.

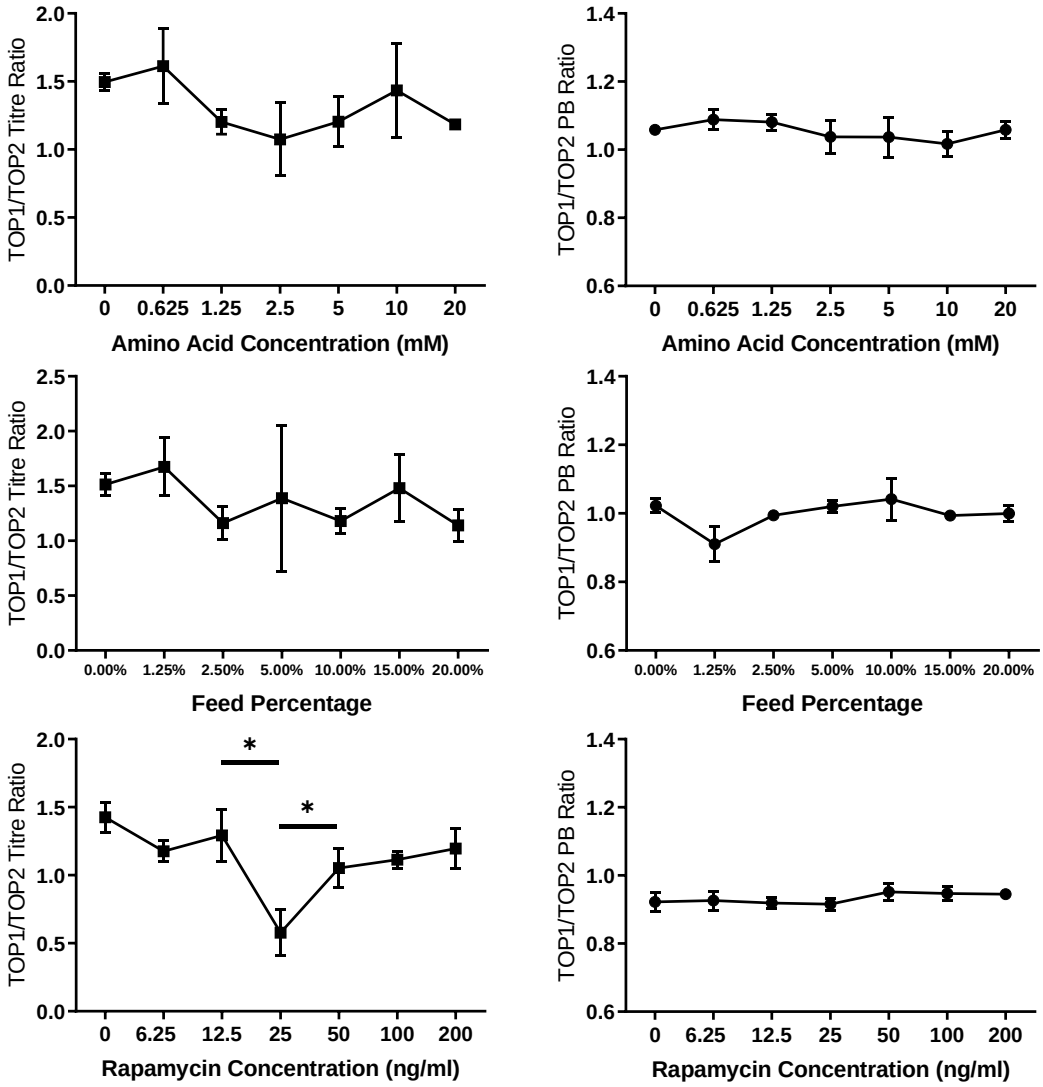


Figure 5.6 - Activation/repression of 5'TOP-SEAP expression in the CHO transient host by treatment with chemical supplements. No significant local maxima are observed in the TOP1-SEAP/TOP2-SEAP ratio upon amino acid titration in A) SEAP titre, or B) cell growth. No significant local maxima are observed in the TOP1-SEAP/TOP2-SEAP ratio upon feed titration in C) SEAP titre, or D) cell growth. E) A local minima in TOP1/TOP2 SEAP titre is observed upon treatment with 25ng/ml rapamycin, in a statistically significant decrease compared to 12.5ng/ml (P=0.047) and 50ng/ml (P=0.030) treatment. F) No significant local minima are observed in the TOP1/TOP2 growth ratio upon rapamycin titration. All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1 of this thesis. Results presented are of technical duplicates from three biological replicates. PB: prestoblue.

To validate that changes in expression characteristics of SEAP imparted by 5'TOP motifs were as a result of the expected mechanism of acute *LARP1*/MTORC1 inhibition, and therefore sensitive to rapamycin treatment, the transient CHO host was transfected with the 5'TOP-*SEAP* library by PEI at a 1:5 DNA:PEI ratio, at $1*10^6$ cells/ml, in 96 deep-well plate shaking culture. 24h post-transfection, all transfected cells were treated with 25ng/ml rapamycin, as in section 2.11. SEAP titre and growth fold change were measured 72h post-transfection, by Sensolyte and Prestoblue assay, as in sections 2.6.2. and 2.4.4.2. The fold changes in SEAP titre and specific productivity associated with transfection of the total TOP-*SEAP* library, compared to the NON-TOP control, are shown in Figure 5.7. A non-significant reduction in average titre is observed upon rapamycin treatment, from 1.60-fold to 1.43-fold. A significant reduction in average specific productivity is observed, from 1.76-fold to 1.42-fold. These results suggest that stimulation of SEAP expression by 5'TOP motifs is at least partially sensitive to rapamycin treatment.

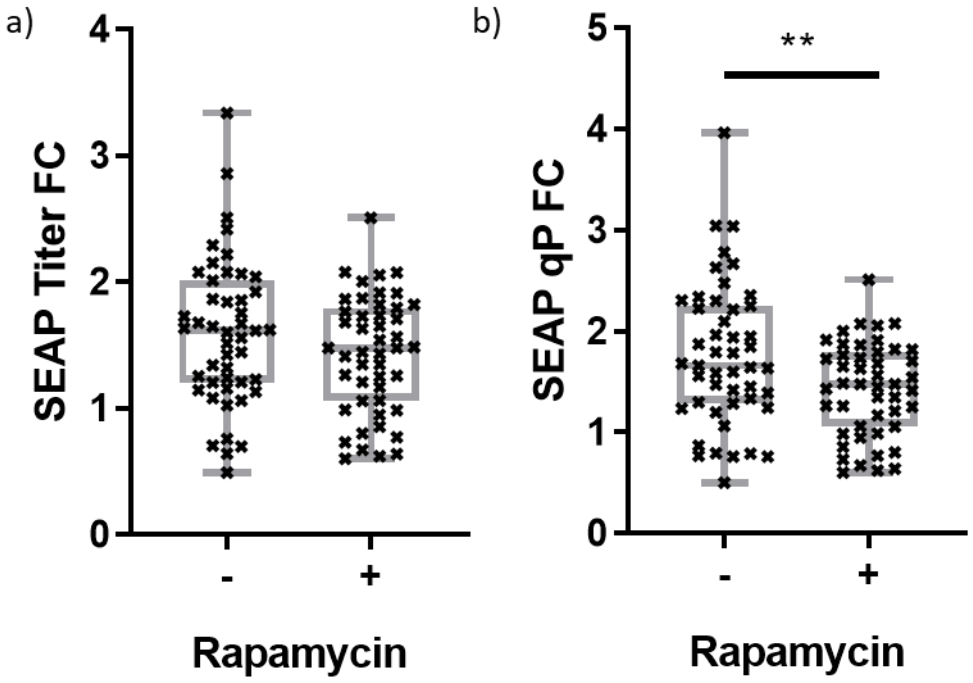


Figure 5.7 – Fold change of TOP-SEAP library expression relative to TOP2-SEAP upon treatment with 25ng/ml Rapamycin 24h post-transfection. A) A non-significant reduction in SEAP titre from 1.60 to 1.43-fold is observed. B) A significant (P=0.0079) reduction in specific productivity from 1.76 to 1.42-fold is observed. Results presented are of technical duplicates from four biological replicates. FC: fold-change; qP: specific productivity.

5.2.4. A library of synthetic promoters do not function as designed in expression of 5'TOP-SEAP

To begin expanding upon and developing the use of 5'TOP motifs to control recombinant protein expression, it was first decided to investigate whether they could be integrated with synthetic promoters, designed in Brown *et al.*, 2014, for both maximising, and predictably titrating expression. A total of seven proximal promoters were inserted upstream of the Ef1α core in the TU1-Ef1α-*SEAP* vector, in place of the Ef1α proximal promoter. Four had a user-defined strength of transcription (5RPU, 10RPU, 80RPU, 100RPU), two were based on tandem repeats of a TFRE (10X ERSE, 10X NFκB), and finally the CMV proximal promoter was used. This molecular cloning was performed by restriction/ligation between the EcoRI/AflII sites as in section 2.3.6., the synthetic promoters used found in section 2.13.5. and Table , and a representative plasmid map found under 'TU1-Ef1α' in Appendix A. A schematic of the Proximal Promoter-TOP-SEAP library used in this section is shown in Figure 5.8.

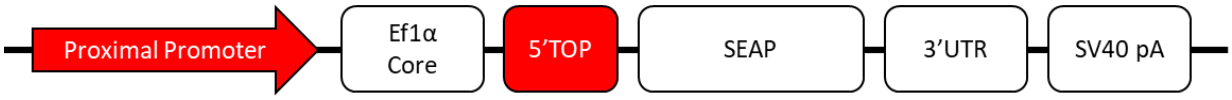


Figure 5.8 – A schematic of the Proximal Promoter-TOP-SEAP library used in this thesis.

Name	Sequence
P-CMV	AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA TGGCCCGCCTCGTGACCGCCCAACGACCCCCGCCCATTGACGTCAAT AATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACAT CAAGTGTATCATATGCCAAGTCCGGCCCCCTATTGACGTCAATGACG GTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGAC TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGG TGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGAC TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCC CGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT
100RPU	TGGGACTTTCCACCTTAGATGACACAGCAATCAGATTTGCTTGCGTGA GAAGATATAGGATGACACAGCAATCTAGACTGGGACTTTCCACTGAT

	ATTTTGCGCAATTGACCTAATGACACAGCAATAGTATGTGGGGCGGG GATCTAACTGGGACTTTCCAAAGGTCTTACCGGAAGTTGTTAGAATG ACACAGCAATGGATTCATATCCTGGGACTTTCCAGTATACTGCTTGCG TGAGAAGATGATCATGGGACTTTCCATGTACAAAAGGTC
80RPU	TGGGGCGGGGAAGTATGATGACACAGCAATTGATCATGGGACTTTCC ACTAGACTGCTTGCGTGAGAAGAAAGGTCTTACCGGAAGTTGACCTA ATGACACAGCAATGTTAGATGCTTGCGTGAGAAGACTGATATGGGAC TTTCCAGTATACTGGGGCGGGGATCTAACTGGGACTTTCCACAGATTA TGACACAGCAATTGTACAAAAGGT
20RPU	TTACCGGAAGTTGACCTATGCTTGCGTGAGAAGAGTTAGATGGGGCG GGGAAAGGTCTTTTGCGCAATTCAGATTTTACCGGAAGTTTATAGGAT GACACAGCAATTGTACAAAAGGTCTATATAAGCAGAGCTCGTTTAGT GAACCGTCAGATCGCCTAGATACGCCATCCACGCTGTTTTGACCTCCA TAGAAGAC
10RPU	TTACCGGAAGTTAAGGTCTTTTGCGCAATTCAGATTGACCTATTACCG GAAGTTTATAGGTGGGGCGGGGAGTTAGATTTTGCGCAATTTGATCAT TACCGGAAGTTTGTACAAAAGGT
5RPU	TATAGGAAGGTCTTACCGGAAGTTCCTTAGCTGATAGTATACCAGATT TTTTGCGCAATTCTAGACTGATCATCTAACGACCTATTACCGGAAGTT AGTATGTGTACAAAAGGT
10xERSE	ACCAATGGCCAGCCTCCACGAAAGGTCACCAATGGCCAGCCTCCAC GAAGTATGACCAATGGCCAGCCTCCACGACAGATTACCAATGGCCA GCCTCCACGACCTTAGACCAATGGCCAGCCTCCACGAGACCTAACCA ATGGCCAGCCTCCACGAGTATACACCAATGGCCAGCCTCCACGATCT AACACCAATGGCCAGCCTCCACGATGATCAACCAATGGCCAGCCTCC ACGACTAGACACCAATGGCCAGCCTCCACGAAGGT
10xNFκB	TGGGACTTTCCAAAGGTCTGGGACTTTCCAAGTATGTGGGACTTTCCA CAGATTTGGGACTTTCCACCTTAGTGGGACTTTCCAGACCTATGGGAC TTTCCAGTATACTGGGACTTTCCATCTAACTGGGACTTTCCATGATCAT GGGACTTTCCACTAGACTGGGACTTTCCAAGGT

Table 5.2 – The synthetic proximal promoters used in this study. Synthetic proximal promoters were taken from literature (Brown and James, 2015; Brown et al., 2014a, 2017), and substituted with the proximal promoters of both CMV and Ef1α, upstream of their cores, in the AstraZeneca in-house TU1 vector, using EcoRI/NheI and EcoRI/AflII restriction/ligation cloning for the CMV and Ef1α core promoters, respectively.

The 5'TOP motif was then removed from each construct, creating TOP1/NON-TOP variants of each proximal promoter construct. The level of transcription performed by the synthetic promoters could therefore be measured by comparing expression against cells transfected with unmodified Ef1α-*SEAP*, and correct transcription of 5'TOP motifs in these promoter assemblies by a difference in expression between TOP1/NON-TOP variants. These 14 constructs were then transfected into the CHO transient host at $1*10^6$ cells/ml using PEI at a 1:5 DNA:PEI ratio, in 24 shallow-well plate shaking culture, against a TU1-Ef1α-*SEAP* control. VCD was measured by Iprasense 72h and 120h post-transfection, as in section 2.4.4.3. SEAP titres were measured 120h post-transfection by Sensolyte assay as in section 2.6.2., shown as fold-change compared to TU1-Ef1α-*SEAP*, Figure 5.9a, and fold-change in specific productivity, Figure b. The CMV proximal promoter was associated with a significant decrease in SEAP titre of 0.45-fold, and no difference between TOP variants. The synthetic proximal promoters did not boost expression compared to unmodified Ef1α, did not perform relative to one another as designed, with cells transfected with 10RPU associated with the highest titre, and 80RPU outperforming 100RPU, which was not significantly different to 5RPU, and did not produce any significant difference between TOP variants. The 10x ERSE proximal promoter produced no significant change in titre compared to Ef1α-*SEAP*, and no difference between TOP variants. In contrast, a significant difference was observed between the TOP variants of the 10x NFκB proximal promoter, with its' TOP1 variant increasing SEAP titre by 1.58-fold and specific productivity by 1.76-fold. Cells were also pelleted 120h post-transfection, for measurement of relative *SEAP* mRNA abundance by RT-qPCR as in section 2.9.2., results shown in Figure c. The calculated fold-changes in translation efficiency, taken by controlling specific productivity by mRNA abundance, are displayed in Figure d. The relative abundance of *SEAP* mRNA correlates strongly with its' titre, the only difference being that a significant change in abundance is observed between the TOP variants under the 10x ERSE proximal promoter. The 10x NFκB TOP variants retain their significant difference in abundance, from 1.25 to 0.66-fold. Finally, the translational efficiency of every NON-TOP variant, apart from with the CMV proximal, is slightly, though not significantly, higher than their respective TOP1 variant. These results suggest that the synthetic promoters designed in Brown *et al.*, 2014 do not function as intended in the production of SEAP when paired with an Ef1α core promoter. A proximal promoter comprising of 10x NFκB TFREs may be compatible

with the Ef1α core, increasing SEAP expression, and maintaining proper terminal transcription of the 5'TOP sequence. Interestingly, the increase in SEAP titre given by the TOP1 variant appears mainly due to mRNA abundance, as opposed to translation efficiency.

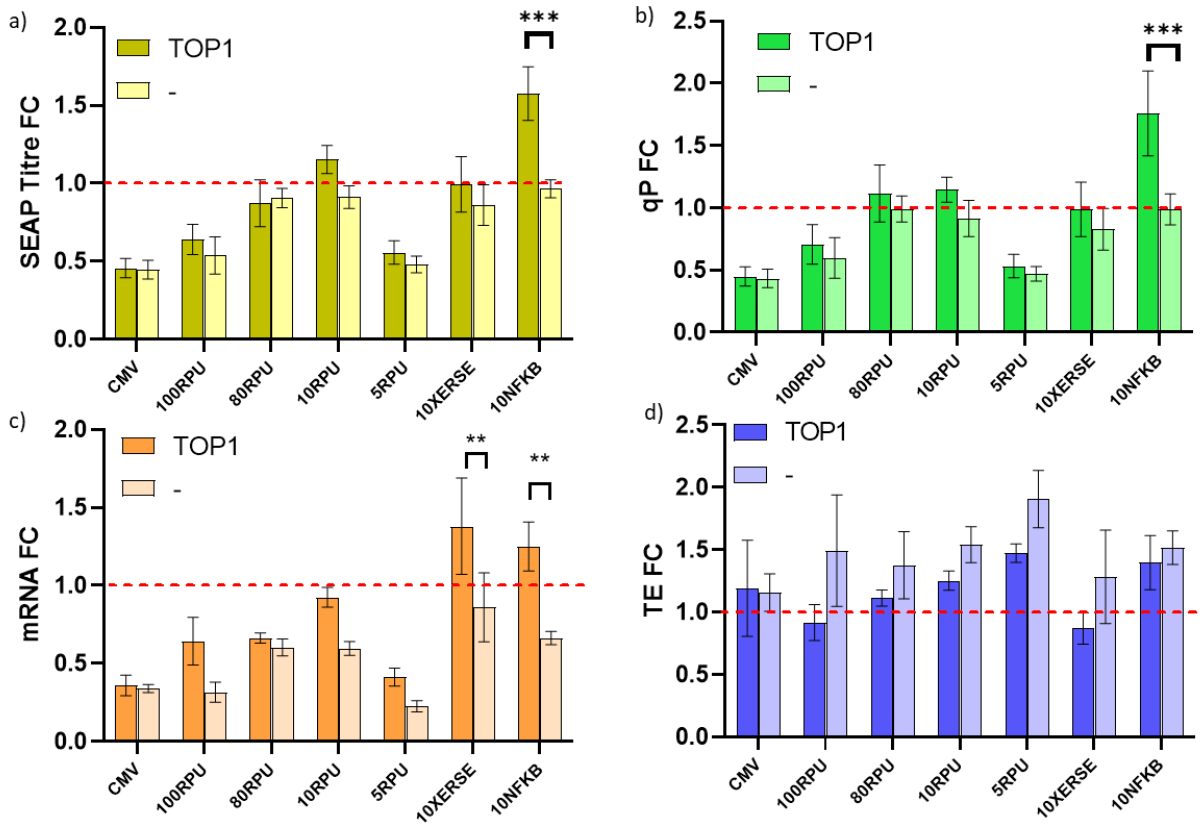


Figure 5.9 – Testing of synthetic proximal promoters with the TU1-Ef1α-SEAP vector, fold changes measured against unmodified Ef1α-SEAP. 10x NFκB shows a significant difference between its' TOP1/NON-TOP variants in A) SEAP titre (P=0.0003) B) SEAP specific productivity (P=0.0015) and C) SEAP mRNA abundance (P=0.0025). 10x ERSE also shows a significant difference in SEAP mRNA abundance between TOP1/NON-TOP (P=0.0064). All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1, and synthetic promoter sequences found in section 2.13.5. and Table 5.2 of this thesis. Results shown are of technical duplicates, from five biological replicates. FC: fold-change; qP: specific productivity; TE: translation efficiency.

5.2.5. The TCT core can integrate TOP motif functionality into the CMV promoter

To determine whether 5'TOP motifs could be integrated into alternative promoters that do not possess a TCT core and retain their function, a library of CMV-TCT-TOP vectors was created, using the TU1-SEAP vector, containing a CMV promoter. The Initiator sequence in the CMV core (GTCA$_{+1}$GA) was replaced by the TCT motif (TC$_{+1}$TTTTT) from the Ef1α. Four spacing variants were created, adding one base pair from the Ef1α core promoter upstream of the TCT motif each time, to vary the distance between the TATA box and transcription start site between 29bp (the distance from TATA box to Inr motif in WT CMV) and 32bp (the distance between the TATA box to transcription start site in WT CMV). Taking the C$_{+1}$ of the TCT motif as the start of a TOP motif, four TOP variants for each spacing variant were created: NONTOP (-), TOP1, TOP8, and TOP36. This molecular cloning was performed by site-directed mutagenesis as in section 2.3.2., and sequence verified by Sanger sequencing. 5'TOP motif sequences can be found in section 2.13.4. and Table 5.1, and a representative plasmid map under 'TU1-TCT-SEAP' in Appendix A. A schematic of the TU1-TCT -TOP-SEAP library is shown in Figure .

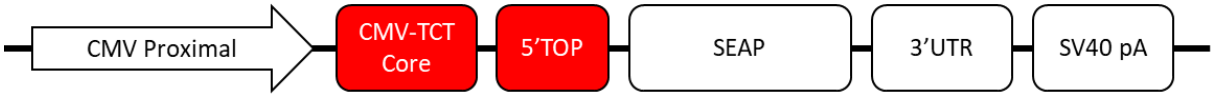


Figure 5.10 – A schematic of the TU1-TCT-TOP-SEAP library used in this thesis.

This total library of 16 constructs, alongside an unmodified TU1-SEAP control, was transfected by PEI at a 1:5 DNA:PEI ratio into transient host CHO cells at 1*10^6 cells/ml, in 24 shallow-well plate shaking culture. VCD was measured by Iprasense 72h and 120h post-transfection as in section 2.4.4.3., and SEAP titre measured 120h post-transfection by Sensolyte assay as in section 2.6.2. Fold change in SEAP titre and specific productivity, compared to TU1-SEAP, are shown in Figure 5.11a and Figure 5.11b. In all spacing variants, titre and specific productivity fold change increase sequentially from cells transfected with TOP1, to NON-TOP, to TOP8, to TOP36. For example, TCT-*SEAP* is associated with 0.117, 0.112, 0.747, and 1.04-fold changes in titre respectively, compared to TU1-*SEAP*. Little difference is observed in specific productivity between the TCT, TCT+2, and TCT+3 spacing variants.

However, transfection of the TCT+1 variant results in a significantly lower titre (0.64-fold) specific productivity (0.58-fold) with TOP36, compared to TCT-TOP36 (titre P=0.0084, specific productivity P=0.027) and TCT+2 (titre P=0.024, specific productivity P=0.035). These data suggest that TOP motifs are capable of imparting changes in recombinant protein titre in the context of a modified CMV promoter, up to levels comparable with unmodified CMV, and that whilst the system is partially permissive, some spacing variants are less effective in this than others.

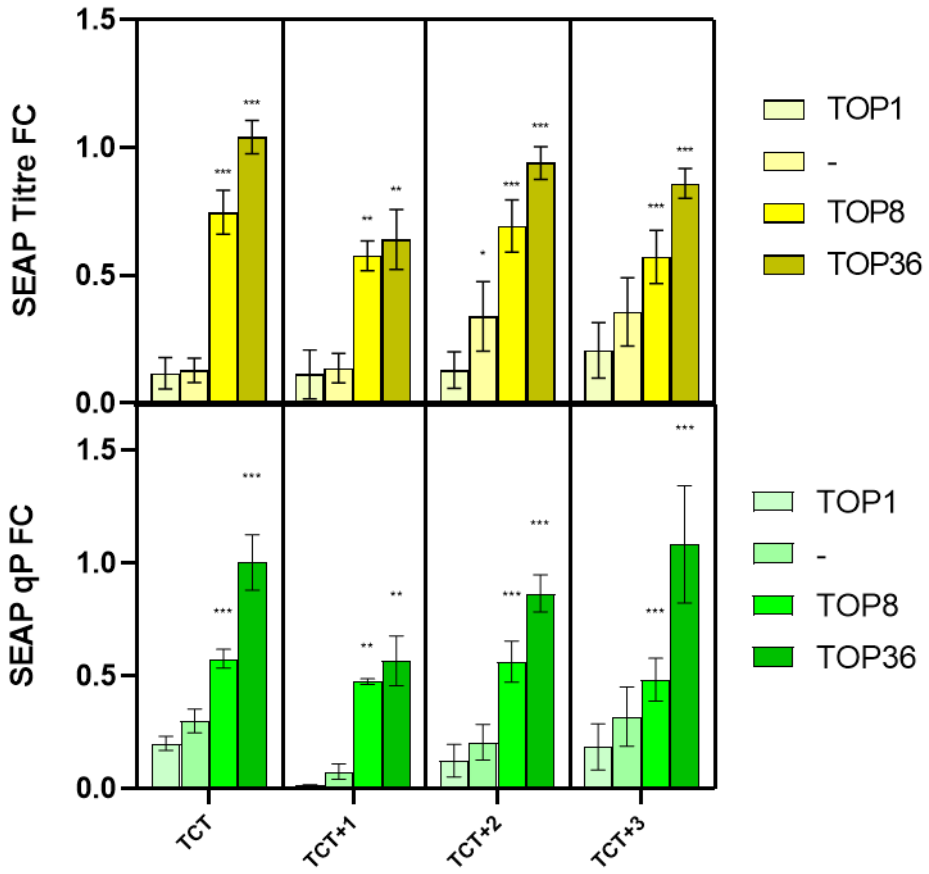


Figure 5.11 – Comparative analysis of the CMV-TCT-TOP-SEAP library, transfected by PEI at a 1:5 DNA:PEI ratio into the CHO transient host in 24-well plates. Fold change expression is measured relative to the TU1-SEAP control. Significance is measured relative to the NONTOP variant of each spacing variant. A) In all spacing variants, a sequential increase in titre is observed from TOP1 to NONTOP, to TOP8, to TOP36. TCT+1-TOP36 displays significantly reduced fold-change in titre compared to its' surrounding spacing variants. A maximal fold change 0f 1.04 is reached by TCT-TOP36. B) Specific productivity follows the same pattern as titre, with fold changes up to 1.08 with TCT+3-TOP36. All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1 of this thesis. Results presented are of technical duplicates from 5 biological replicates. FC: fold-change, qP: specific productivity.

5.2.6. 5'TOP motifs can modulate the titre of codon optimized and deoptimized EPO

To validate the ability of 5'TOP motifs to control the expression characteristics of a recombinant protein other than SEAP, and to investigate whether translation initiation activation by 5'TOP motifs can work synergistically with translation elongation activation, two EPO CDSs were inserted into the TU1-Ef1α vector: the codon deoptimised EPO68, and optimised EPO71. A selection of 8 5'TOPs which had a strong and significant effect on titre and specific productivity in the 5'TOP-SEAP library were then added to this vector. Two strategies were used to attempt the design of synthetic, optimised 5'TOP motifs. Firstly, the highest 10 performing 5'TOP motifs in terms of specific productivity from library 1 were analysed for conserved motifs, and the most significant result used as the new TOP sequence: synthetic TOP (synTOP). Next, the observation from Figure 5.5 that increases in expression are negatively correlated with TOP length was utilised to remove one base from the end of the strongest performing sequence from library 1, TOP36, creating TOP36-1. With the TOP1 and NON-TOP controls, a total library of 24 plasmids was created. These cloning steps were performed by restriction/ligation between the AgeI/SbfI sites as in section 2.3.6., and site-directed mutagenesis as in section 2.3.2., respectively, and verified by Sanger sequencing. All 5'TOP motif sequences can be found in section 2.13.4. and Table 5.1, and a representative plasmid map found under 'TU1-Ef1α' in Appendix A. A schematic of the TU1-TOP-EPO constructs used in this section is shown in Figure 5.12.

Figure 5.12 – A schematic of the TU1-TOP-EPO library used in this thesis.

This library was transfected into transient host CHO cells at $1*10^6$ cells/ml by PEI at a 1:5 DNA:PEI ratio, in 24 shallow-well plate shaking culture. VCD was measured at 72h and 120h post-transfection by Iprasense as in section 2.4.4.3., and supernatant assayed for EPO concentration by ELISA 120h post-transfection, as in section 2.10.2. The measured EPO titres are displayed in Figure 5.13a. Cells transfected with 19 constructs exhibited a significantly higher titre than with NONTOP-EPO68 with 381.5U/ml, up to a maximum of 903.3U/ml with TOP37-EPO71. EPO specific productivity, displayed in Figure 5.13b, follows a very similar pattern, with transfection of 21 constructs associated with significantly higher specific productivity than with NONTOP-EPO68, and range from 24.2U/ml/cell day$*10^6$, up to 59.3U/ml/cell day$*10^6$. Whilst the titre and specific productivity associated with NONTOP constructs increase in EPO71 compared to EPO68, only TOP37 significantly increases titre in both EPO variants, with its' titre increasing from 606.9-903.3U/ml (P=0.0079), and specific productivity increasing from 40.5-59.3U/mL/cell day$*10^6$ (P=0.022). These data imply that while both 5'TOP motifs and codon optimisation are capable of increasing recombinant protein expression, they may not be predictably synergistic in combination.

Cells from these experiments were also pelleted 120h post-transfection for analysis of relative *EPO* mRNA abundance by RT-qPCR as in section 2.9.2., Figure 5.13c. It should be noted that, since EPO68 and EPO71 are different CDSs, their relative abundance was detected by different primers. Their relative abundance is therefore only comparable to themselves, and not across the data sets. Relative to NONTOP-*EPO*68, transfection of two TOP-*EPO*68 constructs gave a significant increase in *EPO*68 mRNA: TOP24 and TOP36 with 3.11-fold and 2.51-fold changes respectively, whereas TOP47-*EPO*68 and TOP48-*EPO*68 were associated with significant decreases of 0.46-fold and o.41-fold respectively. In contrast, all *TOP*-EPO71 constructs increased *EPO*71 mRNA levels compared to the NONTOP-*EPO*71 control, 8 of which were significant, to a maximum significant increase of 2.90-fold with TOP37-*EPO*71. Analysing these data collectively by *EPO* variant in Figure 5.13d, it can be seen that addition of 5'TOP motifs to *EPO*71 has a significantly (P<0.0001) different effect than their introduction to *EPO*68, increasing mRNA abundance by an average of 2.49-fold, as opposed to an average of 1.37-fold with *EPO*68. These results suggest that 5'TOP motifs may be capable of influencing recombinant protein expression by more than one mechanism, and that this mechanism may be different depending on the codon optimality of the recombinant transcript.

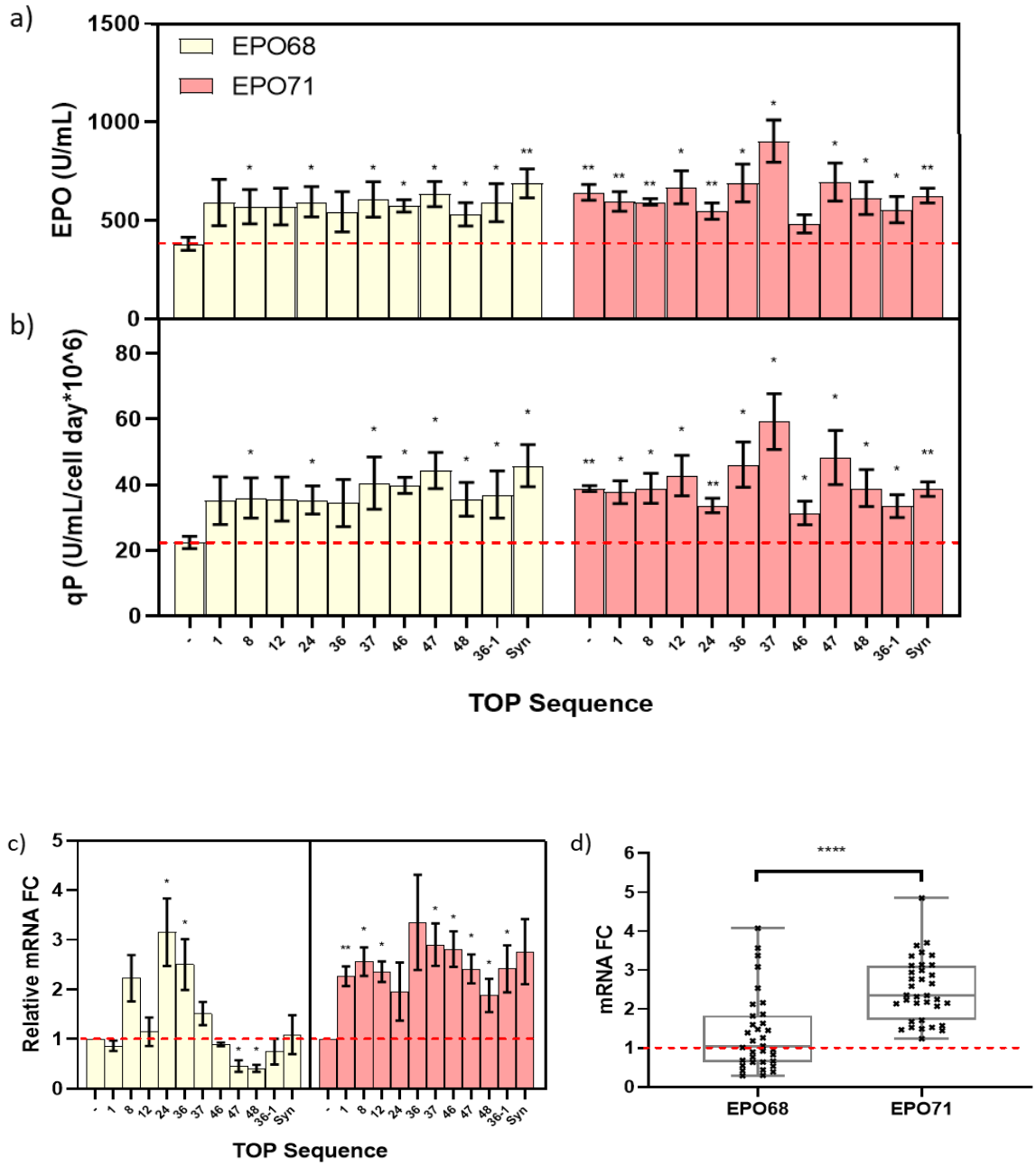


Figure 5.13 – Screening and analysis of TOP-EPO library, transfected by PEI at a 1:5 DNA:PEI ratio into the CHO transient host in 24-well plates. A) 19 constructs exhibit a significant (P<0.05) increase in EPO titre, compared to NONTOP-EPO68 with 381.5U/ml, up to 903.3U/ml with TOP37-EPO71. B) Specific productivites of the TOP-EPO library are significantly correlated with titres (R^2=0.934). 21 constructs significantly (P<0.05) increase specific productivity compared to NONTOP-EPO68, from 24.2U/ml/cell day10^6, up to 59.3U/ml/cell day*10^6 with TOP37-EPO71. C) With EPO68, a variety of significant (P<0.05) fold-changes in mRNA upon addition of TOPs was observed, from 0.41-fold with TOP48, and 3.11-fold with TOP24. All TOP motifs increased the relative abundance of EPO71 mRNA, 8/11 doing so significantly (P<0.05), to a maximum of 2.90-fold with TOP37. D) TOP motifs generally increase EPO71 mRNA abundance, but have no significant effect on EPO68. A significant (P<0.0001) difference is observed in the effect on mRNA abundance upon addition of TOP motifs to EPO71, compared to EPO68, an average of 2.49-fold increase compared to 1.37-fold. All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1 of this thesis. Results presented are of technical duplicates from four biological replicates. qP: specific productivity; FC: fold-change.*

RT-qPCR analysis of this screen offers further insight, in the finding that addition of 5'TOP sequences to EPO68 has mixed effects on mRNA abundance, generally implying an increase in translational efficiency, whereas addition to EPO71 led to an increase in mRNA in every case, surprisingly implying a decrease in translational efficiency. These fold-changes in translation efficiency, calculated by dividing specific productivity by mRNA abundance, compared to each NONTOP-EPO control are show in Figure 5.14. Addition of a 5'TOP motif to EPO71 universally led to a decrease in translational efficiency, significantly so with four sequences: TOP1 (P=0.001), TOP8 (P=0.001), TOP12 (P=0.08), and TOP46 (P=0.006). Meanwhile, addition of 5'TOP motifs to EPO68 led to more mixed effects, with four significant increases given by TOP12 (P=0.022), TOP46 (P=0.008), TOP48 (P=0.016), up to a 5.28-fold increase with TOP47 (P=0.037), and one significant decrease down to 0.54-fold with TOP24 (P=0.034).

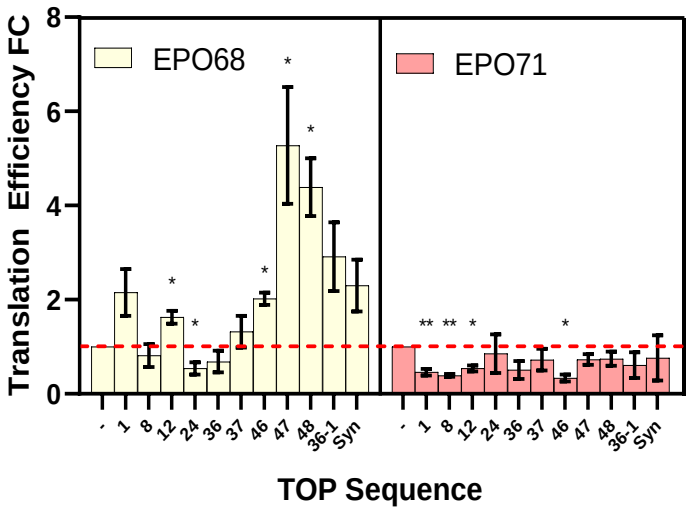


Figure 5.14 – Addition of 5'TOP motifs to codon de/optimized EPO has a variety of effects on translation efficiency. With deoptimized EPO68, effects vary between a significant 5.28-fold increase with TOP47 (P=0.037) and a significant decrease of 0.54-fold with TOP24 (P=0.034). With optimized EPO71, translation efficiency is universally decreased, four 5'TOP motifs doing so significantly, down to a 0.34-fold change with TOP46 (P=0.006). All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1 of this thesis. Results presented are of technical duplicates from four biological replicates. FC: fold-change.

5.2.7. 5'TOP motifs increase the titre of SEAP and a fusion protein in an industrial-standard transient expression laboratory

To validate that control of recombinant titre by 5'TOP motifs could be replicated in industrially-relevant culture and transfection platforms, and with fully industrially relevant products, two libraries were created. Firstly, a selection of 6 TOP motifs were inserted into the TU1-Ef1α-SEAP vector, alongside TOP1 and NONTOP controls. All 8 of these transcription units were subsequently inserted into an MGVT vector, containing an OriP for stimulation of transient production (Daramola et al., 2014). These molecular cloning steps were performed by site-directed mutagenesis, as in section 2.3.2., and golden-gate cloning as in section 2.3.5. respectively, and verified by Sanger sequencing. All 5'TOP sequences used can be found in section 2.13.4. and Table 5.1, and a plasmid map of MGVT found in Appendix A. A schematic of the MGVT-TOP-CDS constructs used in this thesis is shown in Figure 5.15.

Figure 5.15 – A schematic of the MGVT-TOP-CDS constructs used in this thesis.

In an AstraZeneca transient production laboratory, this library was transfected by PEI into the CHO transient host by a proprietary method, and subject to conditions for an AstraZeneca-optimised 7-day transient production process. To assess whether TOP expression could be further activated, two feed schemes were used, one feeding only on day 0 (D0 feed), the other feeding on day 0 and day 4 (D0+4 feed), both with a proprietary feed. Supernatants were collected 7 days post-transfection, and analysed for SEAP titre by Sensolyte assay, as in section 2.6.2. The results of this analysis are shown in Figure 5.16. Figure 5.16a shows the titre fold changes resulting from the D0 feed. Transfection of all seven TOP sequences increased SEAP titre compared to the NONTOP control, six of which were significant, with a maximum 1.88-fold change with TOP36. With the alternative feeding regime, D0+4 feed, displayed in Figure b, transfection of all seven TOP sequences were again associated with increased titre, six significantly, with a maximum of 2.56-fold with TOP37. Comparing the fold change in titre associated with the same TOP sequences between the two feeds in Figure c, the expression of four TOP sequences (TOP8, 12, 36, 37) is significantly activated by the D0+4 feed. Grouping all constructs with a TOP sequence together in Figure d, it can be seen that their expression is

significantly activated by the D0+4 feed compared to the D0 feed, with average titre fold-changes of 1.98 and 1.49, respectively. To assess whether control of titre with the same recombinant protein was replicable across different culture platforms, the titres measured from this experiment were plotted against those given by the same TOPs, in the 96 deep-well plate transfection described earlier in Figure 5.3. When plotted against titres from the D0 feed in Figure e, a significant correlation is noted, with an R^2 value of 0.71. In contrast, plotting of D0+4 feed titres in Figure f yields no significant correlation, with an R^2 value of 0.09. These data show that TOP motifs can be used to control titre in an industry-relevant transient culture platform, increasing titre beyond the natural Ef1α 5'TOP, that these titres may be further activated by different feed schemes, and that whilst the same TOP-CDS constructs can show consistency across culture platforms, this can be disrupted by different feed schemes.

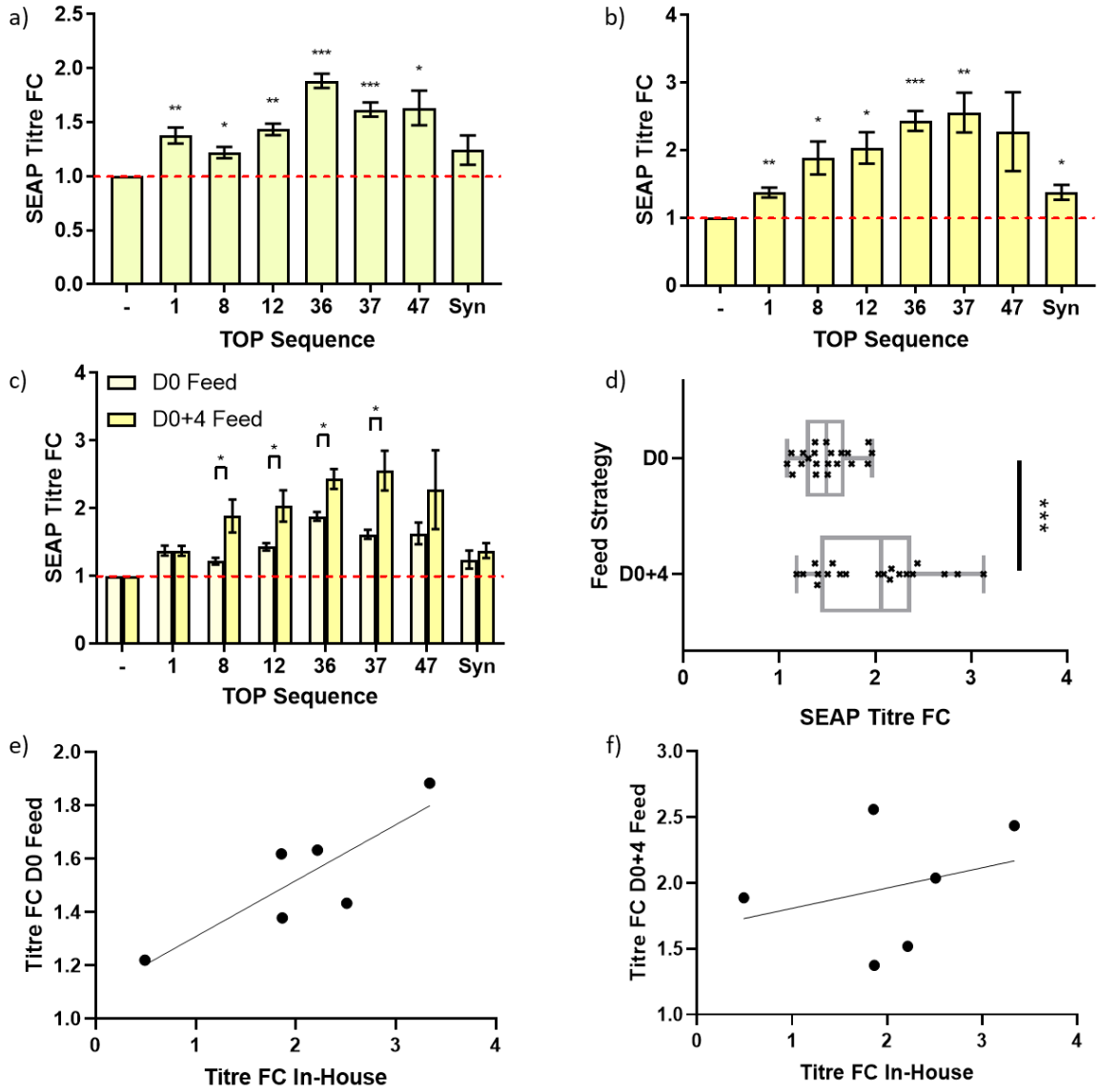


Figure 5.16 – 5'TOP motifs can control SEAP titre in an industry transient expression process. A) 6/7 TOP motifs impart a significant (P<0.05) increase in SEAP titre compared to NONTOP controls on a D0 feed, up to a 1.88-fold change with TOP36. B) 6/7 TOP motifs impart a significant (P<0.05) increase in SEAP titre compared to NONTOP controls on a D0+4 feed, up to a 2.56-fold change with TOP37. C) 4/7 TOP motifs are significantly (P<0.05) activated by the D0+4 feed compared to the D0 scheme. D) TOP-SEAP titres are significantly (P=0.0009) activated by the D0+4 feed compared to the D0 feed. E) TOP-SEAP titres from the D0 feed display a significant correlation (P=0.045, R^2=0.71) with their corresponding titres from a 96 deep-well plate transfection, whereas F) D0+4 fed cultures do not (P=0.56, R^2=0.09). All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1 of this thesis. The results shown are of two technical replicates, from three biological replicates. FC: fold-change; D: day.

To validate that 5'TOP motifs could be used to control titre of an industrially relevant protein, a single-chain fusion protein (scFP)was inserted into the TU1-Ef1α vector, the same 8 TOP variants introduced as in the previous experiment, and inserted into the MGVT vector containing OriP. These molecular cloning steps were performed by restriction/ligation

between AgeI/SbfI sites as in section 2.3.6., site-directed mutagenesis as in section 2.3.2., and golden-gate cloning as in section 2.3.5. respectively, and verified by Sanger sequencing. All 5'TOP sequences used can be found in section 2.13.4. and Table 5.1, and the MGVT plasmid map found in Appendix A.

These 8 constructs were similarly transfected with PEI into the CHO transient host by a proprietary method, and subject to the AstraZeneca 7-day transient production process, with the D0 and D0+4 feed schemes. Titres of the protein were measured 7 days post-transfection by Octet assay, as in section 2.10.3. Titres measured from the D0 feed are shown in Figure 5.17a. All TOP sequences were associated with increased titre compared to the NONTOP control at 61.4mg/l, TOP1 and TOP8 showing significant increases with 111.1mg/l and 109.5mg/l, respectively. Data from the D0+4 feed, Figure 5.17b, showed that cells transfected with all TOP sequences again exhibited increased scFP titre compared to the NONTOP control at 72.4mg/l, TOP37 and TOP36 showing significant increases with 153.6mg/l, and 127.2mg/l, respectively. Comparing titres derived from the two feeding regimes in Figure 5.17c, titre of scFP is increased by the D0+4 feed with all constructs, including the NONTOP control, but this increase is only significant with TOP37, increasing from 83.8mg/l to 153.6mg/l. Finally, the total TOP-scFP library shows significant expression activation by the D0+4 feed compared to the D0 feed, increasing from an average of 88.9mg/l to 124.2mg/ml, as shown in Figure 5.17d. These results show that TOP sequences can be used to control the titre of an industrially relevant biotherapeutic protein, and that in one case, further additive increases in titre may be achieved by applying different feeding schemes.

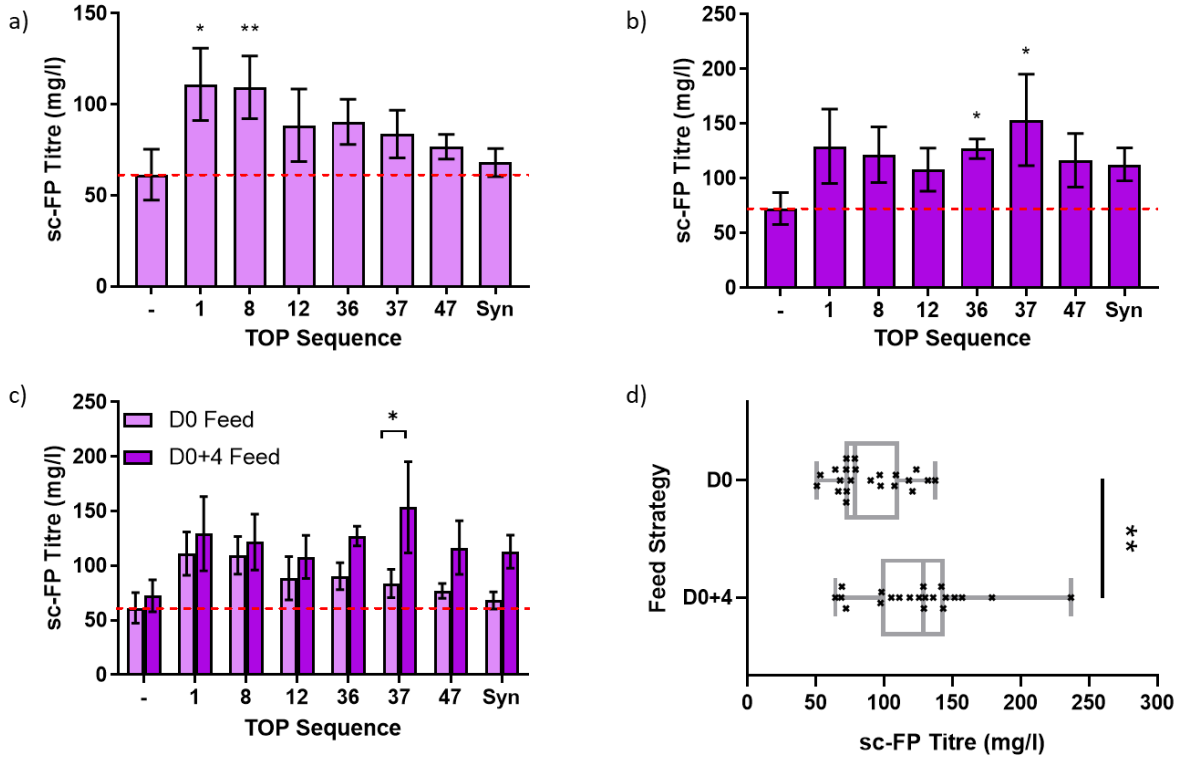


Figure 5.17 – 5'TOP motifs can control titre of an industrially relevant DTE molecule in an industrial transient expression process. A) All TOP motifs increase the titre of scFP compared to a NONTOP control with a D0 feed, two doing so significantly: TOP1 (P=0.025) and TOP8 (P=0.0051). B) All TOP motifs increase the titre of scFP compared to a NONTOP control with a D0+4 feed, two doing so significantly: TOP36 (P=0.033) and TOP37 (P=0.027). C) Expression of TOP37-scFP is significantly (P=0.023) activated by the D0+4 feed. D) As a group, expression of the TOP-scFP library is significantly (P=0.0015) increased by the D0+4 feed. All 5'TOP sequences referenced can be found in section 2.13.4. and Table 5.1 of this thesis. Results presented are of technical duplicates from three biological replicates. D: day.

5.3. Discussion

Throughout the results presented in this chapter, it is evident that the effect of 5'TOP motifs on recombinant protein expression is setting-dependent. The first such discrepancy came when screening the 5'TOP-*SEAP* library. Although PEI transfection was first utilised to target more consistent cell growth across the plate, it also leads to significant activation of the 5'TOP motifs' effect on titre. Although correlation was noted between their effect on specific productivity (R^2=0.64), electroporation and PEI led to identification of different 5'TOP motifs as significantly increasing expression, for instance 23 5'TOP motifs increased specific productivity with PEI, and 10 with electroporation, of which only 5 were in common. Therefore, one method had to be chosen as the model from which successful sequences could be identified. The 96 deep-well plate electroporation method has been shown to produce high transfection efficiencies of greater than 90% (Cartwright *et al.*, in print). However, its' variable effect on cell growth renders it a difficult system for reliably measuring specific productivity. Transfection by PEI led to far more consistent growth, considerably activated 5'TOP expression compared to electroporation, and is the most commonly used method for high-yielding transient biopharmaceutical production (Daramola *et al.*, 2014; Hacker *et al.*, 2013; Rajendra *et al.*, 2015a). Therefore, it was chosen as the more appropriate model system for selection of 5'TOP motifs to take forward into future studies.

Activation of MTORC1, and therefore of acute activation of 5'TOPs, is mediated through both growth factors and nutrients, particularly glucose, leucine, arginine, and glutamine (Ben-Sahra and Manning, 2017; Saxton and Sabatini, 2017; Yao *et al.*, 2017), with both required for its' full activation (Valvezan and Manning, 2019), and inhibited by rapamycin (Roux and Topisirovic, 2018). However, supplementation titration of culture in Figure 5.6 with both a commercially available feed, and with specifically targeted amino acid mix failed to specifically activate 5'TOP-*SEAP* expression. The same figure shows the successful implementation of the effect in reverse: specific, acute repression of 5'TOP expression by 25ng/ml rapamycin. One explanation may be that 5'TOP motifs are already active in basal CHO cell culture, with nutrient-rich media and ideal conditions for cell growth. Figure 5.3 shows 5'TOP motifs are sufficiently activated in basal CHO cell culture to give a 3.34-fold increase in SEAP titre with the right 5'TOP sequence. However, this explanation is

contradicted by analysis of the Omics dataset from which the 5'TOP library was derived, comparing the Log_{10} of various expression characteristics of TOP and NON-TOP genes, presented in Figure 5.18. Figure 5.18a, b, and c show that TOP genes have on average a higher protein abundance, mRNA abundance, and total translational activity that their NON-TOP counterparts. However, the translational efficiency (how many amino acids are added to the respective gene's protein, per hour, per mRNA) of TOP genes is slightly lower than NON-TOP. This is consistent with current models of 5'TOP repression, wherein mRNA abundance is kept high by the stabilising effect of *LARP1* binding and 40s ribosome sequestering (Gentilella *et al.*, 2017). Moreover, Omics datasets from other cell lines show that 5'TOP genes can exist in a translationally-enhanced state, as shown in Figure 5.19, where an Omics dataset derived from NIH3T3 cells (Schwanhausser *et al.*, 2011) shows that translation efficiency of TOP genes (and a ribosomal protein subset) are significantly increased compared to the total gene dataset, in contrast to the CHO Omics dataset, where no such activation is observed.

Furthermore, significant activation of TOP expression was achieved in Figure 5.16 and Figure 5.17 by differences in proprietary feed, with both SEAP and scFP. Collectively, the data in this thesis imply that whilst TOP motif expression is sufficiently activated to produce significant increases in titre, it may not be fully activated in basal CHO cell culture. TOP motif expression can be further stimulated by chemical manipulation via a proprietary feed. This effect was not replicated by simple nutrient supplementation, but the root cause of the activation cannot be analysed without knowing the constituent parts of the proprietary feed.

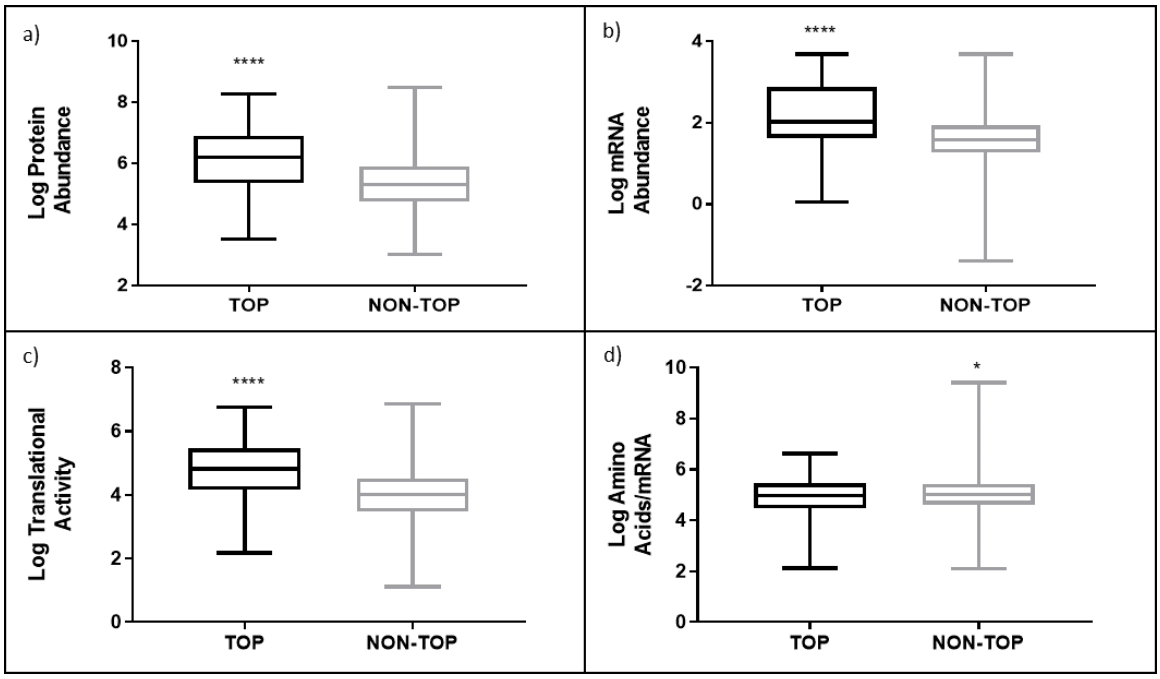


Figure 5.18 – Translation efficiency of TOP genes is not activated in basal CHO cell culture. TOP genes exhibited higher A) protein abundance (P<0.0001), B) mRNA abundance (P<0.0001), and C) translational activity (P<0.0001) that NON-TOP genes. However, as shown in D) translation efficiency is significantly (P=0.049) higher in NON-TOP than TOP genes. Data from AstraZeneca.

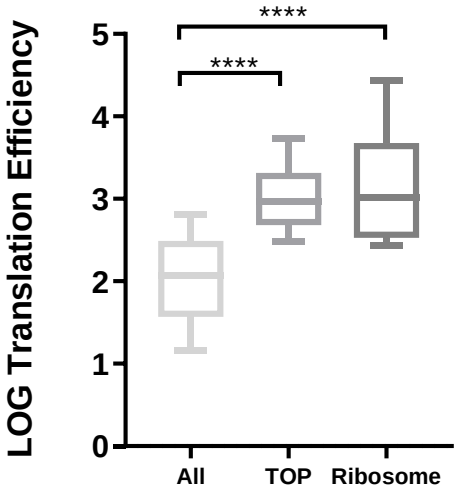


Figure 5.19 – Translation efficiency of TOP genes is activated in NIH3T3 culture. TOP genes, as well as a ribosomal protein subset, both exhibited significantly (P<0.0001) higher translational efficiency than the total gene dataset (Schwanhausser *et al.*, 2011).

In an attempt to integrate the translational control imparted by 5'TOP motifs with the transcriptional control granted by synthetic promoters, a selection of user-defined synthetic proximal promoters taken from Brown et al., 2014 were inserted upstream of the Ef1α core promoter. Figure 5.9 shows that this approach was unsuccessful: the promoter strengths did not align to user definition. This can be explained by the promoter architecture of Ef1α. The transcription start-site element of Ef1α is a TCT motif, in contrast to CMV, which uses an Inr motif. TCT motifs, and the resultant 5'TOP motifs they transcribe, are a rare core element, occurring in ~1% of mammalian promoters (Ngoc et al., 2017). Importantly, TCT core promoters are not bound by the TFIID/TBP complex which is necessary for Inr transcription, and indeed are insensitive to TBP knockdown (Wang et al., 2014). Instead, TCT core promoters are bound by TRF2, facilitating TATA-independent transcription (Duttke et al., 2014). It has been postulated that this functional difference allows for specialised control of transcription (Lenhard et al., 2012), which may have facilitated evolution of the Bilateria (Duttke et al., 2014). The RNA Polymerase II core promoter is described as punctilious (Vo ngoc et al., 2017): sequence elements are essential and specific for function, and it has been shown that enhancer regions for TCT promoters do not necessarily stimulate Inr core promotors, and vice-versa (Zabidi et al., 2015). Therefore, the synthetic promoters used in this study did not act as designed because they were designed against the Inr CMV core (Brown et al., 2014a). The success of the 10xNFκB promoter demonstrates, however, that synthetic promoters could be designed against the Ef1α core, and even that it may have responsive TFREs in common with CMV. Taking the opposite approach, by integrating a TCT core into the CMV promoter yielded promising results, as seen in Figure 5.11, in contrast to similar attempts made in literature with endogenous genes (Parry et al., 2010). Relatively strong expression from an Inr-less promoter, and differential yields between TOP variants suggest TCT-based, TRF2-dependent transcription. Whether or not this TCT core is rendered receptive to Inr-based proximal control by placement in an Inr promoter will be explored in Chapter 6 of this thesis.

The results presented in this chapter show that 5'TOP motifs are a viable and powerful tool for control of biopharmaceutical expression, especially in the light of Figure 5.16 and Figure 5.17, showing their efficacy both in an optimised transient biopharmaceutical production setting, and with a directly industrially relevant biotherapeutic protein. Their effect on expression, however, appears not to be predictable. This is demonstrated in Table , where the

average titre fold-changes imparted by four 5'TOP motifs across different protein products and culture conditions are shown.

Having established by results presented in Figure 5.3 that 5'TOP motifs can grant significant increases in SEAP titre and specific productivity, a more nuanced result was gained from screening of the 5'TOP-EPO library, in Figure 5.13. Addition of TOP sequences to the deoptimised EPO68 universally increased titre and specific productivity. Similarly, the optimised EPO71 carried a significantly higher titre without a TOP sequence, compared to NONTOP-EPO68. However, only TOP37 significantly increased the titre of EPO71, compared to the NONTOP control. Moreover, TOP37 and NONTOP were the only variants which saw a significant increase in titre between EPO68 and EPO71. Whilst TOP37 may share a sequence-specific affinity with EPO71 (an effect which will be discussed below), these findings generally imply that although codon optimisation or addition of TOP motifs individually benefit expression, their effects may not combine in a synergistic, modular fashion. When translation efficiency is sufficiently optimised, by initiation or elongation, these data suggest that a bottleneck is formed downstream, for instance in folding or secretion, rendering expression insensitive to further increases in translation efficiency. Further investigation using codon optimised and deoptimised proteins known to be easy (e.g. SEAP, GFP) and difficult to express (e.g. an aggregating mAb) may elucidate this question.

Another explanation for this bottleneck could be the saturation of translational machinery, for instance initiation factors or ribosomes, by sufficiently high levels of EPO expression (Jackson *et al*., 2010). This hypothesis could be investigated by additional stimulation of expression by cellular engineering by translational, and non-translational, components. If overexpression of non-translational components, for example XBP1, fails to relieve this expression bottleneck, and translational components succeed, it would support this hypothesis.

Two mechanisms of stimulation of EPO titre can be inferred from Figure 5.14, where codon-optimized EPO68 is increased in terms of translation efficiency, whereas the optimized EPO71 is increased in terms of mRNA abundance. Both of these mechanisms of expression manipulation are consistent with models of TOP expression. In the case of EPO71, MTORC1 may not be fully activated, meaning TOP mRNA is sequestered, increasing half-life and decreasing translation efficiency (Gentilella *et al.*, 2017). With EPO68, MTORC1 may be

activated or deactivated, replicating the effect observed with EPO71, or increasing translation efficiency (Hong *et al*., 2017). These observations are consistent with the hypothesis of downstream expression bottlenecks: if EPO translation, engineered through 5'TOP motifs or codon optimisation, is elevated enough to cause bottlenecks in folding and induce the UPR, autophagy may be induced to counterbalance this effect (Bernales et al., 2006; Hussain et al., 2014), requiring the deactivation of MTORC1 (Saxton and Sabatini, 2017). As with all molecular biology tools, 5'TOP motifs are not capable of alleviating bottlenecks downstream of their point of action, and these data demonstrate that control of translation may be rendered ineffective if appropriate engineering strategies post-translation are not implemented.

Protein	SEAP			EPO		scFP	
Condition	96DWP 72h	AZ D0 Feed	AZ D0+4 Feed	EPO68 24SWP 120h	EPO71 24SWP 120h	AZ D0 Feed	AZ D0+4 Feed
TOP8	0.49	1.22	1.89	1.49	0.92	1.78	1.68
TOP12	2.51	1.43	2.04	1.5	1.04	1.44	1.49
TOP36	3.34	1.88	2.44	1.43	1.07	1.47	1.76
TOP37	1.86	1.62	2.56	1.59	1.4	1.36	2.12

Table 5.3 – 5'TOP motifs produce highly variable fold-changes in titre across different product proteins and conditions. Heatmaps showing strength of titre fold-change are calibrated separately for each product protein.

The effect exerted by 5'TOP motifs on titre appears firstly dependent on the protein being expressed. TOP8, for instance, appears ineffective at enhancing SEAP titre, giving the lowest titre in all three conditions tested, down to a 0.49-fold decrease in 96 deep-well plates. However, it performs reasonably well producing scFP, producing the highest titre with the D0 feed. TOP36, in contrast, excels at SEAP production, scoring the highest titre fold changes in two conditions, and second-highest in the third, up to an impressive 3.34-fold increase in 96 deep-well plates, but is only moderately effective at producing EPO and scFP, scoring three second places and one last place. TOP12 is a moderate-to-low performer for SEAP and EPO, scoring two second and three third places, but is drastically worse at producing scFP, being associated with the lowest titre in both conditions.

Secondly, effects on titre appear dependent on cell culture conditions. TOP37 is generally one of the highest-performing presented in this study, producing the highest titre in 4 out of the 7 conditions presented in Table , including a 2.12-fold increase in scFP titre with the D0+4 feed. However, with the same protein and the D0 feed, it is associated with the lowest titre.

In an ideal case, a TOP sequence could be synthetically designed to possess a user-defined effect on expression, titratable across a range of titres. Whilst this may be possible, the data presented in this chapter suggest that modelling of the TOP's surrounding plasmid, coding sequence, expression system, and culture conditions may be required to reliably predict its' effect on expression.

Whilst the use of 5'TOP motifs for controlling translation presented in this thesis are predictable nor titratable in its effect, it may still be a useful target in a transient screening system. Desirable traits of the tools develop in this chapter include dramatic changes include the ability to dramatically increase transient titre, for example of a biopharmaceutical scFP by >2-fold, integration with multiple promoter assemblies, and relatively simple insertion of such short sequences into expression plasmids. Combined, these characteristics make them an attractive tool for high-throughput transient testing.

5.4. Conclusion

In this chapter, I have investigated and developed the novel use of 5'TOP motifs as a tool for controlling biopharmaceutical expression. Their power to affect recombinant protein expression was repeated across multiple culture conditions and proteins, including the increase in productivity of a biotherapeutic protein in an industry optimised 7-day transfection platform, where TOP37 was associated with an increase in titre from 72.4mg/l, to 153.6mg/l, compared to a NONTOP control. Promoter architecture requirements for their proper expression may limit their use beyond the Ef1α promoter, though this may be circumvented by the modification of Inr to TCT promoter cores, novel within the context of an industrial Ef1α promoter. With currently available data, a predictive model of 5'TOP motif's effect on titre cannot be constructed. However, due to their effect on recombinant protein expression, and ease of implementation, they represent a viable and attractive tool for controlling biopharmaceutical expression.

5.5. References

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. 19, 568–576.

Aoki, K., Adachi, S., Homoto, M., Kusano, H., Koike, K., and Natsume, T. (2013). LARP1 specifically recognizes the 3′ terminus of poly(A) mRNA. FEBS Lett. 587, 2173–2178.

Aylett, C.H.S., Sauer, E., Imseng, S., Boehringer, D., Hall, M.N., Ban, N., and Maier, T. (2016). Architecture of human mTOR complex 1. Science (80-.). 351, 48–52.

Ben-Sahra, I., and Manning, B.D. (2017). mTORC1 signaling and the metabolic control of cell growth. Curr. Opin. Cell Biol. 45, 72–82.

Bernales, S., McDonald, K.L., and Walter, P. (2006). Autophagy counterbalances endoplasmic reticulum expansion during the unfolded protein response. PLoS Biol. 4, 2311–2324.

Brown, A.J., Sweeney, B., Mainwaring, D.O., and James, D.C. (2014). Synthetic promoters for CHO cell engineering. Biotechnol. Bioeng. 111, 1638–1647.

Buel, G.R., and Blenis, J. (2016). Seeing mTORC1 specificity. Science (80-.). 351, 25–26.

Courtes, F.C., Vardy, L., Wong, N.S.C., Bardor, M., Yap, M.G.S., and Lee, D.Y. (2014). Understanding translational control mechanisms of the mTOR pathway in CHO cells by polysome profiling. N. Biotechnol. 31, 514–523.

Dadehbeigi, N., and Dickson, A.J. (2015). Chemical manipulation of the mTORC1 pathway in industrially relevant CHOK1 cells enhances production of therapeutic proteins. Biotechnol. J. 10, 1041–1050.

Daramola, O., Stevenson, J., Dean, G., Hatton, D., Pettman, G., Holmes, W., and Field, R. (2014). A high-yielding CHO transient system: Coexpression of genes encoding EBNA-1 and GS enhances transient protein expression. Biotechnol. Prog. 30, 132–141.

Dreesen, I.A.J., and Fussenegger, M. (2011). Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. Biotechnol. Bioeng. 108, 853–866.

Duttke, S.H.C., Doolittle, R.F., Wang, Y.L., and Kadonaga, J.T. (2014). TRF2 and the evolution of the bilateria. Genes Dev. 28, 2071–2076.

Eliseeva, I., Vorontsov, I., Babeyev, K., Buyanova, S., Sysoeva, M., Kondrashov, F., and Kulakovskiy, I. (2013). In silico motif analysis suggests an interplay of transcriptional and translational control in mTOR response. Translation 1, 18–24.

Fonseca, B.D., Lahr, R.M., Damgaard, C.K., Alain, T., and Berman, A.J. (2018). LARP1 on TOP of ribosome production. Wiley Interdiscip. Rev. RNA 1–16.

Gandin, V., Masvidal, L., Hulea, L., Gravel, S.P., Cargnello, M., McLaughlan, S., Cai, Y., Balanathan, P., Morita, M., Rajakumar, A., et al. (2016). NanoCAGE reveals 5′ UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs. Genome Res. 26, 636–648.

Gentilella, A., Morón-Duran, F.D., Fuentes, P., Zweig-Rocha, G., Riaño-Canalias, F., Pelletier, J., Ruiz, M., Turón, G., Castaño, J., Tauler, A., et al. (2017). Autogenous Control of 5′TOP mRNA Stability by 40S Ribosomes. Mol. Cell 67, 55-70.e4.

Geoghegan, D., Arnall, C., Hatton, D., Noble-Longster, J., Sellick, C., Senussi, T., and James, D.C. (2018). Control of amino acid transport into Chinese hamster ovary cells. Biotechnol. Bioeng. 115, 2908–2929.

Hacker, D.L., Kiseljak, D., Rajendra, Y., Thurnheer, S., Baldi, L., and Wurm, F.M. (2013). Polyethyleneimine-based transient gene expression processes for suspension-adapted HEK-293E and CHO-DG44 cells. Protein Expr. Purif. 92, 67–76.

Hong, S., Freeberg, M.A., Han, T., Kamath, A., Yao, Y., Fukuda, T., Suzuki, T., Kim, J.K., and Inoki, K. (2017). LARP1 functions as a molecular switch for mTORC1-mediated translation of an essential class of mRNAs. Elife 6.

Hopkins, K.C., Tartell, M.A., Herrmann, C., Hackett, B.A., Taschuk, F., Panda, D., Menghani, S. V., Sabin, L.R., and Cherry, S. (2015). Virus-induced translational arrest through 4EBP1/2-dependent decay of 5′-TOP mRNAs restricts viral infection. Proc. Natl. Acad. Sci. 112, E2920–E2929.

Hussain, H., Maldonado-Agurto, R., and Dickson, A.J. (2014). The endoplasmic reticulum and unfolded protein response in the control of mammalian recombinant protein production. Biotechnol. Lett. 36, 1581–1593.

Iadevaia, V., Liu, R., and Proud, C.G. (2014). MTORC1 signaling controls multiple steps in ribosome biogenesis. Semin. Cell Dev. Biol. 36, 113–120.

Josse, L., Xie, J., Proud, C.G., and Smales, C.M. (2016). mTORC1 signalling and eIF4E/4E-BP1 translation initiation factor stoichiometry influence recombinant protein productivity from GS-CHOK1 cells. Biochem. J. 473, 4651–4664.

Lahr, R.M., Mack, S.M., Héroux, A., Blagden, S.P., Bousquet-Antonelli, C., Deragon, J.M., and Berman, A.J. (2015). The La-related protein 1-specific domain repurposes HEAT-like repeats to directly bind a 5′TOP sequence. Nucleic Acids Res. 43, 8077–8088.

Lahr, R.M., Fonseca, B.D., Ciotti, G.E., Al-Ashtal, H.A., Jia, J.J., Niklaus, M.R., Blagden, S.P., Alain, T., and Berman, A.J. (2017). La-related protein 1 (LARP1) binds the mRNA cap, blocking eIF4F assembly on TOP mRNAs. Elife 6, 1–15.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. 13, 233–245.

Leppek, K., Das, R., and Barna, M. (2018). Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat. Rev. Mol. Cell Biol. 19, 158–174.

Markou, T., Marshall, A.K., Cullingford, T.E., Tham, E.L., Sugden, P.H., and Clerk, A. (2010). Regulation of the cardiomyocyte transcriptome vs translatome by endothelin-1 and insulin: Translational regulation of 5′ terminal oligopyrimidine tract (TOP) mRNAs by insulin. BMC Genomics 11.

Masvidal, L., Hulea, L., Furic, L., Topisirovic, I., and Larsson, O. (2017). mTOR-sensitive translation: Cleared fog reveals more trees. RNA Biol. 14, 1299–1305.

Mayberry, L.K., Allen, M.L., Dennis, M.D., and Browning, K.S. (2009). Evidence for Variation in the Optimal Translation Initiation Complex: Plant eIF4B, eIF4F, and eIF(iso)4F Differentially Promote Translation of mRNAs. Plant Physiol. 150, 1844–1854.

Meyuhas, O. (2015). Ribosomal Protein S6 Phosphorylation: Four Decades of Research. In International Review of Cell and Molecular Biology, p.

Meyuhas, O., and Kahan, T. (2015). The race to decipher the top secrets of TOP mRNAs. Biochim. Biophys. Acta - Gene Regul. Mech. 1849, 801–811.

Miloslavski, R., Cohen, E., Avraham, A., Iluz, Y., Hayouka, Z., Kasir, J., Mudhasani, R., Jones, S.N., Cybulski, N., Rüegg, M.A., et al. (2014). Oxygen sufficiency controls TOP mRNA translation via the TSC-Rheb-mTOR pathway in a 4E-BP-independent manner. J. Mol. Cell Biol. 6, 255–266.

Musa, J., Orth, M.F., Dallmayer, M., Baldauf, M., Pardo, C., Rotblat, B., Kirchner, T., Leprivier, G., and Grünewald, T.G.P. (2016). Eukaryotic initiation factor 4E-binding protein 1 (4E-BP1): A master regulator of mRNA translation involved in tumorigenesis. Oncogene 35, 4675–4688.

Nandagopal, N., and Roux, P.P. (2015). Regulation of global and specific mRNA translation by the mTOR signaling pathway. Translation 3, e983402.

Ngoc, L.V., Cassidy, C.J., Huang, C.Y., Duttke, S.H.C., and Kadonaga, J.T. (2017). The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. Genes Dev. 31, 6–11.

Parry, T.J., Theisen, J.W.M., Hsu, J.Y., Wang, Y.L., Corcoran, D.L., Eustice, M., Ohler, U., and Kadonaga, J.T. (2010). The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes Dev. 24, 2013–2018.

Philippe, L., Vasseur, J.J., Debart, F., and Thoreen, C.C. (2018). La-related protein 1 (LARP1) repression of TOP mRNA translation is mediated through its cap-binding domain and controlled by an adjacent regulatory region. Nucleic Acids Res. 46, 1457–1469.

Rajendra, Y., Kiseljak, D., Baldi, L., Wurm, F.M., and Hacker, D.L. (2015). Transcriptional and post-transcriptional limitations of high-yielding, PEI-mediated transient transfection with CHO and HEK-293E cells. Biotechnol. Prog. 31, 541–549.

Roux, P.P., and Topisirovic, I. (2018). Signaling pathways involved in the regulation of mRNA translation. Mol. Cell. Biol. MCB.00070-18.

Saxton, R.A., and Sabatini, D.M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. Cell 168, 960–976.

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

Tcherkezian, J., Cargnello, M., Romeo, Y., Huttlin, E.L., Lavoie, G., Gygi, S.P., and Roux, P.P. (2014). Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5 ′ TOP mRNA translation. Genes Dev. 28, 357–371.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature 485, 109–113.

Tuxworth, W.J., Shiraishi, H., Moschella, P.C., Yamane, K., McDermott, P.J., and Kuppuswamy, D. (2008). Translational activation of 5′-TOP mRNA in pressure overload myocardium. Basic Res. Cardiol. 103, 41–53.

Valvezan, A.J., and Manning, B.D. (2019). Molecular logic of mTORC1 signalling as a metabolic rheostat. Nat. Metab. 1, 321–333.

Vo ngoc, L., Wang, Y.-L., Kassavetis, G.A., and Kadonaga, J.T. (2017). The punctilious RNA polymerase II core promoter. Genes Dev. 31, 1289–1301.

Wang, Y.L., Duttke, S.H.C., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., and Kadonaga, J.T. (2014). TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes Dev. 28, 1550–1555.

Yamashita, R., Suzuki, Y., Takeuchi, N., Wakaguri, H., Ueda, T., Sugano, S., and Nakai, K. (2008). Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. Nucleic Acids Res. 36, 3707–3715.

Yao, Y., Jones, E., and Inoki, K. (2017). Lysosomal regulation of mTORC1 by amino acids in mammalian cells. Biomolecules 7, 1–18.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature 518, 556–559.

6. Multielement Screen and Concluding Remarks

6.1. Introduction

Besides being predictable and titratable, an additional desirable trait of synthetic molecular biology tools is to be modular, meaning it can work in concert with other tools being deployed elsewhere in the system. The generalised behaviour of two wholly modular tools are shown in Figure 6.1. The effects of the two modular tools combine synergistically, and their product describes the effect when used in combination.

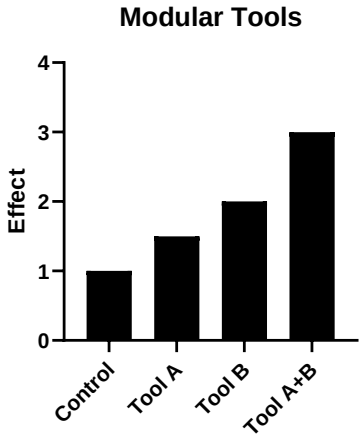


Figure 6.1 – The behaviour of wholly modular synthetic biology tools. The effects of Tool A (1.5-fold) and Tool B (2-fold) combine synergistically, to provide a total 3-fold effect when used in combination.

In practice, almost no synthetic biology tools will be entirely modular in this way. Overlaps and redundancies in mechanism, bottlenecks in expression downstream, and dozens of unpredictable example-specific interactions all prevent tools from cooperating in this strictly synergistic fashion. Nonetheless, it is important to know to what extent and in what circumstances tools behave modularly, as the most powerful expression and process engineering occurs when such tools are deployed together. For example, in Johari et al., 2015, both cellular engineering by protein overexpression, and chemical treatment of cultures are tested as methods of enhancing production of a DTE Fc-fusion protein. From this screen, CypB is identified as a target gene to enhance transient production, increasing titre by ~1.4-fold when cotransfected at 20% of the gene load of the Fc-fusion, mediated through increased IVCD. Two chemical compounds were also identified: PBA and Glycerol, which increased transient titre by ~1.5-fold when used to treat culture at 0.5mM and 1% concentration

respectively, both affecting titre through increasing specific productivity. These three tools were combined into a 12-day biphasic transient production process, which overall gave a 6-fold increase in protein titre, displaying the power of synergistically combining synthetic biology tools. Further demonstration of the power of modular tools, and demonstration of limitations in their modularity, are found in Brown et al., 2019. In this study, various vector elements are first tested individually for their ability to increase SEAP titre, relative to a construct bearing industry-standard parts. A synthetic promoter is identified that increases titre by 2.05-fold, a signal peptide that does to by 1.48-fold, and a method of codon optimisation that does so by 1.38-fold. In combination, all three elements achieved only a 3.02-fold increase in titre, in a non-synergistic combination that failed to outperform a similar construct made from less optimal variants of all three synthetic parts. However, when this bottleneck was relieved by integration with an adapted cell-line for secretion, and identification of ATF6 as an effector gene at an optimal dosage, the components acted in a more synergistic pattern, up to an impressive 9.24-fold increase in SEAP titre with the original transient screening system, increasing up to 12.37-fold in a six day fed-batch transient process, as shown in Table 6.1.

Synthetic Tool	SEAP Titre Fold Change
Promoter	2.05
Signal Peptide	1.48
Codon Optimisation	1.38
Cell Line	2.20
Effector Gene	1.61
Predicted Synergistic Effect	**14.83**
Observed Synergistic Effect	**12.37**

Table 6.1 – The synergistic effect of molecular biology tools identified by Brown et al., 2019. The predicted effect upon perfect synergy of each component (14.83-fold) is closely matched by the observed effect (12.37-fold), indicating a high degree of modularity.

Finally, modular tools can also be used for more creative purposes than the synergistic maximisation of protein titre. For example in Yeo *et al.*, 2017, different synthetic antibiotic selection markers are attenuated using mutated IRESs in alternative cell lines. Using Zeomycin as a selection agent, as opposed to Hygromycin increased product titres from 9.80mg/l to 29.1mg/l, and selection marker attenuation of this IRES-Zeomycin brought titres as high as 353.5mg/l.

6.2. Results

6.2.1. Construction of the multielement SEAP library

In order to test the modularity of the synthetic biology tools identified in this thesis, and to measure the range of expression they enabled in combination, a multi-element library was constructed. This library varied the core promoters (Ef1α, CMV-TCT), proximal promoters (100RPU, 20RPU, 10xNFκB), TOP motif (TOP2, 8, 36, 37), and stability element (SE20), in the TU1-*SEAP* vector. Restriction/ligation and site-directed mutagenesis molecular cloning were both used to create this library, as in sections 2.3.6. and 2.3.2., which was verified by Sanger sequencing. Since the EF1α core has previously been shown not to perform as expected with the Brown *et al.*, 2014 synthetic promoters, this combination was not used. The 100RPU and 20RPU promoters were instead combined with the CMV-TCT core promoter, and the 10xNFκB proximal promoter combined with the Ef1α core. Apart from these exceptions, every component was used in combination with every other, ensuring a thorough test of their modularity. This made for a total library of 40 constructs, in addition to the 2 controls of TU1-*SEAP*, and TU1-Ef1a-*SEAP*, summarised in Figure 6.2.

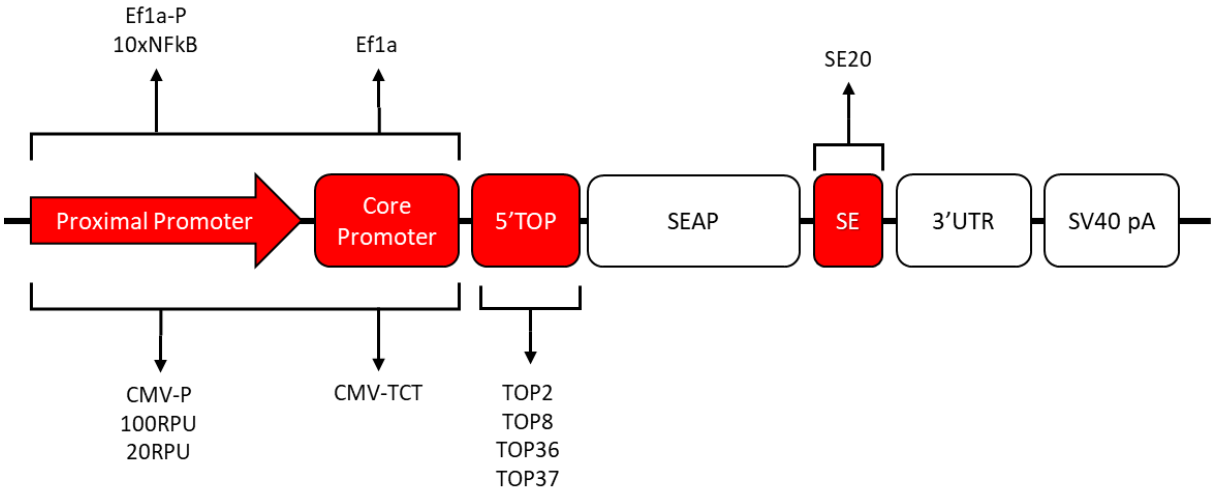


Figure 6.2 - A schematic of the multielement SEAP library. Four elements were varied in this library: proximal promoter, core promoter, TOP motif, and stability element. Each core promoter was only tested with its' cognate proximal promoters: Ef1α with Ef1α-proximal and 10xNFκB , and CMV-TCT with CMV-proximal, 100RPU, and 20RPU. TOP: terminal oligo-pyrimidine; SEAP: secreted alkaline phosphatase; SE: stability element; UTR: untranslated region; pA: PolyA.

6.2.2. Synthetic expression elements in combination are associated with a large range of transient SEAP titres

This total library of 42 constructs was transfected by PEI at a 1:5 DNA:PEI ratio into the CHO transient host at $1*10^6$ cells/ml, in 96 deep-well plate shaking culture. 72h post-transfection, VCD was measured by Iprasense as in section 2.4.4.3., and SEAP titre measured by Sensolyte assay, as in section 2.6.2. The measured SEAP titre, and calculated specific productivity, relative to TU1-*SEAP*, are displayed from highest to lowest in Figure 6.3. Figure 6.3a shows that transfection of 19 constructs was associated with significant increases in SEAP titre compared to TU1-*SEAP*, with a densely-populated range from 1.11-fold with 100RPU-TCT-TOP36-*SEAP*, up to 5.23-fold with 10xNFκB-TOP36-*SEAP*-SE20. Five constructs were associated with significantly decreased titre, down to 0.82-fold with TCT-TOP2-*SEAP*. Similarly in Figure 6.3b, cells transfected with 19 constructs exhibit significantly increased specific productivity, from 1.08-fold with TCT-TOP2-*SEAP*-SE20, up to 5.98-fold with 10xNFκB-TOP36-*SEAP*-SE20. Five constructs were associated with decreases in specific productivity, down to 0.81-fold with TCT-TOP2-*SEAP*. SEAP titre and specific productivities measured in this experiment are significantly correlated ($P<0.0001$, $R^2=0.99$). These data suggest that an expansive range of expression strengths are made attainable by the combined use of the synthetic elements identified in this thesis.

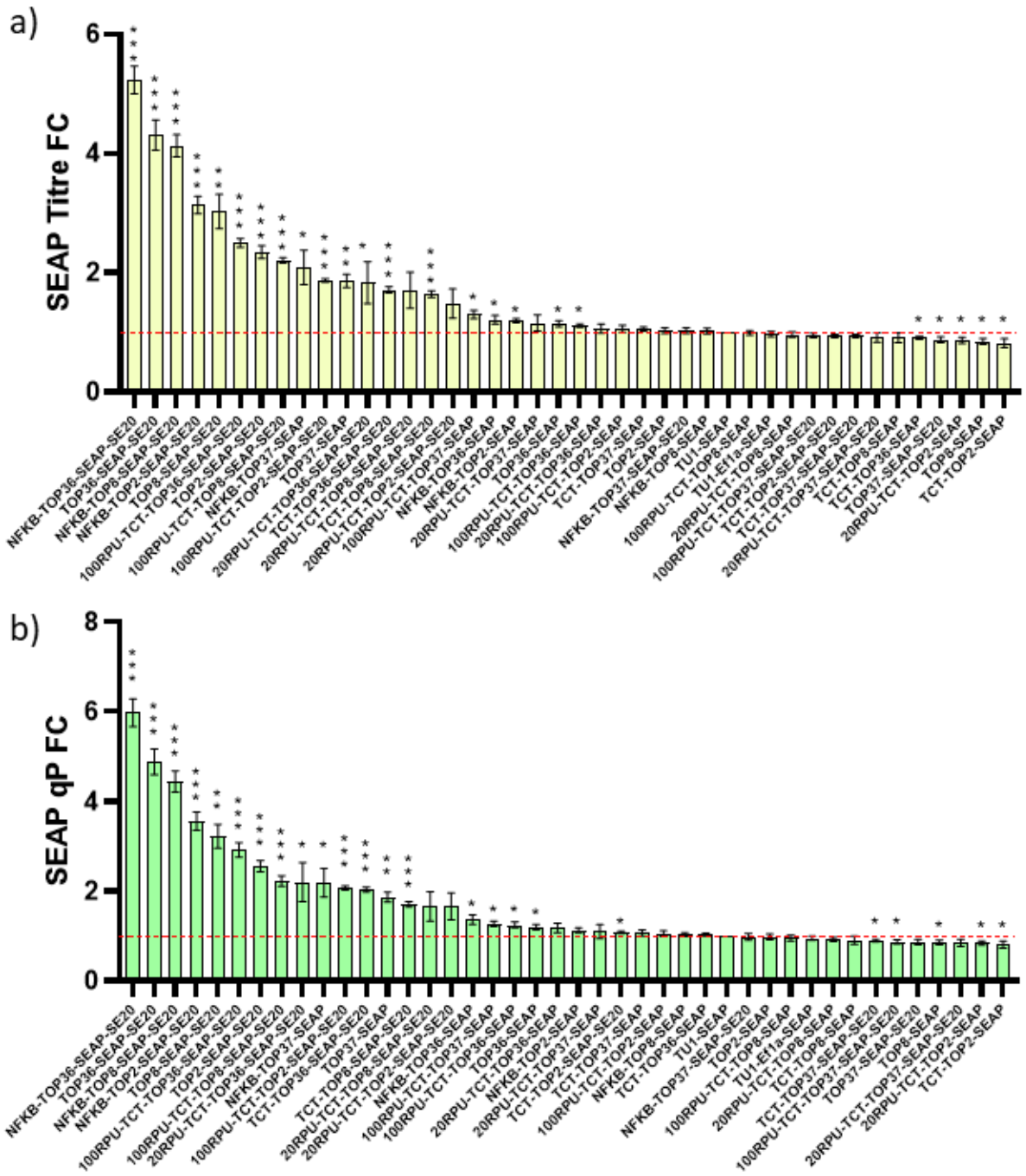


Figure 6.3 – Multielement SEAP library screening. A) SEAP titres of 0.82-fold to 5.23-fold relative to TU1-SEAP are observed across the library, with 19 constructs signficantlty (P<0.05) increasing titre, and 5 significantly decreasing it. B) SEAP specific productivities of o.81-fold to 5.98-fold are observed across the library, with 19 constructs signficantlty (P<0.05) increasing specific productivity, and 5 significantly decreasing it. All synthetic promoter sequences referenced can be found in section 2.13.5. and Table of this thesis, 5'TOP sequences in section 2.13.4. and Table 5.1, and stability elements in section 2.13.3 and Table 4.2. Results presented are of technical duplicates from four biological replicates. FC: fold-change; qP: specific productivity.

To examine whether SE20 performs the function of titre stimulation consistently across different constructs, the SEAP titre observed with each construct from the previous experiment, with and without SE20 are shown in Figure 6.4. The addition of SE20 was associated with a significant increase in titre with 14 out of 20 constructs, and a non-significant increase in titre from TCT-TOP2-*SEAP*, from 0.82-fold to 0.94-fold. Interestingly, the addition of SE20 decreases the titre of every construct containing TOP37, with 2 out of 5 being significant. These data suggest that SE20 generally stimulates the expression of multiple SEAP-containing constructs, but can be rendered suppressive by sequence-specific effects.

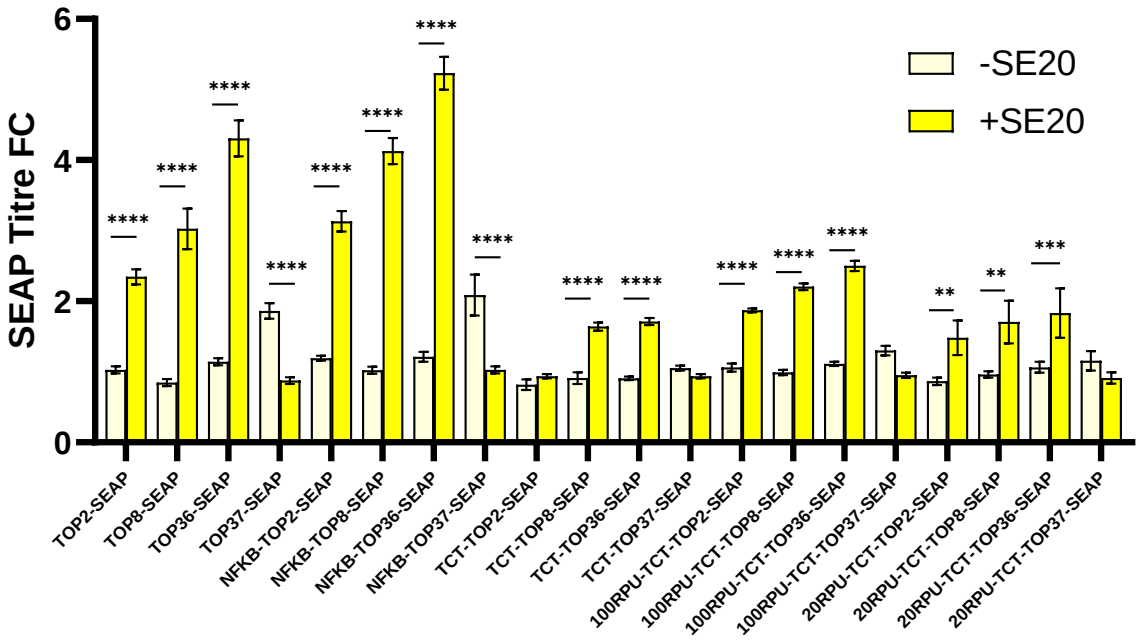


Figure 6.4 – SE20 enhances the titre of all SEAP constructs, apart from those bearing TOP37. 14/20 constructs' SEAP titre is significantly (P<0.05) activated by addition of SE20, and 2 significantly decreased, both of which contain TOP37. All synthetic promoter sequences referenced can be found in section 2.13.5. and Table of this thesis, 5'TOP sequences in section 2.13.4. and Table 5.1, and stability elements in section 2.13.3 and Table 4.2. Results presented are of technical duplicates from four biological replicates. FC: fold-change; SE: stability element.

To identify whether the 10xNFκB proximal promoter consistently increases titre when placed upstream of the Ef1α core promoter, the titre fold change imparted by all constructs containing the Ef1α proximal promoter compared to 10xNFκB , from the experiment described above, are shown in Figure 6.5. In all 8 such constructs, 10xNFκB was associated with a higher titre than the Ef1α proximal promoter, 5 doing so significantly. A consistent ratio is seen between 10xNFκB /Ef1α constructs, with an average of 1.20 and standard deviation of 0.096. These data show that the 10xNFκB proximal promoter is consistently capable of increasing SEAP titre, compared to the unmodified Ef1α promoter, within the context of an Ef1α core.

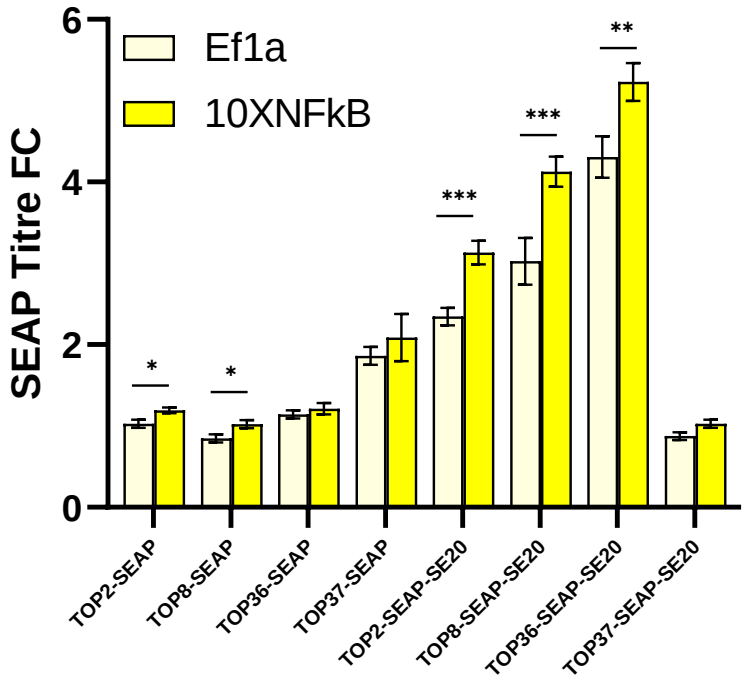


Figure 6.5 – The 10xNFκB proximal promoter consistently increases SEAP titre from the Ef1α core promoter. 5/8 constructs give significantly (P<0.05) higher titre with the 10xNFκB proximal promoter than modified Ef1α. All synthetic promoter sequences referenced can be found in section 2.13.5. and Table of this thesis, 5'TOP sequences in section 2.13.4. and Table 5.1, and stability elements in section 2.13.3. and Table 4.2. Results presented are of technical duplicates, from four biological replicates. FC: fold-change.

It was assessed whether the CMV-TCT promoter core developed in Chapter 5 of this thesis is receptive to predictable control of transcription by the synthetic proximal promoters developed in Brown *et al.*, 2014. All constructs with this core from the previously-described experiment are presented together in Figure 6.6, showing the difference in SEAP titre associated with the CMV, 20RPU, and 100RPU proximal promoters. No significant change is measured between the CMV and 20RPU proximal promoter in any construct. However, the 100RPU promoter is associated with a significant increase in titre compared to CMV in 6 of the 8 constructs. These data suggest that whilst the CMV-TCT core does not produce the pre-defined strengths of transcription granted by proximal promoters designed for the unmodified CMV core, expression from it may still be stimulated by the stronger proximal promoters.

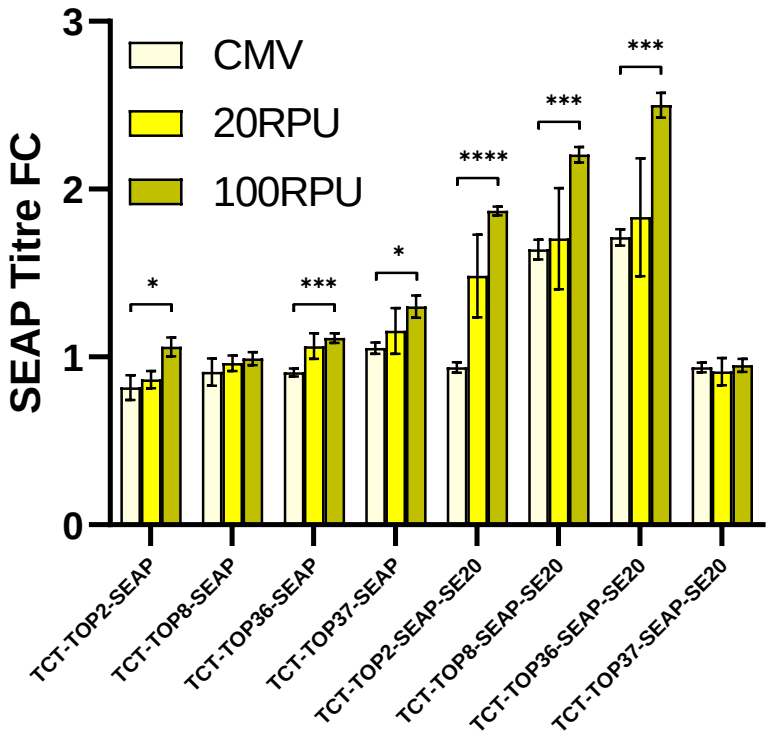


Figure 6.6 – The 100RPU proximal promoter can stimulate expression from CMV-TCT cores. The SEAP titre from 6/8 constructs is significantly (P<0.05) higher with the 100RPU than the CMV proximal promoter. No significant changes are observed between the CMV and 20RPU proximal promoters. All synthetic promoter sequences referenced can be found in section 2.13.5. and Table of this thesis, 5'TOP sequences in section 2.13.4. and Table 5.1, and stability elements in section 2.13.3. and Table 4.2. Data presented are of technical duplicates from four biological replicates. FC: fold-change.

To assess the modularity of the tools used in this assay, Table shows the fold changes in titre associated with the constituent parts of the highest-expression constructs for each promoter core variant. The expected fold-change imparted by a modular combination of these elements and the observed fold change in titre imparted by the construct containing all three are both shown. In both contexts, the observed effect outperforms the fold-change predicted by the multiplied product of each individual element. This indicates firstly that a bottleneck in SEAP expression has not yet been reached, which would otherwise compromise the modularity of the tools. Secondly, it shows that there is little to no functional redundancy between the tools. Indeed, their ability to collectively outperform the expected product of their effects may suggest co-activation of one another. For instance, an increase in mRNA stability imparted by SE20 may further facilitate the activation of translation given by a 5'TOP sequence.

Ef1a Core		CMV-TCT Core	
10xNFκB	1.25	100RPU	1.67
TOP36	1.20	TOP36	1.11
SE20	2.47	SE20	1.15
Predicted Modular Fold Change	**3.72**	**Predicted Modular Fold Change**	**2.13**
Observed Fold Change	**5.51**	**Observed Fold Change**	**2.50**

Table 6.2 – Comparing the individual effect on SEAP titre of the most efficacious single tools, within both an Ef1α and CMV-TCT core context, with their predicted combined effect in a modular system, calculated by the product of all their individual effects, and their observed effect when combined together.

6.3. Discussion

Over the course of this thesis, two tools have been developed for use in controlling titre of biopharmaceutical recombinant proteins: the stability element, and 5'TOP motifs, with the addition of a modified CMV core promoter to accommodate the latter. In combination with the previously designed tool of synthetic promoters, a vector library was constructed that achieved a densely-populated range of SEAP titre from 0.81-fold to 5.23-fold change relative to the industry-standard TU1-*SEAP* vector. This is firstly notable because it outperforms the fold-change in SEAP expression achieved in Brown *et al.*, 2019 by vector elements alone, before a bottleneck was encountered that had to be relieved by cellular engineering, at around a 3-fold increase in titre (Brown *et al.*, 2019). This may be due to differences in transfection

platform, as the lipofectamine used in Brown *et al.,* 2019 is well documented to carry a higher transfection efficiency than the more industrially-relevant PEI used in this study (Hacker *et al.,* 2013; Wang *et al.,* 2018), which could lead to differences in expression dynamics. This difference could also be explained by alterations in the process targeted by the respective studies. In Brown *et al.,* 2019, transcription is targeted by synthetic promoters, translation elongation by codon optimisation, and folding/secretion by signal peptides. This study shares the transcriptional targeting, but also targets translation initiation with 5'TOP motifs, and mRNA half-life with the stability element. These two processes could form a more restrictive bottleneck in SEAP expression that those targeted in Brown *et al.,* 2019. However, the effector protein which best relieves this bottleneck in Brown *et al.,* 2019 is *ATF6*Ac, a transcription factor that stimulates production of foldases and chaperones, implicating a bottleneck further downstream. More insight could be gained from this by further integration of synthetic elements from this study with previously developed tools, for example measuring the effects of 5'TOPs and signal peptides together.

The 5'TOP sequences used in this screen display some consistency, and therefore predictability, in their effect on expression, but are also clearly affected by other sequence features. TOP37, for instance is the highest-performing 5'TOP motif in terms of titre in all constructs lacking SE20. In all constructs containing SE20, it produces the lowest titre of all the TOP variants. TOP8, meanwhile, produces no significant difference in titre compared to TOP2 in all constructs lacking SE20, but produces a higher titre in all construct variants containing SE20, significantly ($P<0.05$) so in 3 out of 5 cases. Figure 6.4 shows that addition of SE20 consistently increases SEAP titre, with the exception of constructs containing TOP37. As the stability elements were identified by a descriptive study, and no mechanistic basis of their action is known, it is difficult to infer the cause of this sequence-specific interaction. No complementarity is found between TOP37 and SE20 sequence. In the absence of a mechanistic explanation, the best conclusion to draw is that the ability of 5'TOP motifs to influence recombinant protein titre may be sensitive to 3'UTR sequence elements. Excluding TOP37, the addition of SE20, whilst consistently increasing titre, has a stronger effect on constructs with an Ef1a, as opposed to a CMV-TCT promoter core. As shown in Figure 6.7, the average titre increase given by addition of SE20 to a vector containing an Ef1α core promoter is 3.34-fold,

significantly higher than the 1.81-fold increase with CMV-TCT, implying that transcripts expressed from an Ef1α core benefit to a greater extent from extension of their half-life.

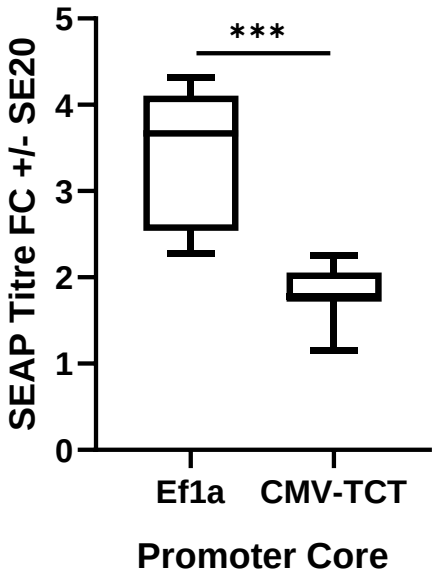


Figure 6.7 – Addition of SE20 has a greater positive effect on Ef1α than CMV-TCT core constructs. The mean titre fold change given by addition of SE20 to Ef1α core constructs is 3.34-fold, significantly (P=0.0001) higher than those with a CMV-TCT core. Due to its' specific negative interaction with SE20, TOP37 constructs are excluded from this dataset. The SE20 sequence can be found in section 2.13.3. and Table 1. Results displayed are of technical duplicates, from four biological replicates. FC: fold-change; SE: stability element.

Figure 6.5 shows that the 10xNFκB proximal promoter retained a consistent ability to slightly increase expression from the Ef1α core promoter in all SEAP constructs, these data suggesting it as a predictable tool for leveraging this effect. More mixed results were gained from the CMV-TCT core promoter, shown in Figure 6.6. The 100RPU proximal promoter did show a relatively consistent ability to upregulate expression compared to the CMV proximal. Given that its design includes a number of highly-responsive TFREs, its ability to increase transcription is not surprising. However, the 20RPU proximal did not align to its' defined function, to produce significantly less transcriptional activity than CMV (Brown *et al.*, 2017). Together, these data suggest that whilst the synthetically engineered CMV-TCT core is more responsive to synthetic promoters developed for the CMV core than Ef1α, it does not possess the ability to entirely reconcile the precise synthetic control of transcription given by these synthetic promoters with control of translation by 5'TOP motifs.

Whilst each tool presented here displays different extents of predictability, notwithstanding the question of whether the stability element will similarly stimulate the expression of alternative proteins, it can be are argued that they, as a collective, render expression strength and characteristics titratable. Through them, titres from 0.82-5.32 fold over controls are made

accessible, through the diverse mechanisms of controlling transcription, translation initiation, and mRNA stability.

6.3.1. Future Work

Whilst the synthetic molecular biology tools presented in this thesis are powerful, modular, and titratable when considered together, more work would be required to develop them into truly predictable tools. In this section, I will present several questions that need to be answered for each element to be developed in this way, and propose tests to each end.

The most pertinent question in relation to the stability element is whether SE20 represents a one-size fits all solution for extension of mRNA half-life. Whilst it retained its positive influence on titre, the extend of this stimulation was significantly different between the Ef1α and CMV-TCT core promoter, suggesting a measure of dependency on its' context. Secondly, the other tool developed in this thesis, 5'TOP motifs, have different optimum sequences depending on the protein being expressed. SE20 has only been shown to stimulate SEAP titre, and it is unknown whether a different stability element may be required for alternative recombinant proteins. These questions can be addressed by wider screening of stability elements, under different promoters (e.g. SV40), with different 5'UTRs and 3'UTRs, and with different recombinant proteins (e.g. EPO, GFP, an scFP).

The specific titre-inhibitory interaction between TOP37 and SE20 is of note, due to its implication of specific 5'TOP motifs and stability elements' potential to interfere with one another. To understand and predict this, the expression characteristics of this interaction should first be understood. Expression cassettes both with and without TOP37 and SE20 could be analysed by RT-qPCR, to understand their combinatorial effect on mRNA processing. Further investigation might involve measurement of their mRNA stability by inhibition of transcription, or more direct inference of their translation efficiency by polysome profiling. Alternatively, to determine the sequence elements necessary for the interaction, a series of mutational studies could be undertaken, investigating which sequence elements abolish the effect when altered.

The primary questions to be answered concerning 5'TOP motifs as a tool relate to the nuanced results from the EPO screen presented in Chapter 5. Firstly, is the bottleneck in expression which prevented most 5'TOP motifs from increasing the titre of codon-optimised EPO

conserved across difficult to express molecules? Secondly, is the observed effect on mRNA, with 5'TOP motifs increasing the abundance of an optimised transcript, and having no effect on the non-optimised mRNA, conserved across difficult to express molecules? Thirdly, is this effect accompanied by induction of the UPR with optimised, high-production transcripts, as hypothesised in Chapter 5? These questions could be addressed by testing the effect on 5'TOP motifs on the expression of optimised and deoptimised transcripts of proteins known to be easy to express (e.g. GFP, SEAP) and difficult to express (e.g. an aggregation-prone scFP), coupled with RT-qPCR analysis of their transcript abundance. To assess the UPR hypothesis, accumulation of UPR-associated proteins such as *XBP1S* could be measured in parallel, by western blot (Johari *et al.*, 2015).

Studies from thesis show that the effect on expression of particular 5'TOP sequences is highly-context dependent. Changes in effect have been noted in different vector contexts, with different recombinant proteins, in different culture conditions, and with different 3'UTR sequences downstream. If control of recombinant protein expression by 5'TOP motifs is to be rendered predictable, a model must be constructed which incorporates all of these conditions. This is difficult to achieve, and there is currently no published literature or datasets addressing the effect of 5'TOP motifs on recombinant protein expression. To begin approaching this problem, a database could be built, noting the conditions of transient expression processes, e.g. the cell type, vector construct, recombinant protein, feed scheme, growth profile, etc. This database could be analysed by a dimensionality reduction algorithm, such as principal component analysis, to identify trends in the data. For instance, TOP-X may tend to be a high-producer of easy to express, non-glycosylated proteins in a 7 day process under a day 0 feed scheme, whilst TOP-Y may tend to stimulate production of difficult to express proteins, in a 14-day fed-batch culture where the cells are kept in stationary phase. If such patterns do emerge, they could be used to make predictions of appropriate 5'TOP motifs to use for future expression of biotherapeutic proteins. In support of this proposal, replicability of the effect of 5'TOP motifs with some similar conditions has been demonstrated in this thesis, by the correlation of titre fold changes given by TOP sequences to SEAP using a 7-day industrial transient process, and a 72h in-house transient expression.

6.3.2. How to Use This Toolkit

In Chapter 1 of this thesis, a workflow was proposed for the rationalisation of CLD process design, by an iterative process of transient screening and targeted application of synthetic biology tools. In it, analysis of expression after round one of transient screening of a new molecule would be used to inform synthetic design of an expression system for round two, allowing for the use of effectively targeted synthetic molecular biology tools.

In this thesis, two such viable molecular biology tools have been developed: 5'TOP motifs for control of translation, and stability elements for control of mRNA stability. In an ideal case, the effect of 5'TOP motifs and stability elements could be user defined: predictably modifying expression characteristics to desired parameters. For instance, if the light chain of a mAb needed to be expressed at a 4:1 stoichiometry to the heavy chain, 5'TOP motifs could be assigned to grant its' transcript a 4-fold greater rate of translation. At their current stage of development, neither 5'TOP motifs nor stability elements represent tools with the necessary predictability for this system. This imposes a small change: for optimisation of their respective target functions, multiple solutions should be screened, as shown in Figure 6.8. Fortunately, the relative simplicity of both of their implementation, incorporating between 5-34nt into a synthetic vector, render their production and testing in a high-throughput system viable. The large range of expression strengths they bestow demonstrate that they represent a powerful synthetic, titratable, modular, if not predictable, molecular biology tool.

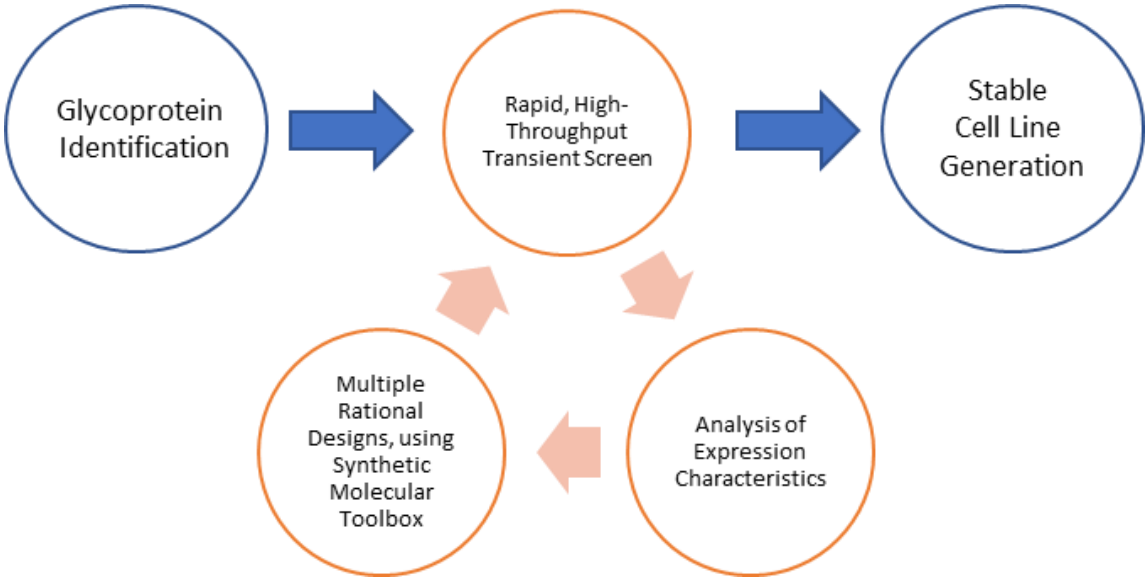


Figure 6.8 - A schematic of the generalised use of the molecular biology tools developed in this thesis. Whilst a single predictable user-defined solution is not yet possible, parallel testing of multiple tool variants may still be deployed in this system, to reduce R&D risk, and ensure a high success rate in cell line development for glycoprotein production.

6.4. Conclusion

Synthetic molecular biology tools present a powerful solution, to rationalising and de-risking the design and implementation of CLD for biopharmaceutical production. A lack of tools influencing mRNA dynamics have limited the scope of this toolbox.

Investigation of novel RNA-binding effector proteins *C1orf35* and *HuR* as such tools gave a set of initially promising results and interesting targets for future research. The association *C1orf35* with stimulation of cell growth is consistent with the limited knowledge available on it. Similarly, observation of *HuR* stabilising mRNA and subsequently enhancing specific productivity are consistent with literature, and may imply the potential use of ARE-binding proteins in biopharmaceutical expression. However, the incongruities inherent to effector gene screening in cell lines between transient and stable expression systems rendered these tools inappropriate for a system predicated on transient-to-stable transferability of outcomes.

The 3'UTR stability elements, specifically SE20, have been shown and validated to increase production of the SEAP model recombinant protein across several vector contexts and culture conditions. Moreover, its' mechanism of action has been confirmed as extension of mRNA half-life. This use of the elements discovered in Oikonomou et al., 2014 is entirely novel. It

implies both the value in investigating the enhancement of biopharmaceutical titre via the 3'UTR, and the potential power of omics-forward approaches to element discovery, even when a molecular mechanism isn't apparent. The immediate next step should be to investigate whether SE20 presents a more general solution to mRNA stabilisation and titre enhancement, or whether its' effects are more product, or context-dependent.

5'TOP motifs have been shown to exert a high degree of control over the expression of recombinant protein. This has been validated across a number of vector contexts, culture platforms, and recombinant proteins, including an industry-standard transient screen of an scFP biotherapeutic molecule. Though 5'TOP motifs are relatively well characterised, and related molecular mechanisms such as mTOR overexpression and rapamycin treatment have been utilised in biopharmaceutical production, this direct use of 5'TOP motifs within expression vectors is entirely novel. The next step in the development of this tool will be to develop an understanding of the context-dependence of their effect on expression, thus making possible their predictable integration into a rationalised CLD process.

When combined, together with synthetic promoter tools for control of transcription, these novel tools enable a wide array of expression strengths, enabling a range of transient SEAP titres from 0.82-fold to 5.23-fold compared to an industry-standard vector, and behave in a synergistic modular fashion. Whilst a number of questions and developments will need to be solved to render each fully predictable in new contexts, this body of work has contributed two powerful, valuable synthetic biology tools toward the rationalisation of biopharmaceutical development and production.

6.5. References

Brown, A.J., Gibson, S.J., Hatton, D., and James, D.C. (2017). In silico design of context-responsive mammalian promoters with user-defined functionality. Nucleic Acids Res. 45, 10906–10919.

Brown, A.J., Gibson, S.J., Hatton, D., Arnall, C.L., and James, D.C. (2019). Whole synthetic pathway engineering of recombinant protein production. Biotechnol. Bioeng. 116, 375–387.

Hacker, D.L., Kiseljak, D., Rajendra, Y., Thurnheer, S., Baldi, L., and Wurm, F.M. (2013). Polyethyleneimine-based transient gene expression processes for suspension-adapted HEK-293E and CHO-DG44 cells. Protein Expr. Purif. 92, 67–76.

Johari, Y.B., Estes, S.D., Alves, C.S., Sinacore, M.S., and James, D.C. (2015). Integrated cell and process engineering for improved transient production of a "difficult-to-express" fusion protein by CHO cells. Biotechnol. Bioeng. 112, 2527–2542.

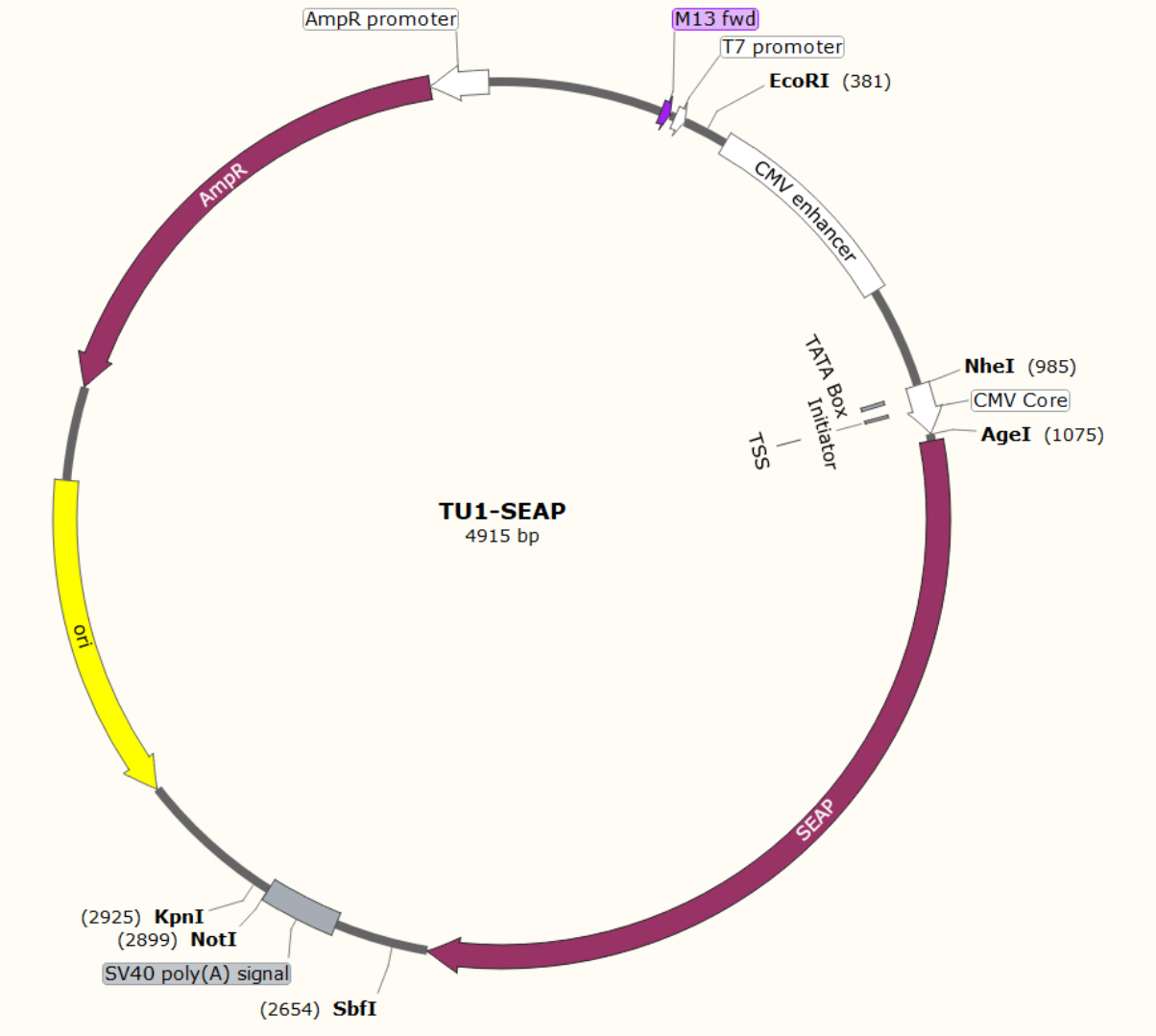
Wang, W., Guo, X., Li, Y. mei, Wang, X. yin, Yang, X. jun, Wang, Y. fang, and Wang, T. yun (2018). Enhanced transgene expression using cis-acting elements combined with the EF1 promoter in a mammalian expression system. Eur. J. Pharm. Sci. 123, 539–545.

Yeo, J.H.M., Ho, S.C.L., Mariati, M., Koh, E., Tay, S.J., Woen, S., Zhang, P., and Yang, Y. (2017). Optimised Selection Marker and CHO Host Cell Combinations for Generating High Monoclonal Antibody Producing Cell Lines. Biotechnol. J. 12, 1–11.

Appendix A: Vector Maps and Construction

TU1

The basal TU1 vector is shown below, expressing SEAP as its' CDS. The unique restriction sites, intended for cloning of vector elements in the transcription unit vector system (Patel *et al.*, in press) are highlighted. Alternative genes were inserted by restriction/ligation cloning at the AgeI/SbfI sites.

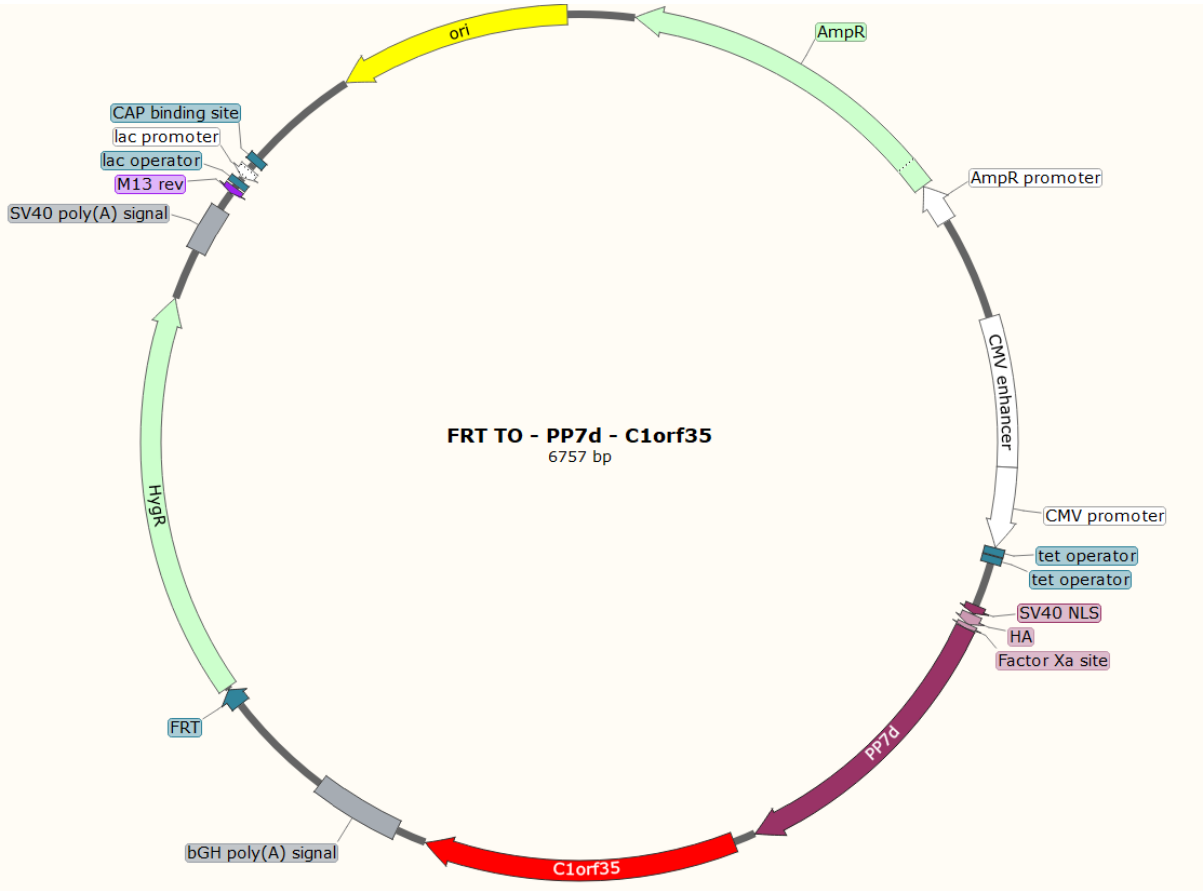


TU1-*SEAP*-6xPP7bs

TU1-*SEAP*-6xPP7bs, shown below, was generated by insertion of 6xPP7bs into the 3'UTR of TU1-*SEAP*, by Gibson Assembly.

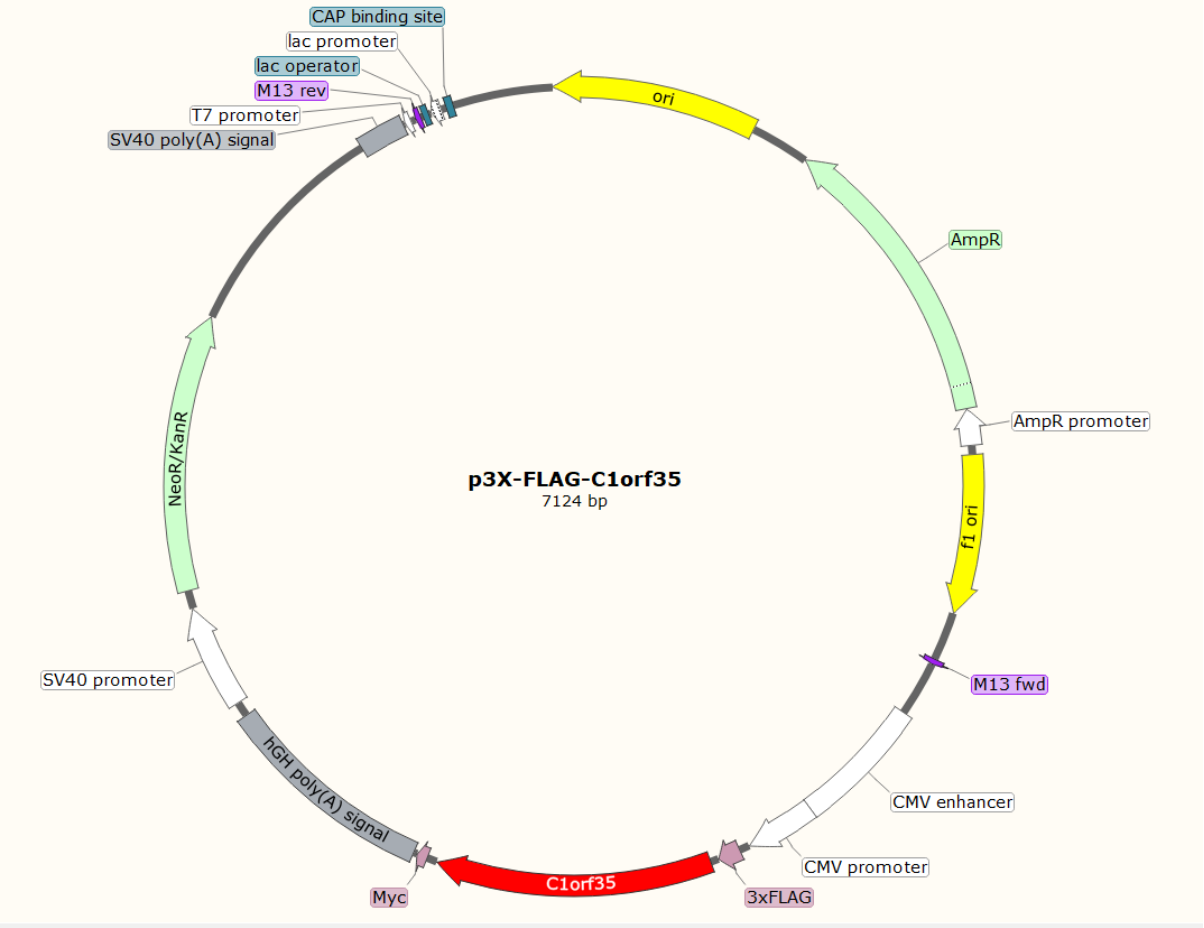
FRT TO – PP7d Fusion

Shown below is the FRT TO-PP7d-fusion expression vector, in this case encoding the PP7d-*C1orf35* fusion. This vector was first generated by insertion of PP7d, plus the SV40 NLS and HA tag, into the FRT TO vector, by Gibson Assembly. Genes were then inserted downstream of PP7d by Gibson Assembly.



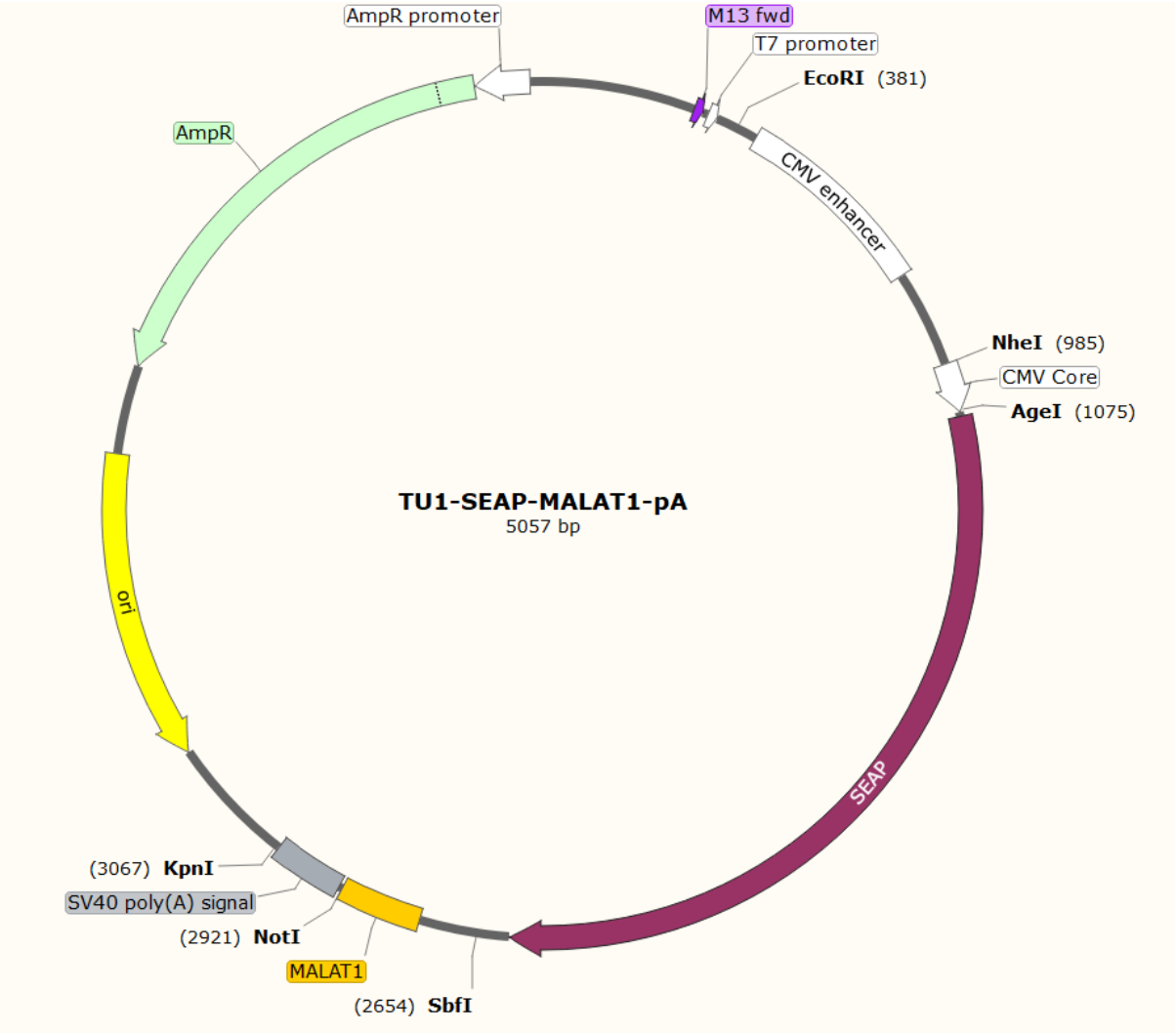
P3X-FLAG-*C1orf35*

To generate this construct, the *C1orf35* CDS was inserted downstream of the 3xFLAG peptide in the p3X vector, by Gibson Assembly.



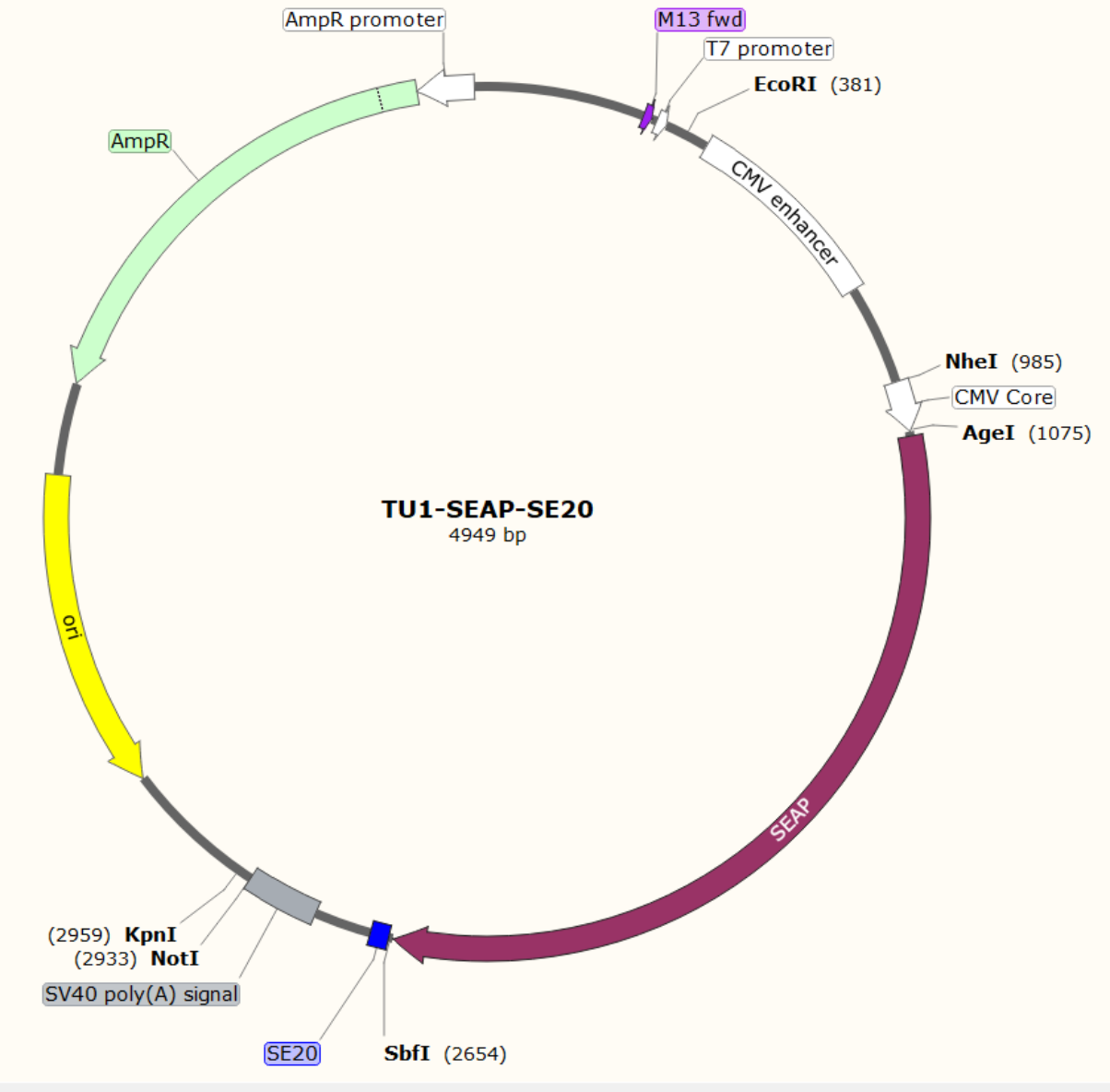
TU1-*SEAP*-Triple Helices

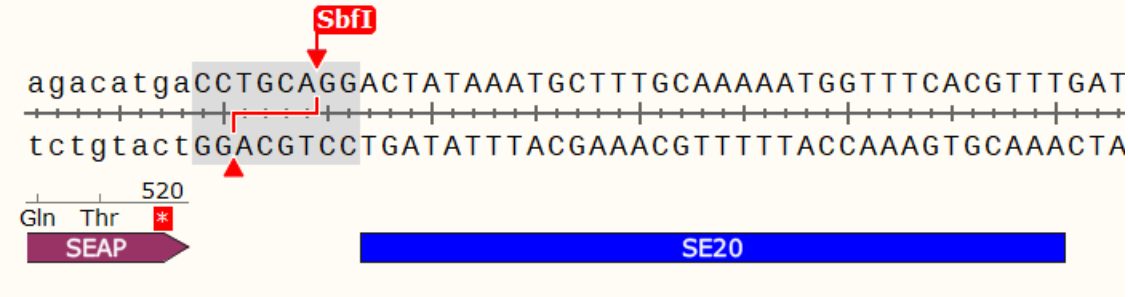
The triple helices library was constructed by insertion of triple helices (in the example shown, MALAT1) downstream of the TU1-*SEAP* 3'UTR, in place of the SV40 PolyA, by SbfI/NotI restriction/ligation cloning. For constructs containing a PolyA, the SV40 PolyA was inserted downstream of the triple helix, by NotI/KpnI restriction/ligation cloning.



TU1-*SEAP*-Stability Elements

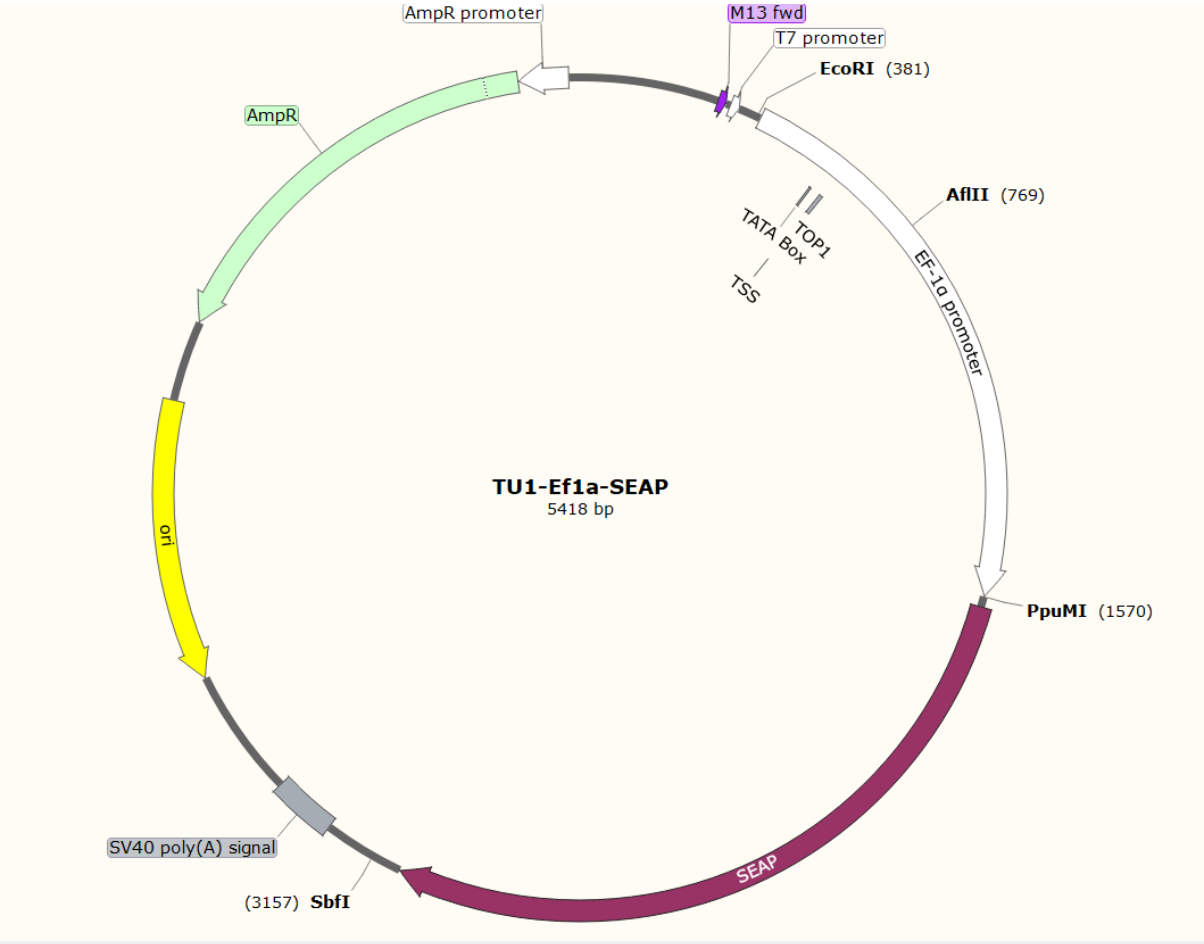
Stability elements were inserted directly downstream of the SEAP CDS and SbfI restriction site, in the TU1-*SEAP* vector, by site-directed mutagenesis.

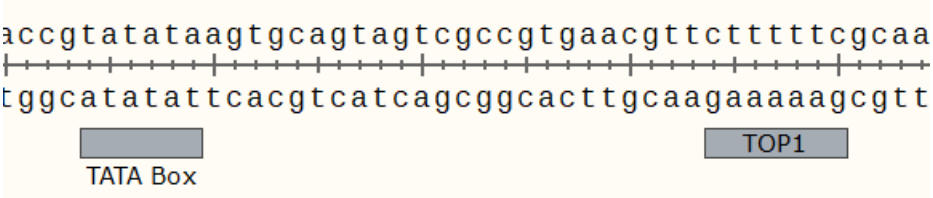




TU1-Ef1α

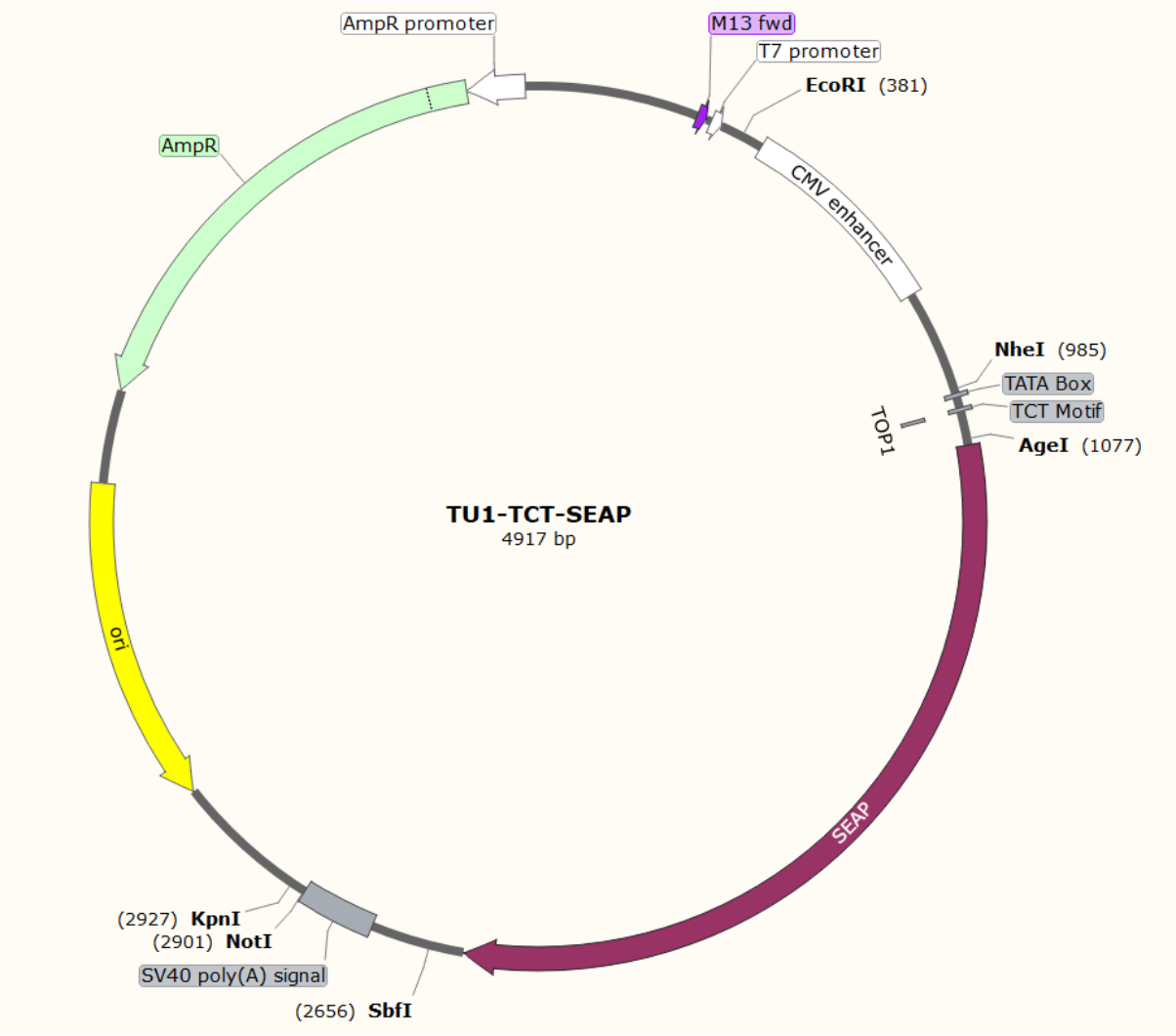
The TU1-Ef1α-*SEAP* expression vector (also referred to as TOP1-*SEAP*) was generated by insertion of the Ef1α promoter in place of the CMV promoter and 5'UTR in the TU1-*SEAP* vector, by EcoRI/PpuMI restriction/ligation cloning. Alternative genes were inserted by PpuMI/SbfI restriction/ligation cloning. 5'TOP sequences were inserted in place of TOP1, shown below, by site-directed mutagenesis. The Ef1α region upstream of the TATA box was designated as the proximal promoters, and synthetic proximal promoters inserted by EcoRI/AflII restriction/ligation were cloned into this site, replacing the Ef1α proximal promoter.

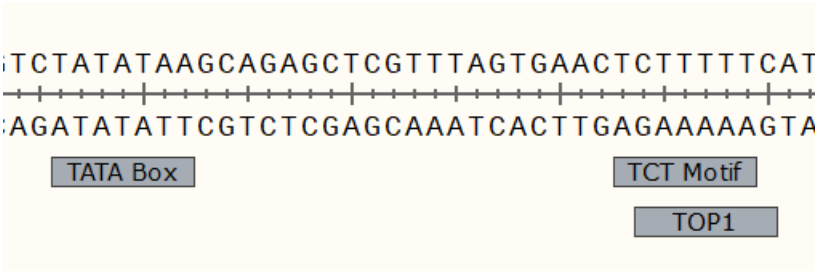




TU1-TCT-*SEAP*

The TU1-TCT expression vector was generated by insertion of the TCT motif, in place of the Inr motif, in the CMV core of TU1-*SEAP*. In the spacing variants, base pairs upstream of the TCT motif in the Ef1α promoter were inserted upstream of the TCT motif in this core. Taking the C_{+1} as the start of the 5'TOP motif, different sequences were inserted by site-directed mutagenesis.





MGVT

To generate MGVT-Ef1α constructs, TU1-Ef1α transcription units (from BsaI sites upstream of the Ef1α promoter, and downstream of the SV40 PolyA) were inserted into the MGVT vector by Golden Gate cloning.