



The
University
Of
Sheffield.

**Design and Characterisation of Multi-Gene
Expression Vectors for CHO Cell
Engineering**

Yash Dhananjai Patel

Thesis Submitted for the Degree of Doctor of Philosophy

November 2018

Department of Chemical and Biological Engineering

University of Sheffield

Declaration of Originality

I, Yash Dhananjai Patel, in accordance with the University regulations, hereby declare that this thesis was solely composed by myself and that the data presented is the product of my own efforts. Results generated in collaboration have been clearly stated and acknowledged. This document has not been submitted in part or whole for any other degree or personal qualification.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor David C. James for providing me with the opportunity to study for a PhD, and for the guidance, advice, project direction and supervision throughout the entire process. I would also like to thank Dr. Adam J. Brown for the technical advice and considerable support during the project. Additionally, I am very grateful to MedImmune for funding and facilitating my project. I am particularly thankful to the Cell Line Development team at MedImmune, specifically Dr. Diane Hatton and Dr. Jie Zhu for technical advice and guidance in the thesis writing process. I would also like to acknowledge Dr. Guglielmo Rosignoli for the flow cytometry technical support and data analysis advice.

I would like to thank all the past and present PhD students, postdocs and technicians within the David James lab for all the great support both professionally and personally. I would like to particularly acknowledge Dr. Joe Cartwright for advice on the project and taking the time to help me. Furthermore, thank you to Dr. Katie Syddall, Dr. Joanne Noble-Longster, Dr. Devika Kalsi, Dr. Darren Geoghegan, Dr. Claire Arnall and Dr. Nicholas Barber for all the great times both in and out of the labs, your friendships have made this experience memorable. I would also like to share my appreciation to all my friends at the Sheffield University Rowing Club for giving me a place away from the labs where I can relax and do a sport I enjoy very much. A very special thank you goes to Dr. Géraldine Baekelandt who has been incredible during the final stages of my PhD with her words of wisdom and emotional support to help me get across the line.

Finally, a big thank you goes to my family, who have been my backbone through this process, with their unwavering encouragement, patience and support. Thank you for making my ambitions of attaining a doctorate a reality.

Table of Contents

1. Introduction	1
1.1. The Biopharmaceutical Industry	1
1.2. Overview of the current production technology of biopharmaceuticals	2
1.2.1. Host cell factories.....	3
1.2.1.1. Alternative host cell factories.....	3
1.2.1.2. Mammalian host cell factories	5
1.2.2. Recombinant gene expression technology	6
1.2.2.1. Antibiotic-based selection systems	6
1.2.2.2. Dihydrofolate reductase (DHFR)-based selection system	7
1.2.2.3. Glutamine synthetase (GS)-based selection system.....	7
1.3. Overview of the biopharmaceutical and bioprocessing production platform ..	8
1.4. Demands within industry and limitations of current technology	10
1.5. Overview of synthetic biology	12
1.6. Introduction of multi-gene expression vectors as tool for cell engineering...	13
1.6.1. Current approaches and limitations to cell engineering in mammalian cells	14
1.6.2. Applications of multi-gene expression vectors in a biopharmaceutical environment.....	15
1.6.3. DNA cloning approaches to assemble multi-gene expression vectors ...	16
1.7. Synthetic Genetic Elements	20
1.7.1. Synthetic Promoters	20
1.7.2. Synthetic Transcription Terminators.....	23
1.7.2.1. Polyadenylation signal-mediated termination.....	24
1.7.2.2. Co-transcriptional cleavage element-mediated transcription termination	25
1.7.2.3. Pause element-mediated transcription termination	26
1.7.2.4. Ribozyme-mediated transcription termination.....	27
1.7.2.5. Existing Synthetic Terminator Elements	28
1.7.3. Context dependent synthetic genetic element design.....	28
1.8. Thesis Aims & Overview	31
2. Materials & Methods.....	35

2.1. Synthesis and cloning of transcription unit and multi-gene expression vectors	35
2.1.1. Gene and oligonucleotide de novo synthesis	35
2.1.2. Restriction digest – ligation plasmid cloning	35
2.1.3. Golden Gate assembly	36
2.1.4. Transformation of plasmid DNA for cloning, multi-gene expression vector construction or DNA amplification	36
2.1.5. Restriction digest colony screen of cloned plasmids	37
2.1.6. Plasmid DNA amplification	37
2.1.7. Modification of synthetic components using PCR	38
2.2. Cell culture methodologies to screen transcription units and multi-gene expression vectors	38
2.2.1. Cell culture revival, maintenance and cryopreservation	38
2.2.2. High throughput 96 well plate-based transient expression	39
2.2.3. Generation of stable pool cell lines using targeted integration	40
2.2.4. Fed-batch overgrow and cell culture characteristic analysis	41
2.3. Fixation of transfected CHO cells	42
2.4. Fluorescent protein detection within transfected CHO cells by flow cytometry	42
2.4.1. BD LSR Fortessa flow cytometer	43
2.4.2. Attune NxT flow cytometer	44
2.4.3. Gating strategy for flow cytometry quantification	45
2.5. Real-time quantitative PCR analysis of transfected CHO cells	48
2.5.1. Cell pellet storage and RNA extraction of transfected CHO cells	48
2.5.2. Reverse transcription of extracted RNA samples	48
2.5.3. Quantitative PCR analysis	49
2.6. Quantification of secreted alkaline phosphatase by absorbance	49
2.7. Quantification of <i>Cypridina</i> luciferase by luminescence	50
3. Design and Characterisation of a Multi-Gene Expression Vector for CHO Cell Engineering	51
3.1. Introduction	52
3.2. Review of Design Attributes for a multi-gene expression vector system	54
3.3. Results	59
3.3.1. Design of a multi-gene expression vector	59

3.3.1.1. Designing 4 bp linkers for controlling gene assembly.....	59
3.3.1.2. <i>In silico</i> design of genetic components for assembly of a multi-gene expression vector.....	60
3.3.1.2.1. Design of the transcription unit.....	60
3.3.1.2.2. Design of recipient vector backbone.....	62
3.3.1.3. Overview of cloning and modification of a multi-gene expression vector.....	64
3.3.2. Testing the cloning technique for assembling multi-gene expression vectors	65
3.3.3. Developing an optimised detection method to quantify expression of three fluorescent proteins	68
3.3.3.1. Optimising the parameters of the flow cytometer for fluorescent protein expression	69
3.3.3.2. Assessing the dynamic range of fluorescent protein quantification by flow cytometry	70
3.3.4. Identifying an optimal DNA load of multi-gene expression vectors to avoid cytotoxicity.....	73
3.3.5. Investigating the gene expression dynamics within a multi-gene expression vector.....	75
3.4. Discussion	81
4. Application of Synthetic Promoters in a Multi-Gene Expression Vector to Stoichiometrically Control Gene Expression in CHO cells.....	87
4.1. Introduction	88
4.2. Results	93
4.2.1. Adapting and cloning of a synthetic promoter library into the transcription unit for multi-gene expression vector construction.....	93
4.2.2. Characterising the modified synthetic promoter library by fluorescent protein expression	96
4.2.3. Developing and assessing assays for characterising positional effects and quantifying titrated gene expression	101
4.2.3.1. Development of a qPCR assay to quantify transcriptional activity	102
4.2.3.1.1. Designing and testing primer efficiency for qPCR.....	102

4.2.3.1.2. Determining the dynamic range and accuracy of transcriptional activity quantification by plasmid DNA titration.....	104
4.2.3.2. Determining the dynamic range and accuracy of fluorescent protein expression by DNA titration	109
4.2.3.3. Comparing gene expression quantification accuracy of qPCR and flowcytometry	113
4.2.4. Characterising combinatorial functionality of synthetic promoters	115
4.2.5. Normalising for variable mRNA dynamics of <i>eGFP</i> , <i>mCherry</i> and <i>tagBFP</i> to quantify transcriptional activity	121
4.2.6. Characterising transcriptional strength-dependent positional effect within a multi-gene expression vector.....	124
4.2.7. Demonstrating titration of gene expression and determining stoichiometric ratio within a multi-gene expression vector	130
4.2.7.1. Construction of 27 different multi-gene expression vector variants	131
4.2.7.2. Quantifying gene expression stoichiometric ratios of the multi-gene expression vector library by qPCR.....	134
4.2.7.3. Quantifying gene expression trends of the multi-gene expression vector library by flow cytometry.....	142
4.2.8. Identifying transcriptional repression trends within the context of a MGEV	147
4.3. Discussion.....	150
5. Proof of Concept Applications of Multi-gene Expression Vectors as a Research Tool	159
5.1. Introduction	160
5.1.1. Design and screening of synthetic transcription terminator elements...	160
5.1.2. Engineering mammalian cells to enhance cellular phenotype and monoclonal antibody production.....	162
5.2. Results	164
5.2.1. Design and screening of synthetic transcription terminator elements using a multi-gene expression vector.....	164
5.2.1.1. Design of the synthetic transcription terminator element library...	164

5.2.1.2. Construction of multi-gene expression vectors to screen the transcription terminator library	170
5.2.1.3. Screening of the transcription terminator library	173
5.2.1.3.1. Screening of polyadenylation elements	174
5.2.1.3.2. Screening of co-transcriptional cleavage elements	177
5.2.1.3.3. Screening of CHO-K1 homologues	179
5.2.1.3.4. Screening of pause and ribozyme elements	181
5.2.2. Enhancing monoclonal antibody production in a stable CHO cell line by co-expressing XBP-1s using a multi-gene expression vector	183
5.2.2.1. Adapting the pExp-Vec-GG recipient vector for compatibility with MedImmune's targeted integration platform	183
5.2.2.2. Design and construction of multi-gene expression vectors co-expressing a DTE monoclonal antibody and XBP-1s.....	184
5.2.2.3. Establishing targeted integrated stable pools expressing the prototype multi-gene expression vector variants.....	188
5.2.2.4. Growth profile of <i>XBP-1s</i> engineered stable pools during a fed-batch overgrow	189
5.2.2.5. Monoclonal antibody expression trend of <i>XBP-1s</i> engineered stable pools during a fed-batch overgrow	192
5.2.2.6. Measuring mRNA levels of <i>XBP-1s</i> within engineered stable pool variants... ..	195
5.3. Discussion	196
5.3.1. Design and screening of a transcription terminator library using a multi-gene expression vector	197
5.3.2. Application of multi-gene expression vectors to stably engineered CHO cells for enhancing biopharmaceutical production	202
6. Conclusion and Future Work	207
6.1. Conclusion.....	207
6.1.1. Design and characterisation of a multi-gene expression vector for CHO cell engineering	207
6.1.2. Application of synthetic promoters in a multi-gene expression vector to titrate gene expression at stoichiometric ratios in CHO cells	208

6.1.3. Proof of concept applications of multi-gene expression vectors as a research tool	211
6.2. Future Works	212
6.2.1. Improved design of the multi-gene expression vector for CHO cell engineering	213
6.2.2. Further characterisation of gene expression dynamics within a multi-gene expression vector	214
6.2.3. Achieving modularisation within a polycistronic cassette of a multi-gene expression vector	215
6.2.4. Validating the application of multi-gene expression vectors for stable CHO cell engineering	216

List of Figures

Figure 1-1: Recombinant protein expression summary.	2
Figure 1-2: Biopharmaceutical production workflow.....	8
Figure 1-3 - Approaches to clone multiple DNA fragments.....	18
Figure 1-4: The formation of the pre-initiation complex during transcription.	21
Figure 1-5: Structure of a synthetic promoter.	22
Figure 1-6: Polyadenylation signal-mediated termination.	25
Figure 1-7: Co-transcriptional cleavage (CoTC) element-mediated transcription termination.	26
Figure 1-8: Pause element-mediated transcription termination.	27
Figure 2-1: Gating strategy for singularly expressed recombinant fluorescent proteins.	46
Figure 2-2: Gating strategy for expression of multiple recombinant fluorescent proteins.	47
Figure 3-1: Golden gate assembly reaction.....	56
Figure 3-2: Excitation and emission spectra for eGFP, mCherry and tagBFP.	58
Figure 3-3: Control of transcription unit assembly during multi-gene expression vector construction.	60
Figure 3-4: Standardised transcription unit layout.....	62
Figure 3-5: pExp-Vec-GG recipient vector backbone for multi-gene expression vector assembly.	63
Figure 3-6: Genetic elements and layout of transcription unit encoding for fluorescent proteins.	65
Figure 3-7: Assembly order of three transcription units for multi-gene expression vector testing.	66
Figure 3-8: Successful assembly and expression of three transcription units and pExp-Vec-GG backbone plasmid.	67
Figure 3-9: Removing false positive tagBFP signal during eGFP detection.	69
Figure 3-10: Dynamic range of eGFP, mCherry and tagBFP quantification by flow cytometry.....	71
Figure 3-11: Differential level of fluorescent protein detection in Medi-CHO.....	72
Figure 3-12: Growth characteristics of transiently expressed multi-gene expression vector variants.	74

Figure 3-13: Recombinant gene expression dynamics within a multi-gene expression vector.	77
Figure 3-14: Single cell distribution of positional effect for three multi-gene expression vector variants.	80
Figure 4-1: Transcription unit composition for assessing synthetic promoter activity.	93
Figure 4-2: PCR modification of synthetic proximal enhancer element.....	94
Figure 4-3: Cloning of synthetic promoter proximal elements into the transcription unit.....	96
Figure 4-4: Quantifying synthetic promoter performance within a transcription unit.	99
Figure 4-5: Schematic summarising the transfection conditions to analyse the dynamic range of detection of qPCR and flow cytometry.....	101
Figure 4-6: Testing primer efficiency for quantitative PCR.	103
Figure 4-7: Investigating the dynamic range of qPCR-mediated detection by titrating plasmid DNA.....	105
Figure 4-8: Relative promoter activity of eGFP, mCherry and tagBFP by qPCR analysis.	107
Figure 4-9: Determining dynamic range and sensitivity of flow cytometer –mediated detection by titrating plasmid DNA.	110
Figure 4-10: Relative promoter activity of eGFP, mCherry and tagBFP by flow cytometry analysis.....	112
Figure 4-11: Comparing relative fold change of qPCR and flow cytometer-mediated detection.	114
Figure 4-12: Measuring synthetic promoter self-interaction by flow cytometry.	116
Figure 4-13: Measuring dual synthetic promoter cross-interaction by flow cytometry.	118
Figure 4-14: Measuring triple synthetic promoter cross-interaction by flow cytometry.	120
Figure 4-15: Differential mRNA dynamics eGFP, mCherry and tagBFP at different transcriptional activities.	122
Figure 4-16: mRNA external calibration curves.	123
Figure 4-17: Quantifying transcriptional activity dependent positional effects within a multi-gene expression vector by qPCR.	125

Figure 4-18: Expression fold change quantifying positional effect from fluorescent protein expression.	128
Figure 4-19: Schematic describing the MGEV configuration.	131
Figure 4-20: Verifying successful construction of the remaining 24 multi-gene expression vector variants.	133
Figure 4-21: Absolute mRNA copies of GS gene within multi-gene expression vector variants.	135
Figure 4-22: Quantified gene expression levels within multi-gene expression vector variants by qPCR – Low strength combination subcategory.	137
Figure 4-23: Quantified gene expression levels within multi-gene expression vector variants by qPCR – Medium strength combination subcategory.	138
Figure 4-24: Quantified gene expression levels within multi-gene expression vector variants by qPCR – High strength combination subcategory.	139
Figure 4-25: Quantified gene expression levels within multi-gene expression vector variants by qPCR – Mixed strength combination subcategory.	140
Figure 4-26: Quantified fluorescent protein expression of multi-gene expression vectors by flow cytometry – Low strength combination subcategory.	143
Figure 4-27: Quantified fluorescent protein expression of multi-gene expression vectors by flow cytometry – Medium strength combination subcategory.	144
Figure 4-28: Quantified fluorescent protein expression of multi-gene expression vectors by flow cytometry – High strength combination subcategory.	145
Figure 4-29: Quantified fluorescent protein expression of multi-gene expression vectors by flow cytometry – Mixed strength combination subcategory.	146
Figure 4-30: Frequency distribution of transcriptional repression within the MGEV library.	148
Figure 4-31: Position specific trends in transcriptional repression within the MGEV library.	149
Figure 4-32: The transcription factor regulatory element (TFRE) composition of the synthetic promoters tested in the multi-gene expression vector library.	156
Figure 5-1: Design schematic summarising the commercially available polyadenylation variants synthesised and tested.	167
Figure 5-2: Design schematic summarising the synthetic polyadenylation library synthesised and tested.	168

Figure 5-3: Design schematic summarising the synthetic co-transcriptional cleavage library synthesised and tested.....	169
Figure 5-4: Design schematic summarising the synthetic pause-mediated transcription termination library synthesised and tested.	170
Figure 5-5: Schematic summarising the multi-gene expression vector used for terminator library screening.	171
Figure 5-6: Restriction digest colony screen of multi-gene expression vector variants containing the control and synthetic polyadenylation elements.....	173
Figure 5-7: Screening of commercially available polyadenylation element within a multi-gene expression vector.	175
Figure 5-8: Screening of synthetic polyadenylation elements within a multi-gene expression vector.....	176
Figure 5-9: Screening of synthetic co-transcriptional cleavage elements within a multi-gene expression vector.	178
Figure 5-10: Screening of CHO-K1 polyadenylation and co-transcriptional cleavage homologous elements within a multi-gene expression vector.	180
Figure 5-11: Screening of synthetic pause and ribozyme elements within a multi-gene expression vector.....	182
Figure 5-12: Restriction digest colony screen verifying successful cloning of hygromycin resistance gene and LoxP site into the pExp-Vec-GG recipient vector.	184
Figure 5-13: Prototype-3 multi-gene expression vector for engineering CHO cells with XBP-1s.	186
Figure 5-14: Restriction digest colony screen verifying successful cloning of prototype multi-gene expression vectors.....	187
Figure 5-15: Viable cell concentration and cell viability of TI stable pools during a 13 day fed-batch overgrow.....	190
Figure 5-16: Integral viable cell density of TI stable pool during a 13 day fed-batch overgrow.....	191
Figure 5-17: Monoclonal antibody expression titres of XBP-1s engineered stable pools during a 13 day fed-batch overgrow.	192
Figure 5-18: Specific productivity of monoclonal antibody expression within XBP-1s engineered stable pools during a 13 day fed-batch overgrow.....	194

Figure 5-19: Fold change of XBP-1s mRNA copies of targeted integrated stable pool variants.	195
Figure A-1: Vector maps showing the transcription units encoding for eGFP, mCherry and tagBFP.	235
Figure A-2: Map showing successful DNA sequencing of pMGEV-GBC.	236
Figure A-3: Gating strategy for singular expression mCherry using the BD LSR Fortessa flow cytometer.	237
Figure A-4: Gating strategy for singular expression tagBFP using the BD LSR Fortessa flow cytometer.	238
Figure A-5: Vector maps showing multi-gene expression vector variants constructed.	240
Figure B-1: Gating strategy for singular expression of eGFP using the Attune NxT flow cytometer.	243
Figure B-2: Gating strategy for singular expression of mCherry using the Attune NxT flow cytometer.	244
Figure B-3: Gating strategy for singular expression of tagBFP using the Attune NxT flow cytometer.	245
Figure B-4: Gating strategy for expression of multiple fluorescent using the Attune NxT flow cytometer.	246
Figure B-5: qPCR Standard curves for eGFP, mCherry, tagBFP and glutamine synthetase for mRNA copy quantification.	247
Figure B-6: Measuring dynamic range of qPCR by quantified mRNA copies.	248
Figure B-7: Measuring dynamic range of flow cytometry by quantified integrated median fluorescent intensity.	249
Figure B-8: Quantifying positional effects within the MGEV library -1.	250
Figure B-9: Quantifying positional effects within the MGEV library -2.	251
Figure B-10: Quantifying positional effects within the MGEV library -3.	252
Figure C-1: Restriction digest colony screen of multi-gene expression vector variants containing the synthetic co-transcriptional cleavage elements.	253
Figure C-2: Restriction digest colony screen of multi-gene expression vector variants containing the synthetic pause and ribozyme elements.	254
Figure C-3: Restriction digest colony screen of multi-gene expression vector variants containing the CHO-K1 polyadenylation and co-transcriptional cleavage homologues, and commercially available polyadenylation elements.	254

Figure C-4: Secreted alkaline phosphatase standard curves.262

List of Tables

Table 1-1: Discussion of advantages and disadvantages of Gibson assembly, Gateway cloning and Golden Gate assembly.....	19
Table 2-1: Optimised settings for the BD LSR Fortessa flow cytometer.	43
Table 2-2: Example compensation matrix generated by detection on the BD LSR Fortessa flow cytometer.	44
Table 2-3: Optimised settings for the Attune NxT flow cytometer.	44
Table 2-4: Example compensation matrix generated by detection on the Attune NxT flow cytometer.	45
Table 3-1: Complementary 4 bp linkers for multi-gene expression vector cloning. .	59
Table 3-2: Multi-gene expression vector variants constructed by varying fluorescent protein gene order.	75
Table 3-3: Relative fold change in fluorescent protein expression dependent on gene position within the polycistronic cassette.	78
Table 3-4: Mean expression ratios within the polycistronic cassette of the multi-gene expression vector variants.	81
Table 4-1: Expected Amplicon size for each synthetic proximal element.....	95
Table 4-2: Optimised settings for the Attune NxT flow cytometer.	97
Table 4-3: Relative fold change in fluorescent protein expression compared against 5RPU.....	100
Table 4-4: Primers designed and used for real-time quantitative PCR of glutamine synthetase, <i>eGFP</i> , <i>mCherry</i> and <i>tagBFP</i> genes.	102
Table 4-5: Average expression fold change by qPCR between 100 and 800 ng of transfected DNA.....	106
Table 4-6: Average relative promoter activity from qPCR analysis.....	108
Table 4-7: Average relative promoter activity from flow cytometry analysis.....	113
Table 4-8: Multi-gene expression vector variants constructed to characterise positional effects.	124
Table 4-9: Fold Change of relative transcriptional activity to quantify positional effects of the multi-gene expression vector variants.	126
Table 4-10: Average expression fold change of integrated median fluorescent intensity to quantify positional effect of multi-gene expression variants.	129

Table 4-11: Summary of synthetic promoter combination within each multi-gene expression vector variant constructed.	132
Table 4-12: Gene expression stoichiometric ratios of 27 multi-gene expression vector variants.	142
Table 5-1: Identification of conserved functional groups of transcription terminators and the design space for synthetic elements.....	165
Table 5-2: Library of 36 multi-gene expression vector variants for terminator library screen.....	172
Table 5-3: Design of prototype multi-gene expression vectors co-expressing a DTE monoclonal antibody and <i>XBP-1s</i>	185
Table 5-4: Viable cell concentration and viability of TI stable pools post-suspension adaptation.	188
Table A-1: Cost and service analysis of de novo synthesis of transcription units by third party suppliers.....	233
Table A-2: Coding DNA sequence for <i>eGFP</i> , <i>mCherry</i> and <i>tagBFP</i>	234
Table A-3: Ordinary one-way ANOVA – Viable Cell Concentration.....	239
Table A-4: Ordinary one-way ANOVA – Viability.....	239
Table B-1: List of primers used for PCR modification of synthetic promoter library.	241
Table C-1: Transcription termination library.	255

List of Abbreviations

ADC	Antibody-drug conjugate
att	Site-specific attachments
bGH	Bovine growth hormone
bp	Base pair
CAP	CEVEC's amniocyte production
CDS	Coding DNA sequence
cDNA	complementary Deoxyribonucleic acid
chef1 α	Chinese hamster elongation factor-1 alpha
CHO	Chinese hamster ovary
CL	<i>Cypridina</i> luciferase
CoTC	Co-transcriptional cleavage
Cre	Cause recombination
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
Ct	Cycle threshold
DNA	Deoxyribonucleic acid
DTE	Difficult-to-express
dhfr	Dihydrofolate reductase
<i>E.coli</i>	<i>Escherichia coli</i>
eGFP	Enhanced green fluorescent protein
ER	Endoplasmic Reticulum
ERS	Endoplasmic Reticulum stress
FBOG	Fed-batch overgrow
FSC	Forward scatter
GHT	Glycine, hypoxanthine and thymidine
GLP-1	Glucagon-like peptide 1
<i>glnS</i>	Glucosamine-6-phosphate synthase
GMP	Good manufacturing practice
GnT	Acetylglucaminyltransferases
GS	Glutamine synthetase
HC	Heavy chain
hCMV-MIE	Human cytomegalovirus- major intermediate-early

HDV	Hepatitis δ virus
HEK	Human embryonic kidney
hGH	human growth hormone
HIV	Human immunodeficiency virus
HSV-tk	Herpes simplex virus thymidine kinase
HT	High throughput
iMFI	Integrated median fluorescent intensity
IRES	Internal ribosome entry site
IVCD	Integral viable cell density
LC	Light chain
LCR	Ligase cycling reaction
LoxP	Locus of crossover in P1 bacteriophage
mAb	Monoclonal antibody
MAZ	MYC-associated zinc finger protein
MFI	Median fluorescent intensity
MGEV	Multi-gene expression vector
MSX	Methionine sulfoximine
MSA	Mouse serum albumin
MTX	Methotrexate
NCBI	National centre for biotechnology information
NHEJ	Non-homologous end joining
NOMAD	Nucleic acid ordered assembly with directionality
NS0	Mouse myeloma cells
pA	Polyadenylation
PAS	Polyadenylation signal
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PMT	Photomultiplier tube
PTM	Post-translational modifications
PER.C6	Human embryonic retinoblastoma
qP	Specific productivity
qPCR	Real-time quantitative polymerase chain reaction
RNA	Ribonucleic acid
RNA pol II	RNA polymerase II

RTA	Relative transcriptional activity
SEAP	Secreted alkaline phosphatase
SGV	Single gene vectors
SSC	Side scatter
SV40	Simian virus 40
tagBFP	Tag blue fluorescent protein
TF	Transcription factor
TFRE	Transcription factor regulatory elements
TGE	Transient gene expression
TI	Targeted integration
TSS	Transcription start site
tPA	Tissue plasminogen activator
TNF α	Tumour necrosis factor alpha
TU	Transcription unit
UNS	Unique nucleotide sequences
UPR	Unfolded protein response
UTR	Untranslated region
VCC	Viable cell concentration
XBP-1s	X-box binding protein-1 spliced

Abstract

Chinese hamster ovary (CHO) cells are the predominant host cell factory for biopharmaceutical production. However, the increased complexity of next generation therapeutics have made them difficult-to-express (DTE). As a result, additional multi-gene engineering is required to improve the phenotypic capabilities of the host cell. This study explores the development of a synthetic biology tool to facilitate complex multi-gene engineering approaches specifically for biopharmaceutical production in CHO cells.

The first step was to successfully design and develop a robust multi-gene expression vector (MGEV) system. This was achieved by identifying the MGEV system design space which yielded an *in silico* designed standardised transcription unit (TU) template and a pExp-Vec-GG recipient vector. A set of complementary linkers were also designed to promote directional cloning of multiple TUs in a MGEV using Golden Gate assembly. This bespoke MGEV system was tested *in vitro* to demonstrate a rapid, reproducible and robust design and cloning method as shown by successful construction of 70 MGEVs across the entire study. When quantifying gene expression performance within the MGEV by co-expressing fluorescent proteins, a repression of downstream TU gene expression within the polycistronic cassette was observed. This was hypothesised to be a consequence of transcriptional interference.

Moreover, titrating multiple recombinant genes simultaneously allows for complex multi-gene engineering strategies with precise control. This was achieved by applying a validated synthetic promoter library to the MGEV system. Generally, the synthetic promoters performed as expected yielding a relatively low, medium and high level of transcriptional activity within both a single gene plasmid and a MGEV. Further characterisation of the co-functionality of the promoters demonstrated varying levels of gene repression potentially caused by promoter squelching. Additionally, the positional-mediated interference within the polycistronic cassette was determined to be transcriptional strength dependent. Therefore, the gene repression within a MGEV is dynamic and unpredictable due to a cumulation of different transcriptional interference mechanisms. As a result, a database of titrated gene expression

stoichiometric ratios of 27 MGEVs using various promoter strength combinations was generated to guide future multi-gene engineering strategies.

Finally, the functionality of the MGEV system as a tool for developing synthetic genetic components and CHO cell engineering for biopharmaceutical production were demonstrated. This included rapidly testing a library of *in silico* designed synthetic transcription terminators. The screen identified a potential synthetic polyadenylation (pA) element, but also highlighted critical design features for functional motifs. Furthermore, CHO cells were stably engineered by co-expressing a DTE monoclonal antibody (mAb) and X-box binding protein-1 spliced (XBP-1s) transcription factor. This demonstrated successful increase in mAb titre and growth rate of the recombinant CHO cell.

In conclusion, this study has illustrated a successful design of a robust MGEV system that can titrate gene expression and be applied to various CHO cell engineering and synthetic biology strategies.

Chapter 1

1. Introduction

1.1. The Biopharmaceutical Industry

The biopharmaceutical industry focuses on the development of drug molecules using biological systems for the treatment of infectious diseases, autoimmune disorders and some cancers. The industry refers to these types of molecules as recombinant therapeutics, biologics or biopharmaceuticals. There are nine main classes of biopharmaceuticals identified within the industry – monoclonal antibodies (mAbs), therapeutic enzymes, fusion proteins, hormones, growth factors, blood factors, anti-coagulants, cytokines and recombinant vaccines (Aggarwal, 2014).

In 2016, the total sales of the biopharmaceuticals (excluding vaccines) was recorded at \$163 billion, which is a 102% increase from 2008. Within those sales, the total sales of mAbs was \$106.9 billion (Kesik-Brodacka, 2017), and contributes towards 40% of the biopharmaceutical market. The significant increase in the sales revenue suggests the future prospect of the industry is positive with continued growth expected in coming years, hence warranting continuous research to improve the manufacturing process. Currently, there are a number of blockbuster mAbs in the market with the most successful molecule being adalimumab, also known as Humira. This molecule is used as a tumour necrosis factor alpha (TNF α) inhibitor for the treatment of rheumatoid arthritis and other related illnesses. However, as the industry continues to expand, so does the pipeline of molecules with the development of next generation mAbs. This includes the emergence of bispecific antibodies, antibody-drug conjugates (ADCs), single chain variable fragments and glycoengineered mAbs (Kesik-Brodacka, 2017). These next generation molecules pose a challenge, as they are more complex

and in turn can challenge the capabilities of the current manufacturing platforms to produce high quality biopharmaceutical at a cost-effective rate. Sometimes these newly designed molecules can be a problematic to produce, resulting in low yields and affecting growth of the biological host from a traditional manufacturing system. These types of molecules have been classed as difficult-to-express (DTE) recombinant therapeutics within the industry.

1.2. Overview of the current production technology of biopharmaceuticals

The production of biopharmaceuticals is achieved by using recombinant DNA technology. The first biopharmaceutical to be approved by drug regulatory agencies was human insulin in 1982, which was expressed in *Escherichia coli* (*E.coli*) (Johnson, 1983). Technology has since continued to progress with the development of new host cell factories and using more advance gene expression systems for more effective biopharmaceutical production. Figure 1-1 shows a simplified flow path summarising how recombinant DNA technology and different host cell factories are used for the production of biopharmaceutical such as mAbs and antibody fragments.

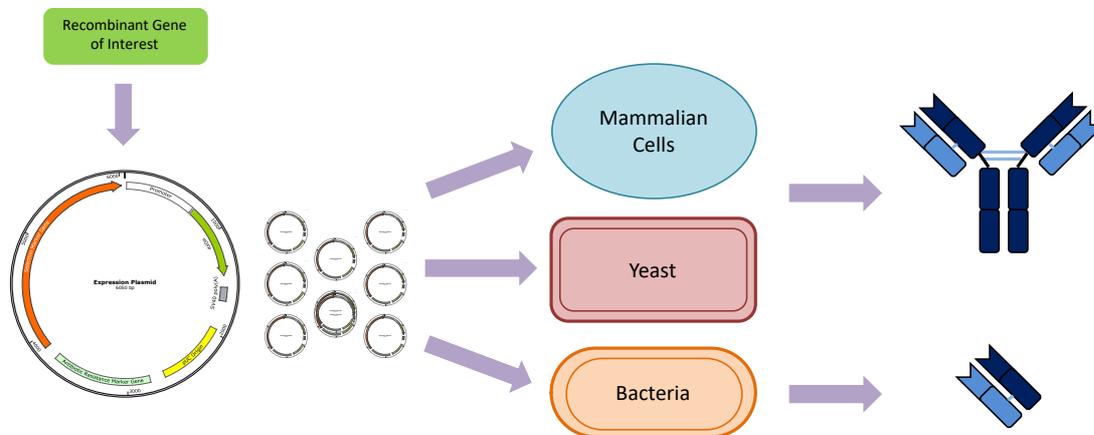


Figure 1-1: Recombinant protein expression summary.

A simplified schematic summarising how recombinant proteins such as mAbs and antibody fragments are expressed using recombinant DNA technology within different host cell factories. Whereby, multiple copies of the plasmid DNA are transferred intracellularly using various mechanisms to employ the host cell machinery to express the recombinant protein of interest.

This section will further explore and review the current host cell factories and gene expression systems used within the industry and academia.

1.2.1. Host cell factories

There are a range of host cell factories available including bacterial, yeast, insect and mammalian cells, along with more synthetic approaches such as cell-free systems. The current preferred host cell factory in industry is mammalian cells, for an array of reasons; including performing human-like protein folding and assembly, and post-translational modifications (PTMs).

1.2.1.1. Alternative host cell factories

A range of alternative host cell factories to mammalian cells have been investigated and employed in industry and academia for the recombinant production of biopharmaceuticals. This includes bacterial, yeast, insect and cell-free systems. As mentioned previously, the first therapeutic molecule to be approved in the market used bacterial host cells for production. Bacterial hosts such as *E.coli* express recombinant proteins in the cytoplasm and these either remain there, migrate to the periplasm, get deposited between the inner and outer membrane or are secreted. Examples of biopharmaceuticals produced by *E.coli* include antibody fragments, cytokines, hormone and growth factors (Berlec and Strukelj, 2013). There are a number of advantages for *E.coli*-based production, including usage of a well characterised host, being cost effective, and a faster and simpler production platform. However, the major limitation of using prokaryotic hosts is their limited capability to perform essential PTMs such as glycosylation and proteolytic maturation for more complex biopharmaceuticals.

Yeast host cells are unicellular and the simplest form of eukaryotes with similar mechanistic capabilities to mammalian cells. The two main strains of yeast used in biopharmaceutical production are *Saccharomyces cerevisiae* and *Pichia pastoris*. These strains have the ability to perform quaternary and multi-component protein folding, as well as PTMs such as glycosylation, unlike bacterial hosts. However, similar to bacterial hosts, yeast strains have well characterised biochemical and cellular functions (Berlec and Strukelj, 2013) and are inexpensive to cultivate. Moreover, yeast strains have a lower susceptibility to contamination from bacteriophage and can generate yields of >1 g/L (Martínez et al., 2012). The main drawback of yeast host cell factories is that the recombinant protein glycosylation profile is non-human-like. For instance, N and O- glycans moieties are significantly

different between yeast and mammalian cells (Martínez et al., 2012). Furthermore, yeast strains generate glycan polymers with higher mannose content (over 30 residues) compared to mammalian cells. This can impact on the biological activity, reduce stability of the molecule and cause unwanted immunogenic reactions in patients (Berlec and Strukelj, 2013).

More recently, synthetic approaches have been explored for biopharmaceutical production such as cell-free expression systems. This technology uses the crude extract of translational machinery (ribosomes, tRNAs, aminoacyl-tRNA synthetases, and initiation, elongation and termination factors), amino acid residue, co-factors, essential ions, an energy source and combined with genetic material to express a recombinant protein. Literature has shown a number of cell extracts have been developed from various sources such as rabbit-reticulocyte lysates, wheat germ embryos, insect cells and human cells. This method might be ideal for production of DTEs and toxic proteins and allows for more control through the protein expression process. However, the raw materials are expensive and the production rate is low and inefficient (Casteleijn et al., 2013). However, more recently, Sutro biopharma have successfully produced an ADC using cell-free system and initiated it in clinical trials. This indicates progression in the technology, whereby lower quantities are acceptable but speed of production was preferred, as an alternative means of biopharmaceutical production (Zimmerman et al., 2014).

Although each of these host cell factories have their limitations, literature has shown host cell engineering can be used to improve these systems to make them viable for biopharmaceutical production by providing the capacity to perform human-like PTMs. For example, N-glycosylation from *Campylobacter jejuni* was successfully engineered into *E.coli* (Pandhal et al., 2012; Spadiut et al., 2014) demonstrating the ability to engineer *E.coli* to glycosylate and potentially make glycoproteins. Similarly, *Pichia* species (GlycoFi technology) and *S. cerevisiae* have been engineered to perform more human-like glycosylation during biopharmaceutical production (Beck et al., 2010; Martínez et al., 2012). These advances improve the potential of using microbial host cell factories for more cost effective and rapid biopharmaceutical production. Currently mammalian cells are still the preferred, however, and still the industry gold standard host cell factory for recombinant protein production.

1.2.1.2. Mammalian host cell factories

A range of mammalian host cell factories have been employed in both academia and industry for biopharmaceutical production including mouse myeloma cells (NS0), human embryonic kidney (HEK) cells, human embryonic retinoblastoma (PER.C6), CEVEC's amniocyte production (CAP) cells and Chinese hamster ovary (CHO) cells. Mammalian cells are the chosen cell factory due to their capacity to perform PTMs of glycoproteins which are essential for bioactivity and minimise unwanted immunogenicity in patients. For example, these host cells can assemble complex multi-domain proteins by formation of disulphide bridges in the endoplasmic reticulum (ER), and perform human-like glycosylation (N & O-glycan moieties) and sialylation (Berlec and Strukelj, 2013) in the ER and Golgi apparatus (Goh and Ng, 2018; Jenkins et al., 2008; Lalonde and Durocher, 2017).

CHO cells are the industry-preferred host cell factory for biopharmaceutical production and were first established in the 1957 by Puck et al. The original derivative cell line was genetically deficient in the synthesis of proline, which also affected the conversion of glutamic acid to glutamine (Kao and Puck, 1968; Puck et al., 1958). A number of derivatives have since been developed, converting the CHO-K1 host from an adherent cell line to a suspension cell line, and are employed by various biopharmaceutical companies and academic labs (Jayapal et al., 2007). The first biopharmaceutical to be manufactured and approved in CHO cells was tissue plasminogen activator (t-PA) in 1987 (Wurm, 2004) and since then, CHO cells have become established as the industry standard host. As a result, the continuous development of CHO production processes has led to robust and well-established platforms for biopharmaceutical production. Therefore, the major advantage of using CHO cell factories is the high titres of 10-15 g/L which they can generate at manufacturing scale (Zhu, 2012). Furthermore, the system uses chemically-defined media and is less receptive to human viral infections, both of which safeguard biopharmaceutical production against undesired contaminants. The CHO cell factory has been the chassis of production for many approved recombinant therapeutics. The platform, therefore, is familiar among drug regulatory agencies, which de-risks future drug molecule approval. A drawback to CHO cell factories is that the cultivation duration is longer than non-mammalian host cell factories and hence is less cost-effective and more time-consuming. Moreover, they are not human derived cell lines

and as such, the glycosylation profile does differ. For example, higher levels of sialylation are observed and generate two immunogenic glycan epitopes on recombinant proteins. However, data has shown that minimal unwanted immunogenic reactions have been triggered in patients upon administration of biopharmaceuticals from CHO cell production (Butler and Spearman, 2014). In conclusion, CHO cell factories do have certain limitations; nonetheless, the many benefits of the system do outweigh the drawbacks and hence they remain the industry and academia standard for biopharmaceutical production (Fan et al., 2012; Jayapal et al., 2007).

1.2.2. Recombinant gene expression technology

Recombinant DNA technology is the methodology used to facilitate the expression of biopharmaceuticals. This is achieved by the introduction of recombinant genetic material by chemical-based or electroporation-based mechanisms into a host cell using the cell's endogenous machinery to express a recombinant protein. However, to achieve continuous stable expression and high titres, the recombinant DNA must integrate into a transcriptionally active loci in the host cell genome by homologous recombination (Costa et al., 2010). Within the biopharmaceutical industry and academia, a range of gene expression systems have been developed for the selection of CHO cells successfully expressing a recombinant protein. The most common expression systems employed are dihydrofolate reductase (dhfr), glutamine synthetase (GS) and antibiotic-based selection systems. All of these expression systems select based on an essential cellular function and use a chemical inhibitor to improve the stringency of selection. Therefore, facilitating the selection and isolation of high expressing stable CHO cells which have a selective advantage over the parental CHO cell (Costa et al., 2010).

1.2.2.1. Antibiotic-based selection systems

There are a range of antibiotic-based selection systems which can be used for cell line selection of recombinant protein expression. Examples of antibiotics used in industry and academia include hygromycin B, puromycin and neomycin (Lanza et al., 2013). One of the most popular antibiotics used for biopharmaceutical production is hygromycin B. This antibiotic is an aminoglycoside that affects the binding of ribosomal components in mammalian cells, in turn inhibiting translation of proteins. The resistance gene expressed is a kinase that inhibits hygromycin B functionality and

allows for selective advantage. Therefore, CHO cells which have successfully integrated and are expressing the hygromycin B resistance gene, along with the tandemly attached recombinant gene of interest, will survive under the selection pressure of hygromycin B supplemented culturing media (Bernard et al., 1985; Lanza et al., 2013).

1.2.2.2. Dihydrofolate reductase (DHFR)-based selection system

Dihydrofolate reductase (gene name - *dhfr*) is the enzyme that converts dihydrofolate to tetrahydrofolate and is essential for nucleotide and amino acid synthesis, therefore, essential to the host cell. The selection process for recombinantly expressing CHO cells is performed by the over-expression *dhfr* alongside the recombinant protein of interest in CHO-DG44 cells that are deficient in dihydrofolate reductase (Fan et al., 2013). Therefore, culturing CHO cells in media lacking glycine, hypoxanthine and thymidine (GHT) will favour selection of CHO cells with *dhfr* expression. To improve the stringency of selection and perform gene amplification, a *dhfr* chemical inhibitor called methotrexate (MTX) is supplemented. This allows for favourable selection of CHO cells with more successfully integrated copies of the recombinant *dhfr* gene and the recombinant gene of interest, yielding higher concentrations of the *dhfr* over the threshold of MTX inhibition, leading to the generation of highly expressing recombinant stable CHO cell lines.

1.2.2.3. Glutamine synthetase (GS)-based selection system

GS is an essential enzyme within the glutamine metabolism pathway in mammalian cells that is involved in converting glutamate into glutamine with the presence of ammonium. Glutamine is an essential amino acid for mammalian cell growth, acting as an energy source, and a nitrogen donor in synthesis of amino acids and nucleotides (Fan et al., 2013). Thus, host cells with elevated levels of *GS* successfully survive in glutamine-free media. NS0 cells are a *GS* negative cell line whereas CHO cells have endogenous *GS*. Indeed, for the *GS* selection system to be effective in CHO cells, a chemical inhibitor called methionine sulfoximine (MSX) is supplemented to improve the stringency of selection, ensuring that only cells with high levels of *GS* expression survive. Therefore, selecting recombinant CHO cells where high levels of recombinant *GS* gene expression is observed by integration into a transcriptionally active loci, in

turn concurrently expressing high levels of the recombinant protein of interest (Bandaranayake and Almo, 2014).

At the moment *GS*-based selection is the most robust system within industry and has been employed in the manufacturing of 50 biopharmaceutical products, ranging from potential clinical trial candidates and existing products in the market. The system is faster at generating high expressing cell lines than DHFR-based selection and does not require gene amplification. Conversely, DHFR-based selection is time consuming since it requires a step-wise gene amplification phase to generate high-producing cell lines with up to 1000 copies. A by-product of requiring amplification is that when the MTX is removed, it can lead to cell line instability (Fan et al., 2013). Furthermore, the lack of antibiotics required for *GS*-mediated selection is favourable with drug regulatory agencies due to minimised risk of contamination post-processing.

1.3. Overview of the biopharmaceutical and bioprocessing production platform

The biopharmaceutical production platform consists of 3 major stages; drug discovery and development, process development and manufacturing. Each of these stages can be further partitioned into sub-stages as shown in Figure 1-2

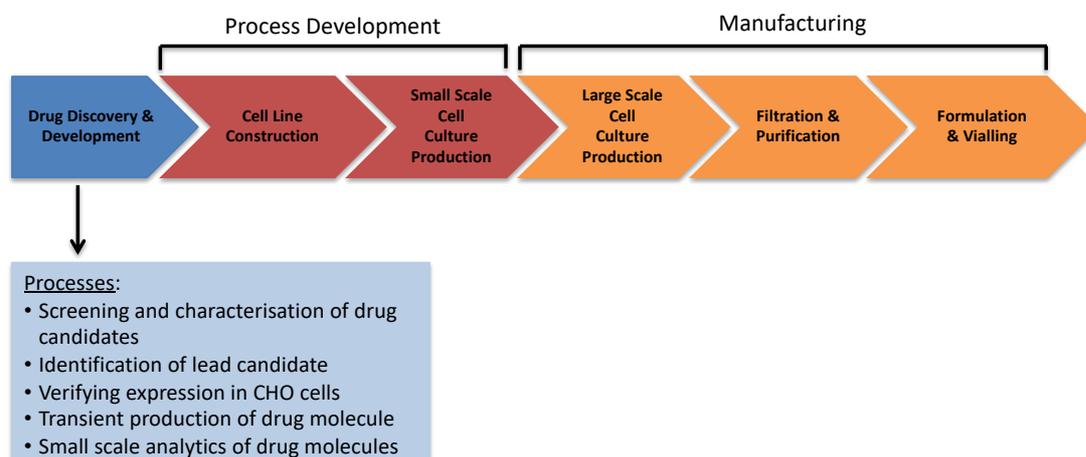


Figure 1-2: Biopharmaceutical production workflow.

Summarises the process steps during the development of biopharmaceuticals from early drug discovery to large scale manufacturing within a biopharmaceutical company for clinical trials or retail in the market.

Drug discovery and development involve a number of screening and verification steps to identify a lead drug candidate with ideal pharmacokinetic, bioactivity, expression, stability and formulation characteristics. This stage normally involves a combination of transient and stable pool-mediated recombinant protein expression. Traditionally, transient gene expression (TGE) is the first expression step in biopharmaceutical production. It involves the introduction of the recombinant gene of interest by either chemical or electroporation-based transfection for short- or long-term protein expression. TGE can be used as a short-term high throughput (HT) screening tool to generate small amounts of protein for rapid screening of drug molecules (Geisse, 2009). Alternatively, it can be used for large-scale long-term expression of a lead drug candidate for early stage material supply of up to 2 g/L by using episomal-based TGE (Daramola et al., 2014). TGE generated material can be used for manufacturability assessment, preclinical testing and analytical method development to support downstream processing and manufacturing stages (Costa et al., 2010). The next step involves the generation of stable pools which are polyclonal expressing recombinant CHO cell lines and are used as an alternative approach to generate large amounts of biopharmaceuticals.

The first step of the process development stage is to establish a clonal stable cell line expressing the lead drug candidate from a stable pool cell line. The stable pool cell line generated is often a heterogeneous population, with varying expression levels due to random integration and varying gene copies within the genome. Furthermore, as a result of the heterogenous nature of the pool, it may demonstrate inconsistent growth, unpredictable yield and varying product quality on scale up. Due to regulatory requirements, these variations are unacceptable and thus a clonal stable cell line has to be generated (Agrawal and Bal, 2012). The key attributes of a clonal cell line are high productivity, high growth rate, predictable stability, adaptable to suspension, functional in varying bioreactor scales and suitable product quality. Historically, a clonal cell line has been isolated by serial limiting dilution and low throughput screening to verify clonality. This method is time consuming and labour intensive and therefore limits the number clones screened (De Jesus and Wurm, 2013). Yet, HT systems have been developed, such as fluorescent-activated cell sorting (FACS) or fully automated systems like ClonePix to isolate single cells. The increased capacity allows for screening and testing more clonal cell lines through multiple rounds of

small-scale bioreactor systems, like the AMBR 15 or shake flasks to eventually identify a clonal cell line demonstrating all the desired characteristics.

This, in turn, is followed by medium-scale material generation using 0.1 to 200 L bioreactors to further optimise the production process. This includes optimisation of media composition, feeding strategies (Sellick et al., 2011) and bioreactor conditions (Li et al., 2010) to extend the bioproduction duration in an aim to maximise yield and generate high quality product during a single fed-batch production (Birch and Racher, 2006). The material generated from the series of optimisation steps, along with the upstream TGE material, is used for early stage downstream process optimisation including developing of purification strategies and identifying a suitable product formulation.

The final stage of the biopharmaceutical production workflow is manufacturing which involves large scale production ranging from 2000 to 20000 L bioreactors. This stage is performed under good manufacturing practice (GMP) and involves large scale optimised purification often entailing an affinity chromatography step followed by orthogonal chromatography to generate a high purity product. The purified material is then formulated and vialled ready for distribution. This stage of the platform is inspected and regulated by the drug regulatory agencies prior to distribution of the biopharmaceutical (Birch and Racher, 2006; Hesse and Wagner, 2000).

1.4. Demands within industry and limitations of current technology

As discussed in the previous sections, CHO cell factories and the *GS* selection system are the industry and academia preferred system for biopharmaceutical production. Therefore, the system is continuously being developed to be cost effective, robust and reproducible by addressing aspects that are poorly understood. Examples of some of the limitations within the current technology are discussed below:

- i) **Predictable and titratable levels of constitutive gene expression for transient and stable production of biopharmaceuticals** – Current recombinant gene expression is typically performed using the human cytomegalovirus- major intermediate early 1 (hCMV-MIE1) promoter which is a strong viral promoter. However, this promoter does have

associations to induction of cell stress, is dependent on the cell cycle and has propensities towards epigenetic silencing. Furthermore, the mechanistic functionality of the promoter is poorly understood in CHO cells (Brown et al., 2014). These features limit the capacity of the promoter to titrate gene expression at a predictable level. This requirement has become more important as recent research has shown that achieving an optimal ratio of heavy (*HC*) and light chain (*LC*) expression for a mAb can significantly improve expression titres (Pybus et al., 2014).

- ii) **Limitations in engineering of CHO cell factories due to limited multi-gene expression cassette capacity** – The development of next generation biopharmaceuticals, such as bispecific antibodies (as mentioned previously), require the capacity to genetically engineer the CHO cell factory to alleviate bottlenecks in the recombinant protein expression pathway. This has been shown by transient co-expression of *XBP-1* (Pybus et al., 2014) and *CypB* (Johari et al., 2015) to enhance titres of DTE molecules. Albeit, co-expression of multiple genes can be performed transiently, there is a limitation in the capabilities of achieving this stably using a single cell line engineering step. Currently, a multi-step engineering of CHO cell factories is performed by firstly, genetically modifying the parental host by, for example, overexpression of *Ero1-L α* followed by generating a recombinant CHO cell expressing a biopharmaceutical (Cain et al., 2013). This approach is time consuming and not flexible for different product specific requirements.

- iii) **Generating predictable and reproducible stable cell lines for efficient drug regulatory approvals** – Drug regulatory agencies verify and validate the product's consistency in the physiochemical properties and biological activity from early stage expression to manufacturing (Li et al., 2010). The product should maintain similar levels of safety and efficacy between production batches. Therefore, the CHO cell factory can be engineered for specific attributes in an aim to achieve control, reproducibility, and predictability during production. As a result, improving the approval process by drug regulatory agencies. (Doblhoff-Dier and Bliem, 1999;

Hesse and Wagner, 2000). For example, it has been documented that signal sequence peptides can lead to mis-cleavage causing either elongation or truncation of the *HC* or *LC* of a mAb, leading to product heterogeneity. Subsequently, improving the predictability and functionality of a signal sequence peptide can improve the homogeneity of the therapeutic protein and de-risk drug approval complications (Gibson et al., 2017)

Addressing the limitations in the current technology has an overall impact of facilitating the industry to efficiently manufacture improved and high-quality biopharmaceuticals at an affordable cost for the betterment of patients and the healthcare industry.

1.5. Overview of synthetic biology

Synthetic biology is a field which has emerged as an exciting area of research in the last two decades. This specific field combines concepts of engineering such as the use of design, creation and control to the biological sciences, with the aim to develop biological systems which exhibit predictable and reproducible behaviour (Jewett and Ellis, 2017; Pasotti et al., 2012). The biopharmaceutical industry is aspiring to apply similar concepts in the production platform, aiming to make the drug manufacturing process more cost effective and ease approval by drug regulatory agencies (Duportet et al., 2014). Synthetic biology spans a wide area of research covering prokaryotic and eukaryotic systems involved in a range of industries including manufacturing biofuels, personalised biological therapies, affordable and deployable diagnostics and vaccines for global health (Pasotti et al., 2012).

An example of ongoing research in the biotechnology and the healthcare industry includes the reprogramming of cells. This is achieved by genome engineering using tools like clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9, controlling gene expression and interrogating gene networks. For example, there is ongoing research in microbial engineering to design new probiotics as a form of therapeutic treatment towards human diseases. *Bifidobacterium longum* has been engineered to express an active form of glucagon-like peptide 1 (GLP-1) as a fusion protein into the colon when residing in the gastrointestinal tract of rats. The secreted GLP-1 has shown to assist in converting intestinal epithelial cells into insulin-

secreting cells within diabetic rats. This is a promising future therapy for the treatment of type 1 and 2 diabetes melitus (Chua et al., 2017; Duan et al., 2015; Wei et al., 2015).

Another example of synthetic biology being applied in biotechnology is the development of biosensors to actively detect biological reactions. Current examples include electrochemical biosensors which are involved in identifying a biological reaction by *in vitro* enzymatic detection. This is employed in bioprocessing with biosensors detecting metabolites like glucose and lactate, where enzymes like glucose oxidase and L-lactate dehydrogenase are used respectively as electrochemical biosensors. Furthermore, with the ability of cellular reprogramming and gene expression control, transcription-based biosensors are being developed which would trigger a synthetic promoter leading to a detectable sensor such as GFP expression and be used to detect molecules like phenylalanine and threonine in *E.coli* (Dekker and Polizzi, 2017; Liu et al., 2015; Mahr et al., 2016)

However, there are still technical hurdles to be overcome with synthetic biology, particularly towards mammalian cell engineering. Approaches need to be identified to achieve scalability, orthogonality and predictability of synthetic circuit behaviour in a more complex mammalian system (Lienert et al., 2014). The following sections will expand on the functionality and application of multi-gene systems and synthetic genetic elements, along with current research published in literature for both aspects.

1.6. Introduction of multi-gene expression vectors as tool for cell engineering

A multi-gene expression vector (MGEV) is a recombinant DNA plasmid (often circular), which comprises of multiple transgenes forming a polycistronic cassette assembled in a tandem series. Literature has demonstrated the successful assembly of up to 11 transgenes in a single plasmid (Weber et al., 2011). In a MGEV, each transgene is paired with its respective promoter allowing for modular control of a series of recombinant genes. Furthermore, all recombinant genes are regulated and maintained under the control of a single selection marker during cell engineering. Currently, the application of MGEVs can be employed in gene therapy, synthetic biology and reprogramming cell fate and function, particularly in prokaryotic and eukaryotic cells (Guye et al., 2013). The latter function applies towards biopharmaceutical production in the ability to facilitate genetic engineering and

reprogramming of CHO cell factories to improve recombinant protein expression, particularly of DTE proteins.

1.6.1. Current approaches and limitations to cell engineering in mammalian cells

There are number of approaches which have been used to perform mammalian cell engineering such as sequential engineering of the parental host, targeted-integration using multiple landing pads for genetic engineering, using internal ribosome entry site (IRES)-based multi-gene engineering and CRISPR/Cas9 genome editing.

Sequential engineering of the parental host involves numerous rounds of integration and cell line generation for multiple recombinant genes using a combination of selection markers to isolate the desired engineered cell from the original parental host. There are limitations with this methodology; limited engineering targets due to a finite number of viable selection markers which can be used simultaneously, and the time consuming and labour-intensive nature of this approach. Alternatively, the parental host can be co-infected using viral delivery methods with multiple plasmids containing the recombinant genes of interest to generate the desired engineered mammalian cell (Sommer et al., 2009). However, using viral delivery mechanisms can limit the DNA payload (Gaidukov et al., 2018) and increase risk of viral contamination, which is a concern in biopharmaceutical production. Nevertheless, companies are developing means to minimise the risk and use lentivirus-mediated biopharmaceutical production.

An alternative approach is to use a multi-landing pad targeted integration method with independent selection markers to be able engineer a cell line with multiple recombinant proteins. For example, a CHO cell line was modified to have multiple landing pads, allowing up to nine copies of recombinant antibody to be integrated and expressed while selection was performed using different antibiotic resistance genes such as hygromycin B and blasticidin (Gaidukov et al., 2018). The advantage of this method is that it allows for more control and precision towards recombinant gene copy integration into the genome. However, it requires a significant amount of host cell modification to have multiple landing pads initially introduced into the genome and consequently, would be a time-consuming method to finally achieve multi-gene cell engineering.

An approach to avoid the limitation of multi-selection markers is to use an IRES-based polycistronic cassette (Sommer et al., 2009) or a 2A translation skip peptide (Gaidukov et al., 2018). This would allow for the integration of multiple recombinant genes under the selection of a single marker. This approach would also be significantly quicker as it would not require multiple cell line generation steps. However, IRES-bicistronic cassettes have shown to have lower expression levels compared to recombinant genes regulated by individual promoters (Bayat et al., 2018). Inefficiency of the IRES-mediated or 2A translation skipping peptide could be the cause of the lower levels of expression. Furthermore, the inefficiency of the sequence could cause undesired N and C- terminus modifications, in turn affecting protein functionality and product quality for biopharmaceutical production (Kriz et al., 2010).

An orthogonal method to cell engineering which has gained traction recently is CRISPR/Cas9. This is a precise genome editing tool which was adapted from type II prokaryotic cells. The tool was developed using *Streptococcus pyogenes* SF370 CRISPR locus including four genes – Cas9 nuclease, non-coding CRISPR RNA (crRNA)(trans-activating CRISPR RNA and precursor CRISPR RNA). The RNA sequences help direct and guide the Cas9 nuclease to cleave double-stranded DNA. Within the sgRNA, genetic engineer targets (can be multiple targets) are included. The target genes can be integrated into the genome by non-homologous end joining (NHEJ) -random integration or homology-directed repair -targeted integration (Cho et al., 2013; Cong et al., 2013; Lee et al., 2016). The precision of the tool is highly advantageous, and the approach allows for increased control through genetic engineering of mammalian cells. The drawbacks to the method include limited plug-and-play functionality for engineering flexibility, limited control on gene expression by synthetic genetic elements like promoters and transcription terminators and some licensing issues which could affect the commercial aspects when applied in a biopharmaceutical environment.

1.6.2. Applications of multi-gene expression vectors in a biopharmaceutical environment

The introduction of MGEVs in the biopharmaceutical industry will increase the capabilities of performing more complex CHO cell engineering in an effective manner. The main goal is to improve the CHO cell factory to increase biopharmaceutical

production or develop new phenotypic characteristics to benefit the production platform. Examples of how multi-gene engineering could be applied using MGEVs are as follows:

- i) MGEVs could be used to co-express a recombinant protein of interest such as a mAb and a series of accessory genes including unfolded protein response (UPR) transactivators and endoplasmic reticulum (ER) chaperones, such as *XBP-1*, *BiP* and *CypB* (Johari et al., 2015; Pybus et al., 2014), or secretion related proteins, such as *SRP14*, *SRP9* and *SRP54* (Le Fourn et al., 2014). These types of accessory genes would assist in minimising bottlenecks within the expression, synthesis and translocation of the mAb or help increase the CHO cell factory capacity to increase expression of DTE molecules.

- ii) MGEVs could also be used to perform more complex cell engineering. For example, genetically modifying the phenotype of the CHO cell factory while concurrently expressing a recombinant protein. For example, modifying glycosylation and sialylation characteristics by introducing N-acetylglucosaminyltransferases – GnT II, IV and V and α 2,6-sialylated glycan residues respectively (Fischer et al., 2015). Other phenotype modifications could include altering growth of the CHO cell factory by co-expressing genes like onco-KIT. This allows for targeted host cell modification by enhancing overall protein synthesis, proliferation and improved titres of recombinant glycoproteins, generating a superior CHO cell factories for the production of next generation biopharmaceuticals (Guye et al., 2013; Mahameed and Tirosh, 2017).

1.6.3. DNA cloning approaches to assemble multi-gene expression vectors

The traditional method of DNA cloning involves restriction digestion by type II endonucleases followed by ligation using a T4 DNA ligase. This approach of cloning has been used for the last 40 years (Chao et al., 2014). However, this method has a number of drawbacks such as:

- i) A limited number of type II restriction endonucleases that have a unique palindromic site for recognition. For example, the recognition site of NotI and EagI is exactly the same, but the direction of cleavage of the DNA strand is different. Furthermore, many of the palindromic recognition sites are very similar (1-2 base variation) which could have some risk with fidelity of cloning.
- ii) The emergence of type II restriction endonuclease recognition sites increases in larger plasmids limiting the number of unique cloning sites available, restricting the ability to clone multiple fragments effectively within a plasmid.
- iii) The number of restriction endonuclease pairs which have 100% co-functionality is limited. For example, various enzymes function optimally within different buffer compositions at a range of temperatures (25 °C to 75 °C) depending on their properties.
- iv) Star activity is when non-specific digestion of the DNA backbone occurs outside of the palindromic recognition site. Star activity can occur with type II restriction endonucleases that are under sub-optimal condition or when the duration of the digest exceeds optimal length.(Cobb et al., 2014).
- v) Restriction digestion-ligation cloning is a sequential approach; digestion of the DNA, gel extraction and purification of fragments, ligation of fragments and transformation in *E.coli*. This process is time consuming and cumbersome when generating larger more complex DNA constructs (Kriz et al., 2010; Torella et al., 2014a).

Due to the limitations mentioned above, this method of cloning is inefficient for MGEV assembly. Yet, there has been significant progression in the last 10 years in the development of new, more effective DNA cloning methods. The range of new viable methods for MGEV assembly can be classified into four categories – *in vitro* and *in vivo* sequence homology, bridging oligonucleotide fragments and alternative restriction digestion-ligation approaches. Figure 1-3 summarises the DNA assembly mechanism for each cloning technique category.

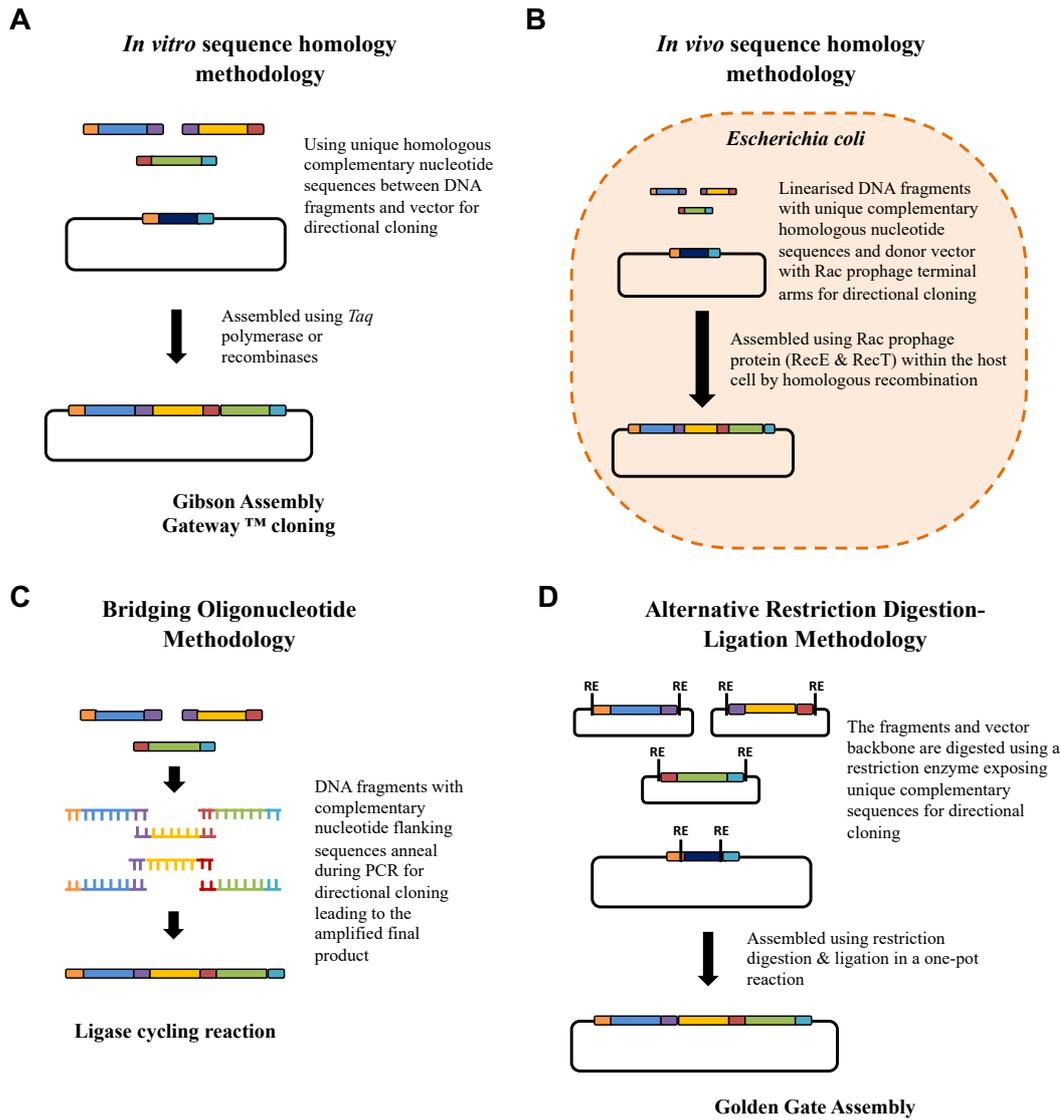


Figure 1-3 - Approaches to clone multiple DNA fragments.

An overview summarising the different approaches developed and described in literature to effectively assemble multiple DNA fragments in a desired order. The approaches can be divided into four categories – *In vitro* sequence homology (A), *In vivo* sequence homology (B), bridging oligonucleotide (C) and alternative restriction digestion-ligation (D) methodologies. In each figure the orange, purple, red and turquoise blocks represent unique complementary nucleotide sequences that control the order of fragment assembly. The blue, yellow and green blocks represent genes of interest to be assembled into a single plasmid. The black circular box represents either a shuttle or recipient vector backbone. The specific mechanisms for cloning are explained in the figure itself.

Current literature was reviewed to identify the different DNA cloning methodologies used for MGEV construction. For example, Guye et al, developed a two tier assembly system for MGEV construction. The first tier used gateway cloning to assemble the transcription units (TUs) comprising of the promoter, recombinant gene and a destination vector. The second tier utilised Gibson assembly to clone multiple TUs together and form a MGEV (Guye et al., 2013). Weber et al also designed a multi-tier system to assemble the TU and MGEV using Golden Gate assembly (Weber et al.,

2011). Whereas, Torella et al developed a methodology using BioBrick™ and Gibson assembly to construct the TU and MGEV respectively (Torella et al., 2014b). Therefore, as the examples indicate along with additional publications reviewed (Andreou and Nakayama, 2018; Halleran et al., 2018; Sarrion-Perdigones et al., 2011), three techniques were repeatedly used for DNA cloning and MGEV constructions, namely Gibson assembly, Gateway cloning and Golden Gate assembly. These methodologies were reviewed to identify their advantages and disadvantages as stated in Table 1-1.

Table 1-1: Discussion of advantages and disadvantages of Gibson assembly, Gateway cloning and Golden Gate assembly.

The table summarises the advantages and disadvantages of Gibson assembly, Gateway cloning and Golden Gate assembly as techniques to construct MGEVs when compared against traditional restriction digestion-ligation cloning.

Cloning Methodology	Advantages	Disadvantages
Gibson assembly (Torella et al., 2014b, 2014a).	<ul style="list-style-type: none"> - High accuracy of assembly due to very specific UNSes - One pot reaction leading to rapid assembly of multiple TUs - Scarless cloning where the UNSes can be integrated between overlapping fragments. 	<ul style="list-style-type: none"> - Complex cloning strategy involving a multi-step process - UNSes have to be specifically designed to minimise secondary structure which could make them susceptible to mRNA cleavage - If UNSes are integrated within the TU sequence, the risk of error or mismatch could cause a frameshift.
Gateway cloning (Chao et al., 2014; Guye et al., 2013; User Guide - Gateway® Technology).	<ul style="list-style-type: none"> - High accuracy of assembly due to specific complementary sequences within the recombination recognition site. - Rapid and HT cloning technique to assemble multiple TUs using recombinases. 	<ul style="list-style-type: none"> - Longer scar sequence (>15 bp) between TUs. - The cloning technique is proprietary to Thermo Fisher Scientific, therefore could involve licensing limitations and compromise cost-effectiveness.
Golden Gate Assembly (Engler et al., 2008, 2009)	<ul style="list-style-type: none"> - One pot reaction leading to rapid assembly of multiple TUs. - The cloning technique is the simplest of the three discussed in the table. - 4 bp overhangs generated due to the mechanism of cleavage by type IIs restriction endonucleases. These sequences can be used to promoter directional cloning by <i>in silico</i> design. Therefore, leading to short scar sequences between TUs. 	<ul style="list-style-type: none"> - The restriction enzyme site – BsaI is frequently present within CDSes or functional genetic elements such as UTRs. - Shorter linker sequences (4bp) can risk mismatched construction.

All three methodologies described in Table 1-1 are viable for the construction of MGEVs. However, Golden Gate assembly is a simpler cloning strategy and aligns well with biopharmaceutical and cell engineering applications in CHO cells, hence was deemed the favourable methodology. This is expanded further in Chapter 3.

1.7. Synthetic Genetic Elements

MGEVs increase our ability to engineer CHO cells in a variety of different aspects (Duportet et al., 2014; Guye et al., 2013; Torella et al., 2014b; Weber et al., 2011). Nonetheless, to achieve the aim of control and predictability during cell engineering and biopharmaceutical production, we can employ various genetic elements such as promoters and transcription terminators within MGEVs. Another aspect of synthetic biology is to enhance the ability to control gene expression by using fully defined and synthetic genetic elements where DNA sequences are fully functional and well characterised. We can use synthetic promoters to control the titratability and level of gene expression. Whereas synthetic transcription terminators can allow us to improve termination and potentially influence gene expression levels or invoke modularity of multiple genes.

1.7.1. Synthetic Promoters

A promoter is a genetic element which regulates the transcription of a gene. The basic structure of a promoter comprises of an upstream enhancer and a core element. The core element is 80 bp in length and consists of conserved binding sequences such as the TATA box, CCAAT-box and GC-box. The element is positioned either 35 bp upstream or downstream of the transcriptional start site (TSS). These motifs recruit and interact with the basal transcription factors (TFs) and forms the pre-initiation complex (PIC). The PIC triggers unwinding of the localised segment of DNA and recruits RNA polymerase II (RNA pol II), the most common polymerase for protein coding gene transcription (Carninci et al., 2006). Figure 1-4 summarises the formation of the PIC and the basal transcription factors involved in eukaryotic transcription.

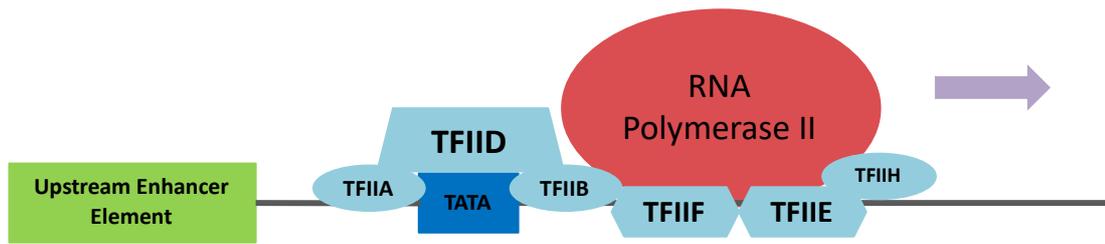


Figure 1-4: The formation of the pre-initiation complex during transcription.

A simplified figure summarising the recruitment of the basal transcription machinery in eukaryotic cells including the general TFs and RNA pol II, along with their interaction with the promoter core element. This figure was adapted from a nature review article (Bywater et al., 2013).

The upstream enhancer element is located at the 5' end of the core element as shown in Figure 1-4. The element comprises of a series of trans-activating TF binding sites. These binding sites are often referred to as transcription factor regulatory elements (TFREs). A TF is a functional regulator for gene expression and can be categorised as an activator or repressor. The interaction between TFs and their respective TFREs can control transcription frequency and regulate the core promoter. The recruited TFs interact with each other and the basal transcription machinery to regulate the formation of PIC (Blazek et al., 2012). The main limitations of a naturally occurring endogenous or viral promoter are its undesirable size (long sequence elements) and unpredictable expression dynamics (due to not being fully annotated and understood). Synthetic promoters address these unfavourable characteristics (Brown et al., 2017).

A synthetic promoter has the same structure and elements as a standard promoter however, all the DNA elements are defined and characterised. Figure 1-5 shows the basic layout of a synthetic promoter including a pre-designed proximal TFRE region and a minimal core element upstream of the TSS.

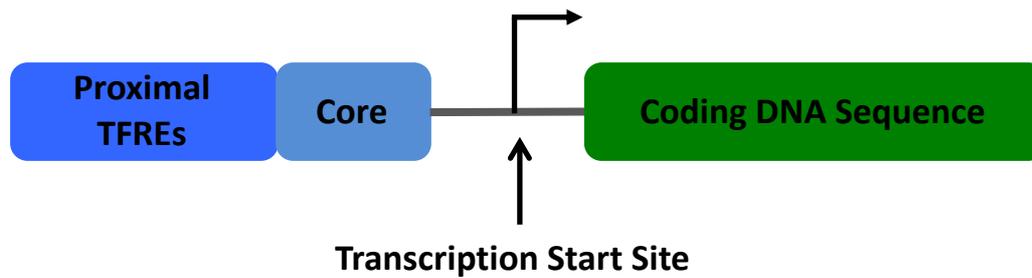


Figure 1-5: Structure of a synthetic promoter.

The figure shows the simplified structure of a synthetic promoter including proximal TFREs which function as a transactivating element and a minimal core element which recruits the basal transcription machinery.

The main method of designing synthetic promoters is within fully synthetic systems consisting of a synthetic promoter and synthetic TFs such as zinc finger (Gaj et al., 2013), transcription activator-like effectors (Perez-Pinera et al., 2013), chimeric (Rössger et al., 2014) and CRISPR transcription factors (Chavez et al., 2015). These systems control gene expression by the unique binding specificity of synthetic TFs to their respective TFREs on the synthetic promoter and minimal cross-interaction with the host cell's endogenous system. The fully synthetic system has demonstrated tuneable inducible gene expression (Brown et al., 2017). Yet, there are drawbacks, with the main flaw in the system being the metabolic burden it may invoke on the host cell by having to recombinantly express the synthetic TF concurrently, which is less desirable for gene therapy and biopharmaceutical production (Brown et al., 2017).

An alternative to a fully synthetic promoter system is a constitutive synthetic promoter which uses the host cell's endogenous TF repertoire. One example designing this class of synthetic promoters, as demonstrated by Brown et al, is a three step process – i) profiling TF expression in the host cell such as CHO cells; ii) identifying TFREs that function mutually exclusively; iii) determining the relative transcriptional activity of the TF-TFRE interaction (Brown et al., 2017). Alternative approaches include randomised DNA sequence or synthetic oligonucleotide repeat assemblies of known *cis*-regulatory elements which have led to synthetic promoters for a range of host systems such as *Corynebacterium* (Yim et al., 2013), *Saccharomyces* (Blazeck et al., 2012), *Pichia* (Stadlmayr et al., 2010), *Streptomyces* (Seghezzi et al., 2011) and mammalian cells (Ferreira et al., 2011) including CHO cells (Brown et al., 2014).

The advantages of using constitutive synthetic promoters over viral/endogenous promoters include a design-based approach for synthetic promoter construction and generating elements of known functionality. Furthermore, using transcriptomic data of the host cell to identify functional TFREs can aid construction of rationally designed promoters as opposed to the complex and undefined nature of endogenous promoters. The process yields promoters with defined and predictable expression levels by varying the combination of TFRE blocks, allowing for tighter control of gene expression with the potential for titratability of recombinant gene expression. Furthermore, a fully defined and characterised synthetic promoter structure including defined TFRE blocks and minimal core reduces the complexity and size of the promoter, generating a more standardised element. In addition, a well characterised library of synthetic promoters allows for identifying combinatorial pairings of promoters where functionality may not be impeded by promoter-promoter interference based on avoiding overlapping TFRE blocks.

Accordingly, using a library of constitutive synthetic promoters designed for CHO cells will facilitate the ability to titrate recombinant gene expression levels. As well as the potential use of combinatorial approaches in a MGEV to have tighter stoichiometric control of multiple recombinant genes when performing CHO cell engineering.

1.7.2. Synthetic Transcription Terminators

Transcription termination involves a range of mechanisms to halt transcription and assist in the dissociation of the RNA pol II from the sense strand at the 3' end downstream of the coding DNA sequence (CDS). It is essential in gene expression to modularise transcription units, avoid transcriptional interference, promote gene expression stability and assist in efficient cellular localisation of the mRNA transcript (Gasnov et al., 2015; Porrua and Libri, 2015; Proudfoot, 2016). Efficient transcription termination has shown to enhance recombinant protein expression by efficiently generating stable mRNA transcripts (West and Proudfoot, 2009). Furthermore, from a more mechanistical approach, efficient dissociation of the RNA pol II would contribute towards modularity within a polycistronic gene cassette.

There are two proposed models for the transcription termination of coding RNA, which are the allosteric and torpedo models (Eaton et al., 2018; Porrua and Libri, 2015). The allosteric model hypothesises that once the polyadenylation signal (PAS) site has been transcribed and the termination complex has assembled, the elongation complex changes conformation by loss of the elongation factors. This causes the dissociation of the RNA pol II from the sense strand. The torpedo model hypothesises that the 3' end cleavage of the mRNA precursor by the polyadenylation (pA) complex allows the entry of the 5'-3' exonuclease called Xrn2. The exonuclease degrades the nascent RNA up to the RNA pol II triggering the dissociation of the polymerase. Literature has also suggested that DNA elements can pause the elongation complex allowing for the exonuclease to catch up and initiate transcription termination (Libri, 2015; Porrua and Libri, 2015).

There are a range of mechanisms and DNA elements which contribute towards the transcription termination process; including pA, co-transcriptional cleavage (CoTC) and pause-mediated termination. Nevertheless, the specific mechanism and regulation of the different aspects of termination are still poorly understood (Proudfoot, 2016). This section will expand on each of the mechanisms and how they might contribute towards transcription termination based on existing literature.

1.7.2.1. Polyadenylation signal-mediated termination

The main function of pA is the modification of the pre-mRNA transcript by the addition of a polyA tail on the 3' end for mRNA stability. However, *in vitro* assays have demonstrated that the PAS triggers a conformational change in the RNA pol II which pauses and eventually dissociates from the sense strand and terminates transcription as shown in Figure 1-6 (Zhang et al., 2015).

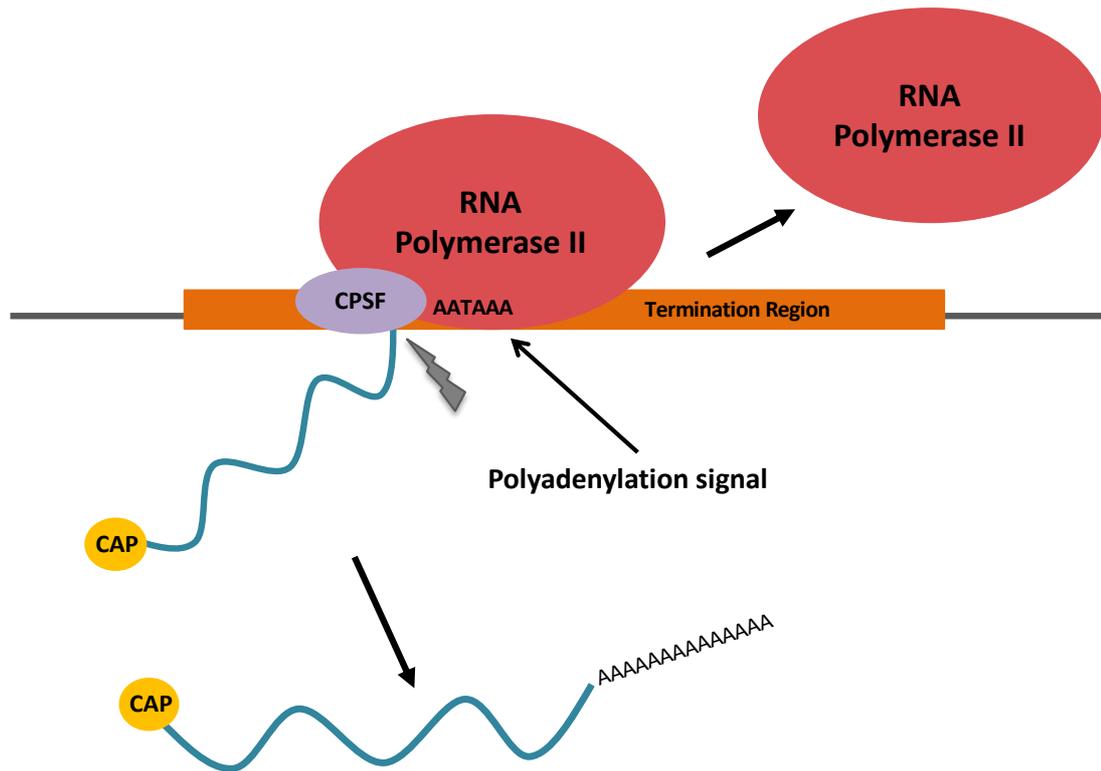


Figure 1-6: Polyadenylation signal-mediated termination.

Figure demonstrating how the elongation complex interacts with the cleavage and polyadenylation specificity factor (CPSF) which performs cleavage and polyadenylation of the pre-mRNA transcript and how the RNA polymerase II is dissociated from the sense strand. The grey symbol represents the site of cleavage of the precursor mRNA transcript.

An alternative mechanism suggests that the PAS slows down/pauses the RNA pol II by the recruitment of the pA basal machinery and a combination of the pre-mRNA transcript being cleaved and the function of the 5'-3' exonuclease causes transcription termination. Once the precursor mRNA is cleaved, the RNA pol II continues to transcribe nascent RNA and the Xrn2 exonuclease degrades this unwanted nascent RNA. When the Xrn2 meets the RNA pol II, termination is triggered (Proudfoot, 2016). The specifics of this mechanism is still unclear and the most effective terminating transcription mode is yet to be clearly determined as conflicting results have been published from different research groups (Fong et al., 2015; Libri, 2015; Zhang et al., 2015).

1.7.2.2. Co-transcriptional cleavage element-mediated transcription termination

A CoTC element is a DNA sequence which is located downstream of the PAS and human genome analysis has inferred it to be a conserved A/T rich element (Nojima et al., 2013). The mechanism for CoTC-mediated transcription termination adopts the torpedo model. This occurs by the CoTC element causing the cleavage of the RNA

strand forming two products – a 5' cleaved pre-mRNA transcript which is polyadenylated and a 3' cleaved nascent RNA strand with an unprotected 5' end phosphate group (Nojima et al., 2013; West et al., 2008). The unprotected 5' end nascent RNA is then the substrate for Xrn2 which leads to the dissociation of the RNA pol II (West et al., 2008) as shown in Figure 1-7.

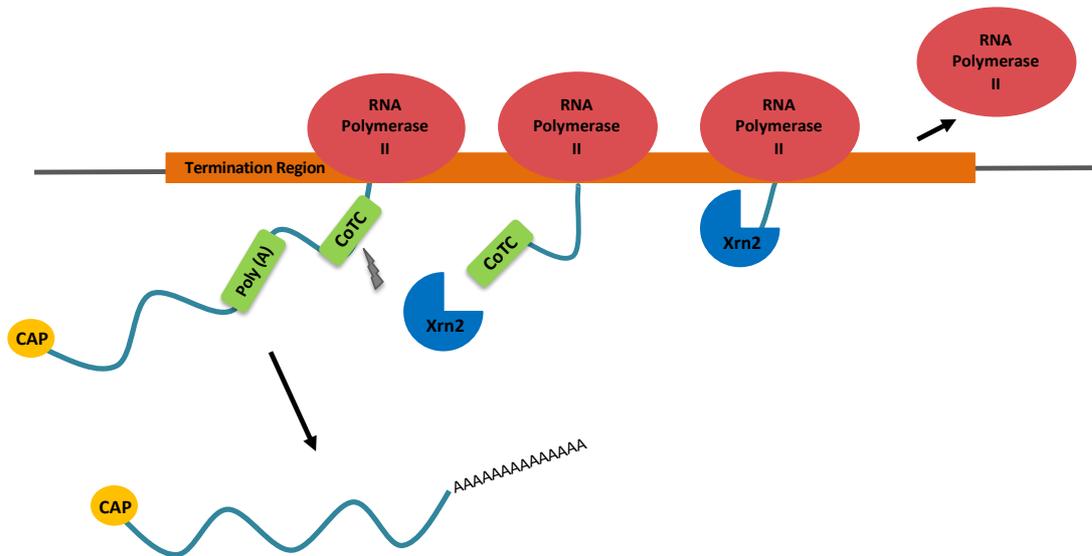


Figure 1-7: Co-transcriptional cleavage (CoTC) element-mediated transcription termination.

Diagram showing how a CoTC element would cause cleavage of the pre-mRNA transcript which is polyadenylated. This causes the recruitment of 5'-3' exonuclease Xrn2 to degrade the unprotected nascent RNA transcript. Once the Xrn2 and RNA pol II interact, it causes the RNA pol II to dissociate from the sense strand. The grey symbol represents the cleavage of the precursor mRNA transcript exposing the 5' end of the nascent RNA.

1.7.2.3. Pause element-mediated transcription termination

A pause element is a DNA sequence which causes the RNA pol II to stop on the sense strand and halt transcription. In effect, this element stops the polymerase from running onto the next gene. One pause element investigated is a MAZ (MYC-associated zinc finger protein) DNA binding element. This element is a G-rich nucleotide sequence – G₅AG₅. Literature has shown that the MAZ element pauses transcription by stopping the RNA pol II and allows the Xrn2 5'-3' exonuclease to degrade the 5' nascent RNA and kinetically catch up to the polymerase initiating transcription termination. This was demonstrated in transfected HeLa cells and compared against using CoTC-mediated termination (Gromak et al., 2006). The specific mechanism by which the MAZ element pauses transcription is still unclear, however studies have shown that it requires a strong PAS for successful termination as shown in Figure 1-8.

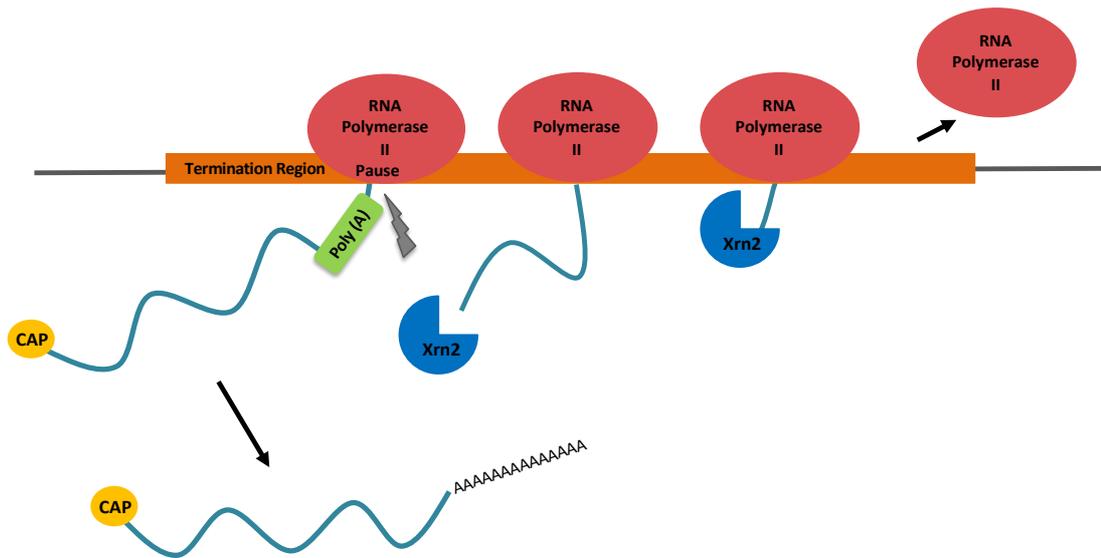


Figure 1-8: Pause element-mediated transcription termination.

Figure demonstrating the hypothesised mechanism of when the RNA pol II encounters the pause element and how cleavage of the pre-RNA transcript recruits the 5'-3' exonuclease (Xrn2) to degrade the nascent RNA. This leads to transcription termination by the torpedo model. The grey symbol represents the cleavage of the precursor mRNA transcript exposing the 5' end of the nascent RNA.

1.7.2.4. Ribozyme-mediated transcription termination

Ribozymes are RNA molecules that form a secondary and tertiary structure which can catalyse the adjacent phosphodiester backbone and therefore have self-cleaving capabilities (Jimenez et al., 2015). There are a growing number of ribozymes identified in a range of different organisms. The six best described ribozymes include hepatitis δ virus (HDV), hammerhead, hairpin, *Neurospora varkud* satellite, glucosamine-6-phosphate synthase (*glmS*) and twister (Jimenez et al., 2015). Literature has suggested that mRNA processing (5' capping, splicing and pA) are interdependent and occur co-transcriptionally (Bird et al., 2005). Furthermore, it is possible that ribozymes contribute towards RNA cleavage as part of the 3' end processing of mRNA and in turn impact on translocation of mRNA to the cytoplasm for translation. Previous research has shown the presence of a self-cleaving ribozyme can still cleave unadenylated pre-mRNA transcript even though the pA process has failed (Bird et al., 2005; Dower et al., 2004).

Hammerhead ribozymes have been comprehensively studied as it was the first ribozyme to have a crystal structure allowing for mechanistic understanding of the molecule. The ribozyme comprises of three helices branching from the catalytic core which is made up of 15 essential nucleotides for functionality (Bird et al., 2005). This

approach to transcription termination employs the efficient and precise characteristics of a hammerhead ribozyme. It is hypothesised that the self-cleaving function will cleave the pre-mRNA allow for 3' end processing, while exposing the unprotected nascent RNA for Xrn2-mediated degradation, leading to torpedo modelled transcription termination to occur.

1.7.2.5. Existing Synthetic Terminator Elements

Similar to synthetic promoters, synthetic transcription terminators are smaller elements comprising of defined conserved functional groups to generate elements which facilitate control and stability of transcription termination. From the literature, it was ascertained that synthetic transcription terminators have been investigated and designed for *E.coli* (Chen et al., 2013) and yeast (Curran et al., 2015). A clear advantage of designing synthetic transcription terminators is reducing the reliance on native scaffold and thus, reducing the homology to the host genome and avoiding unwanted homologous recombination within the host cell genome during recombinant expression (Curran et al., 2015). These elements are designed based on minimal, conserved functional motifs, making sequences shorter, and ergo benefiting applications related to large multi-gene constructs. Finally, the well-defined and characterised nature of these DNA sequences will allow for more predictability and tunability of gene expression. In summation, by discussing the different modes of transcription termination above and identifying the conserved functional elements, we have conceived the potential for designing synthetic transcription terminators for mammalian cells such as CHO cells, which would be highly beneficial particularly in providing more control and predictability towards cell engineering.

1.7.3. Context dependent synthetic genetic element design

When designing synthetic genetic elements for various applications such as initiating or terminating transcription, various context specific properties can influence performance. For example, the cellular background (Brown and James, 2017) in which the genetic part is used within, employing parts within a plasmid or chromosomal context (Brown and James, 2016) and interaction with other genetic elements such as insulator elements (West et al., 2002) can all impact on the elements activity.

In the case of cellular background specific performance, synthetic promoters are designed using a specific cell line's transcriptomic data (e.g. CHO cells), and evaluated by *in vitro* testing of TF and cognate TFRE performance within the same host cell line (Brown et al., 2017). Therefore, promoter activity and performance is context specific to the host cell and transcriptional activity would potentially vary in other host cell types due to different TF and co-regulators abundances (Schlabach et al., 2010). However, comparable promoter performance may be observed between CHO cell subtypes as TF and TFRE cognate behaviour maybe similar, leading to consistent synthetic promoter activity in CHO-K1 derived cell lines (Brown et al., 2014, 2017; Johari et al., 2019).

Conversely, transcription termination is a conserved process in mammalian cells and transcription terminators derived from human, bovine, and viral sources have functioned interchangeably in various mammalian cell types such as HeLa, CHO and HEK (Hunter et al., 2019; West and Proudfoot, 2009). Moreover, polyadenylation-mediated transcription termination (allosteric model) is conserved within eukaryotic cells such as yeast and mammalian cell types (Costa et al., 2010; Curran et al., 2015). Whereas, other mechanisms such as CoTC-mediated termination was identified in mouse and human cells, therefore, could be conserved in mammalian cells. For instance, a CoTC element was first identified downstream of a human β -globin gene and a similar element was located downstream of a mouse serum albumin gene (West and Proudfoot, 2009; West et al., 2006). Therefore, designing synthetic transcription terminators using conserved functional motifs could potentially function interchangeably within the context of different cell backgrounds.

Context dependent performance of some genetic elements can vary within a plasmid and chromosomal integrated format influencing transient and stable gene expression. For example, transcription terminators have been employed in both transient and stable gene expression conditions and perform their function consistently for recombinant gene expression (Costa et al., 2010). Moreover, the inference from human genome derivatives suggests transcription terminators would work consistently within both a plasmid and chromosomal context (White et al., 2013). Therefore, synthetic transcription terminators performance may not be context dependent within plasmid or chromosomal integrated format.

Conversely, the local transcriptional environment within a circular plasmid format and an integrant within the genome can have differing effects on promoter performance influencing gene expression levels. For example, potential interactivity of transactivators or repressors both within the proximal and distal region from the gene of interest or the chromatin state can influence promoter transcriptional activity within the context of stable gene expression (Voss and Hager, 2014). Whereas, within a plasmid context, promoter activity could be influenced by neighbouring promoters and localised competition for common TFs leading to promoter squelching (Huliák et al., 2012). Therefore, the design of promoters (TFRE composition) and introduction of other genetic elements (e.g. insulators) can differ within the context of a plasmid or integrant within the genome to achieve predictable gene expression (Huliák et al., 2012; Yahata et al., 2007).

The introduction of insulators can specifically impact the local gene expression environment and promoter activity. This is because insulators perform two functions – i) barrier from the formation of heterochromatin and ii) blocking distal enhancer TFs from initiating transcription (West et al., 2002). Two commonly known insulators are *gypsy* and 5' chicken hypersensitivity site 4 (cHS4) insulator element identified in *Drosophila* and *Gallus* respectively, whereas other insulator-like elements have been identified in prokaryotes and eukaryotes but poorly characterised (Levy et al., 2017; West et al., 2002). More recently, the application of insulators was used as a means to stop interactivity between promoters in a tandem series within synthetic gene circuits (Liao et al., 2018). Unfortunately, the natural element is quite large (1.2 kb) leading to larger plasmid sizes when utilised which is unfavourable on plasmid transfectability and cellular toxicity (Hornstein et al., 2016; Lesueur et al., 2016). More recently, a core motif of CCCTC-binding factor (CTCF) site of 250 bp within the cHS4 element has shown to be functional, reducing the impact on plasmid size (Uchida et al., 2013; Yusufzai and Felsenfeld, 2004). However, there has been limited characterisation of alternative insulator elements, therefore, a single element would have to be used multiple times within a MGEV context. This increases the risk of undesired recombination events within a MGEV leading to unstable gene expression.

A concern towards consistent stable gene expression is the emergence of multiple repeat sequences. This can often lead to undesirable homologous recombination

leading to potential gene silencing (Brown and James, 2016; Jasin and Rothstein, 2013). Therefore, a consideration specific to designing and using genetic elements is developing different sequence variants or reducing repeat motifs within the composition of the genetic part, in turn, reducing sequence homology and averting homologous recombination-mediated gene expression instability (Brown and James, 2016). Moreover, this also applies in the context of a MGEV where the requirement of multiple promoters, transcription terminators and UTRs to regulate multiple recombinant genes simultaneously is necessary, therefore, repeat elements and motifs are unfavourable (Brown and James, 2016). As a result, specific context of promoter, terminators or MGEV parts need to be considered to achieve the appropriate and predictable function.

1.8. Thesis Aims & Overview

Elements of synthetic biology can be adopted to address some of the limitations and demands within CHO cell line engineering in both the biopharmaceutical industry and academia. As highlighted from previous sections, tools such as constitutive synthetic promoters can allow for titratable, precise and predictable control of recombinant gene expression, and coupled with functional synthetic transcription terminators to help improve efficiency of termination due to potential higher flux of RNA pol II-mediated transcription from synthetic promoters. These synthetic elements combined with a HT means of assembling MGEVs will allow for more complex, stoichiometrically-defined and predictable CHO cell engineering to be performed.

This research project can be divided into three major aims:

1. To design and develop a robust multi-gene expression vector for CHO cell engineering.
2. To demonstrate titratable expression of multiple recombinant genes at defined stoichiometric ratios using synthetically designed components.
3. To test the functionality of the MGEV for different proof of concept applications.

These aims were investigated throughout the duration of the PhD and compiled into the thesis in the following the chapters.

Chapter two details the materials and methods that were used for the various studies discussed in the project. Chapter three explores the design, verification and characterisation of a MGEV for CHO cell engineering. This includes a comprehensive description of the design space and features of a robust, multi-functional and simplistic MGEV system that is applicable in both industry and academia. This was achieved by a combination of literature analysis and collaborating with MedImmune. The outcome led to the *in silico* design of the TU and pExp-Vec-GG recipient vector backbone and then the *de novo* synthesis of the genetic components. Finally, the chapter demonstrates the *in vitro* verification of the designed MGEV system, as well as the successful expression of multiple recombinant genes (fluorescent proteins). This highlighted the development of a robust protocol for assembling functional MGEVs. Further characterisation of the gene expression dynamics within the polycistronic cassette demonstrated position-dependent repression hypothesised to be a consequence of transcriptional interference.

Chapter four explores the application of synthetic promoters within a MGEV in an aim to titrate gene expression of multiple recombinant genes. This was achieved by adapting a validated and defined synthetic promoter library for compatibility with the MGEV system. A subset of synthetic promoters were further characterised for combinatorial functionality and the emergence of promoter squelching. Additionally, further characterisation of the positional effect within the polycistronic cassette demonstrated the effect to be transcriptional-strength dependent. Finally, the chapter characterises a library of 27 MGEVs with various combinations of a low, medium and high strength synthetic promoter to generate defined gene expression stoichiometric ratios, highlighting a database of defined titrated expression of multiple recombinant genes for future multi-gene engineering strategies.

Ultimately, chapter five demonstrates the functionality of a MGEV as a tool for developing synthetic genetic components and engineering CHO cell factories for biopharmaceutical production. This includes an extensive discussion of the design of synthetic transcription terminators applying various termination mechanisms (as discussed previously). The library of synthetic transcription terminators was then rapidly screened to measure the functionality of the terminator using MGEVs. The screen identified one potential synthetic pA element, but also assisted in narrowing the

design space and identifying critical design features for functionality. The titrated co-expression of a UPR TF called X-box binding protein-1 spliced (XBP-1s) alongside a DTE mAb to increase titres was additionally explored. This was based on previous published data and performed in collaboration with MedImmune, using their proprietary targeted integration (TI) expression platform and fed-batch overgrow (FBOG) systems. The data demonstrated the successful stable engineering of CHO cells to increase mAb titre by increasing the growth rate and biomass generation of the host cell population. This, therefore, highlights the functionality of a MGEV as a tool for improving biopharmaceutical production by multi-gene engineering strategies.

Chapter 2

2. Materials & Methods

2.1. Synthesis and cloning of transcription unit and multi-gene expression vectors

2.1.1. Gene and oligonucleotide de novo synthesis

The gene synthesis was outsourced to GeneArt (Life Technologies, Thermo Fisher Scientific, Paisley, UK) using their on-line portal. Coding DNA sequence (CDS) was optimised for *Cricetulus griseus* using GeneArt's proprietary algorithm. The oligonucleotides for the *in silico* designed transcription terminator library (39 variants) were also *de novo* synthesised by GeneArt which included post-synthesis cloning into the transcription unit (TU) -1 plasmid encoding for secreted alkaline phosphatase (*SEAP*).

2.1.2. Restriction digest – ligation plasmid cloning

Various genetic parts such as the proximal element of the promoter, CDS and transcription terminators (both polyadenylation (pA) and terminator elements) were cloned using restriction digest-ligation cloning. This was achieved by digesting 1 µg of DNA with 1 µl of each NEB high fidelity restriction endonuclease (New England Biolabs (NEB), Hitchin, UK) and 3 µl of CutSmart buffer (NEB) and made up to a final volume of 30 µl using nuclease-free water (Qiagen, Manchester, UK). The digests were incubated at 37 °C for 1.5 h. The digested DNA was separated by molecular weight using gel electrophoresis where a 1% agarose gel was stained with

SybrSafe (Life Technologies, Thermo Fisher Scientific) in 1x Tris-Acetate EDTA (TAE) buffer for 1 h at 100 V. The respective insert and vector fragments were excised using a scalpel and purified from the agarose using a Minelute gel extraction kit (Qiagen) according to the manufacturer's protocol. The insert and vector fragments were ligated at either a 6:1 or 3:1 molar ratio of insert: vector using the following equation:

$$\text{Insert fragment (ng)} = \text{ratio excess} \times \frac{\text{length of insert (bp)}}{\text{length of vector (bp)}} \times \text{Vector fragment (ng)}$$

The appropriate mass of insert and vector were mixed and made up to a final volume of 9 μl using nuclease-free water (Qiagen). The ligation reaction was performed using NEB's Quick ligation kit according to the manufacturer's protocol. The ligation reaction was incubated at room temperature (RT) for 10 min.

2.1.3. Golden Gate assembly

Multi-gene expression vectors (MGEVs) were assembled using a Golden Gate assembly kit (NEB). This was achieved using different variants of the recipient vector backbone (pEXP-Vec-GG and pEXP-Vec-GG_TI) combined with a determined number of TU variants encoding different recombinant genes. A MGEV assembly reaction was set up with 75 ng of the recipient vector backbone and 2:1 molar ratio of each TU. The DNA was combined with 2 μl of Golden Gate buffer (10x) and 1 μl of Golden Gate assembly mix and made up to 20 μl with nuclease-free water (Qiagen). The reaction mix was incubated in a thermocycler (Applied Biosystems, Thermo Fisher Scientific) at 37 °C for 1 h followed by 55 °C for 5 min according to the manufacturer's protocol.

2.1.4. Transformation of plasmid DNA for cloning, multi-gene expression vector construction or DNA amplification

An aliquot of 1 to 10 μl of purified DNA or the ligation reaction was transformed in a 50 μl aliquot of Subcloning Efficiency™ DH5 α ™ chemically competent *Escherichia coli* cells (Life Technologies, Thermo Fisher Scientific). This was performed by incubating on ice for 7-10 min, followed by heat shock at 42 °C for 20 secs and immediately incubated on ice for 2 min. An aliquot of 950 μl of unsupplemented Luria-

Betani (LB) broth (Fisher Scientific, Thermo Fisher Scientific) was added to the transformation and incubated at 37 °C for 1 h in a thermomixer (Eppendorf, Stevenage, UK) with orbital shaking at 700 rpm. The transformation was centrifuged at 2000g for 2 min to pellet the cells and 900 µl of supernatant was removed, and the pellet was resuspended in the remaining 100 µl of supernatant. The cell suspension was then plated on 15 ml LB agar (Fisher Scientific, Thermo Fisher Scientific) supplemented with 100 µg/ml ampicillin (Sigma Aldrich, Dorset, UK) and incubated for 16 h (overnight) at 37 °C.

2.1.5. Restriction digest colony screen of cloned plasmids

Two or three colonies per cloned plasmid were screened by inoculating into 10 ml of LB Broth (Fisher Scientific, Thermo Fisher Scientific) supplemented with 100 µg/ml ampicillin (Sigma Aldrich). The cultures were incubated at 37 °C with orbital shaking at 200 rpm between 6 and 16 h. Bacterial culture (8.5 ml) was centrifuged at 6800g for 3 min and the supernatant discarded. The plasmid DNA was extracted and purified using Qiagen's QIAquick Spin Miniprep kit (Qiagen) according to manufacturer's protocol. A total of 600 ng of plasmid DNA was digested using 0.5 µl of various high-fidelity restriction endonucleases (NEB) depending on the plasmid of interest, 1.5 µl of CutSmart buffer (NEB) and made up to a final reaction volume of 15 µl using nuclease-free water (Qiagen). The digests were incubated at 37 °C for 1 h followed by fragment size analysis using a 1% agarose gel stained with SybrSafe (Life Technologies, Thermo Fisher Scientific) in 1 x TAE buffer at 100 V for 1 h by the process of gel electrophoresis. The agarose gel was imaged using an UV transilluminator on a gel imaging system (UVP, Cambridge, UK).

2.1.6. Plasmid DNA amplification

Plasmid DNA amplification was performed at either a mini prep or maxi prep scale. Mini preps were performed by inoculating 10 ml of LB Broth (Fisher Scientific, ThermoFisher Scientific) supplemented with 100 µg/ml ampicillin (Sigma Aldrich) and incubated for 16 h at 37 °C with orbital shaking at 200 rpm. The plasmid DNA was extracted and purified using Qiagen's QIAquick Spin Miniprep kit (Qiagen) according to the manufacturer's protocol. Maxi preps were performed by inoculating a 5 ml starter culture of LB Broth supplemented with 100 µg/ml ampicillin and

incubated for 6 h at 37 °C with orbital shaking at 200 rpm. The starter culture was then used to inoculate a 100 ml culture of LB Broth supplemented with 100 µg/ml ampicillin and incubated for 16 h at 37 °C with orbital shaking at 200 rpm. The cultures were harvested by centrifugation at 6000g for 15 min with the supernatant discarded. The plasmid DNA was extracted and purified using Qiagen's Plasmid Plus Maxi kit (Qiagen) according to manufacturer's protocol. The purified plasmid DNA was quantified, and purity assessed using the Nanodrop spectrophotometer 2000 (Thermo Scientific, Thermo Fisher Scientific)

2.1.7. Modification of synthetic components using PCR

The proximal regions of the synthetic promoter library were modified by introducing a BamHI and NheI restriction site at the 5' and 3' end respectively. This was achieved by designing a series of primers as shown in Appendix B. The fragments were modified and amplified using a Phusion PCR kit (NEB) and thermocycler (Applied Biosystems) according to manufacturer's protocol for 35 cycles at various primer annealing temperatures depending on the primer pair. The amplified DNA was purified using a QIAquick PCR purification kit (Qiagen) according the manufacturer's protocol.

2.2. Cell culture methodologies to screen transcription units and multi-gene expression vectors

2.2.1. Cell culture revival, maintenance and cryopreservation

A vial of MedImmune's proprietary CHO host cell line called Medi-CHO was revived by thawing in a 37 °C water bath and diluted with 43.5 ml of 1x CD-CHO (Gibco™, Thermo Fisher Scientific) supplemented with 6 mM L-glutamine (Gibco™, Thermo Fisher Scientific) media. The diluted cells were centrifuged at 130g for 5 min. The supernatant was discarded, and the cell pellet was resuspended in 10 ml of 1x CD-CHO supplemented with 6 mM L-glutamine. The viable cell concentration was quantified using a ViCell XR (Beckman Coulter, High Wycombe, UK) and the cells diluted to 3×10^5 viable cells/ml. A 30 ml culture was transferred into a 125 ml vented Erlenmeyer flask (Corning, Flintshire, UK) and incubated at 37 °C, 5% (v/v) CO₂ with orbital shaking at 140 rpm for 72 h.

The MedI-CHO cell line was maintained at exponential growth by subculturing every 3-4 days in 1x CD-CHO supplemented with 6mM L-glutamine and incubated at 37 °C, 5% (v/v) CO₂ with orbital shaking at 140 rpm. The culture's cell concentration and viability was quantified using a ViCell XR by using a trypan blue exclusion methodology. The cell line was then subcultured by seeding at 2×10^5 viable cells/ml in a final volume of 30 ml, 60 ml, 100 ml and 200 ml in a 125 ml, 250 ml, 500 ml and 1000 ml vented Erlenmeyer flask (Corning) respectively. The cell line was maintained for up to 20 passages before being discarded.

A working cell bank was established by culturing the MedI-CHO cell line for a minimum of 4 passages before cryopreservation. An appropriate volume of cell suspension was centrifuged at 200g for 5 min. The supernatant was discarded, and the cell pellet was resuspended into 1x CD-CHO mixed with 7.5% (v/v) Dimethyl sulfoxide (DMSO) solution (Sigma-Aldrich) at a final concentration of 1×10^7 viable cells/ml. 1.5 ml of cell suspension was aliquoted into each cryovial equating to 1.5×10^7 cells per vial. The cryovials were transferred into a Mr. Frosty (VWR, Lutterworth, UK) and stored at -80 °C overnight. The Mr. Frosty controls the rate of temperature reduction at 1 °C per minute. The cryovials were then stored in liquid nitrogen for long term storage.

2.2.2. High throughput 96 well plate-based transient expression

A high throughput (HT) 96 well plate-based transfection protocol for rapid transient expression was developed using electroporation. A DNA mix was prepared consisting of 100 to 1000 ng of plasmid DNA (multiple single gene vectors (SGVs) or a MGEV) and nuclease-free d.H₂O (Qiagen) in a final volume of 2.5 µl including a 25% excess. The DNA mix was combined with 7.5 µl nucleofection solution (prepared according to Amaxa SG Cell Line IV 96 well electroporation kit (Lonza, Basel, Switzerland) protocol). The cell suspension was prepared by centrifuging 2.33×10^6 viable cells per well at 200g for 5 min. The supernatant was discarded, and the cell pellet was resuspended in 15 µl per well of nucleofection solution including 25% excess. A 15 µl aliquot per well of cell suspension was transferred to the DNA mix prepared equating to a final DNA-cell suspension mix of 25 µl in a 96 well round bottom plate (Thermo Scientific, Thermo Fisher Scientific). A 20 µl aliquot of the DNA-cell

suspension mix was transferred into a 96 well electroporation plate (Lonza). The plate was then electroporated using the Amaxa Nucleofector 96 Shuttle System (Lonza) with the following protocol settings - SG Cell Line IV, FF-158. Post-electroporation, 80 μ l of pre-warmed 1x CD-CHO (Gibco™, Thermo Fisher Scientific) supplemented with 6 mM L-glutamine (Gibco™, Thermo Fisher Scientific) was added to each well. The transfected cells were cultured in 96 well flat-bottom plates (Thermo Scientific, Thermo Fisher Scientific) in a final culture volume of 200 μ l per well by transferring 20 μ l of transfected culture to 180 μ l of pre-warmed 1 x CD-CHO supplemented with 6 mM L-glutamine per well. The culture was then incubated at 37 °C with 5% (v/v) CO₂ and humidity for a period of 24 or 72 h depending on the experiment.

2.2.3. Generation of stable pool cell lines using targeted integration

A protocol was developed for generation of stable pool cell lines using targeted integration (TI) by Cre-LoxP mediated recombination. The MedImmune proprietary TI host cell line was cultured in 1 x CD-CHO (Gibco™, Thermo Fisher Scientific) supplemented with 6 mM L-glutamine (Gibco™, Thermo Fisher Scientific) and antibiotic selection. The host cell was cultured every 3 to 4 days at a seeding density of 3×10^5 viable cells/ml at 37 °C in 5% v/v CO₂, 85% humidity and 140 rpm orbital shaking. The TI host cell was seeded at 5×10^5 viable cells/ml in 1x CD-CHO supplemented with 6 mM L-glutamine with the antibiotic selection omitted the day before transfection. Transfections were performed using Lonza's Amaxa nucleofector system, whereby 1×10^7 viable cells were aliquoted and centrifuged at 90g for 10 min. The supernatant was discarded, and the cell pellet was resuspended in 200 μ l of nucleofection solution V (Lonza). The MGEV DNA and Cre recombinase vectors were mixed together to a final quantity of 5 μ g and combined with the cell suspension. A 110-120 μ l aliquot of the DNA-cell mixture was transferred into each cuvette (2 cuvettes were used per transfection). The cuvettes were electroporated using the Amaxa nucleofector 4D system (Lonza) using the U024 program. The transfected cells were immediately transferred from the cuvette into 40 ml of pre-warmed 1x CD-CHO supplemented with 6 mM L-glutamine and cultured for 48 h in T175 flasks (ThermoFisher Scientific) at 37 °C in 5% (v/v) CO₂ in a static humidified incubator. After 48 h, selection was performed by adding 40 ml of pre-warmed 1x CD-CHO supplemented with 6 mM L-glutamine and antibiotic supplement. After the

introduction of selection, the transfected cultures were incubated at 37 °C in 5% (v/v) CO₂ in a static humidified incubator for up 18 days.

The recovery of the transfected cells post-selection was monitored by measuring cell viability and density using a ViCell XR (Beckman Coulter). When cell densities were $> 3\text{-}4 \times 10^5$ viable cells/ml and viabilities were $>75\%$, the cultures were transferred into 30 ml of 1x CD-CHO supplemented with 6 mM L-glutamine and antibiotic supplement in a 125 ml Erlenmeyer flask (Corning) and incubated at 37 °C in 5% (v/v) CO₂ and 85% humidity with orbital shaking at 140 rpm. The stable pools were subcultured between 2-4 days after transfer to increase viability to $>90\%$. The pools were then subcultured for a further two passages before being cryopreserved as described previously.

2.2.4. Fed-batch overgrowth and cell culture characteristic analysis

The stable pools generated to co-expression a difficult-to-express (DTE) monoclonal antibody (mAb) by TI using MedImmune's proprietary cell line were characterised for growth and titre. This was achieved by performing a 13 day fed-batch overgrowth (FBOG) using MedImmune's proprietary inhouse culturing media and feed strategy. The viable cell concentration and viability was determined using a ViCell XR (Beckman Coulter) and mAb titres were measured using MedImmune's proprietary Protein A HPLC protocol. The integral viable cell density (IVCD) was calculated at each time point using the following equation:

$$IVCD = \left(\frac{V_0 + V_1}{2} \times \Delta t \right) + IVCD_{t-1}$$

Where V_0 = viable cell concentration ($\times 10^6$ cells/ml) of first sample; V_1 = viable cell concentration ($\times 10^6$ cells/ml) of second sample, Δt = difference in days between first and second sample. The specific productivity of mAb (qP) expression was calculated using the following equation:

$$qP = \left(\frac{T_1 - T_0}{(V_0 + V_1)/2} \right) \div \Delta t$$

Where T_0 = mAb titre (mg/L) of first sample, T_1 = mAb titre (mg/L) of second sample, V_0 = viable cell concentration ($\times 10^6$ cells/ml) of first sample; V_1 = viable cell concentration ($\times 10^6$ cells/ml) of second sample, Δt = difference in days between first

and second sample. Both of these analyses were used to determine the growth and expression characteristics of the different stable pool variants.

2.3. Fixation of transfected CHO cells

Transiently transfected cells after 24 h of culture were fixed for preservation to measure intracellular fluorescence of *eGFP*, *mCherry* and *tagBFP*. This was performed by either sampling the cells into 1.5 ml tubes (Starlab, Milton Keynes, UK) or maintaining them in the 96 well flat-bottom culture plate. The samples were centrifuged at 150g for 3 min and the supernatant discarded. The cell pellets were washed by resuspending them in pre-warmed (37 °C) Dulbecco's phosphate buffered saline (DPBS) solution (Sigma-Aldrich) to a final concentration 1×10^6 cells/ml. The cells were centrifuged again, as previously described, and the supernatant discarded. The cell pellets were gently resuspended and fixed using a pre-chilled (4 °C) paraformaldehyde (PFA) fixative solution comprising of 3.7% w/v PFA and 3% w/v sucrose in PBS, pH 7.4 (Alfa Aesar, Lancashire, UK) at a final cell concentration of 1×10^7 cells/ml. The cells were incubated at 4 °C for 15 min followed by centrifugation. The supernatant was discarded, and the fixed cells were resuspended in pre-chilled (4 °C) DPBS at a final concentration of 1×10^6 cells/ml.

2.4. Fluorescent protein detection within transfected CHO cells by flow cytometry

The expression of intracellular *eGFP*, *mCherry* and *tagBFP* were quantified by flow cytometry. The methodology measures the fluorescent intensity from excitation of the fluorescent protein within living cells. During the project, two flow cytometers were used for the data generated in both chapter 3 and 4. The approach to setting up the flow cytometer was the same for both machines. Initially, the photomultiplier tube (PMT) voltages were adjusted for optimal detection of non-expressing MedI-CHO cells by adjusting the forward scatter (FSC) and side scatter (SSC) for cell size and granularity respectively. The autofluorescence of viable non-expressing MedI-CHO under the respective detection channels for *eGFP*, *mCherry* and *tagBFP* were lowered to 0 by adjusting the respective PMTs. The positive controls (expressing *eGFP*, *mCherry* and *tagBFP* as single fluorescent proteins) were measured on the machine to adjust the PMTs such that fluorescence was within range and to eliminate cross detection of a

fluorophore within other channels when coupling with compensation, for accurate fluorescent protein detection during co-expression. The detection of fluorescent proteins was measured over 10,000 events per sample after gating for viable single cells excluding doublets. The specific settings for each of the machines used in the experiments are described in the following sections below.

2.4.1. BD LSR Fortessa flow cytometer

The LSR Fortessa flow cytometer (BD Bioscience, Wokingham, UK) was used for the fluorescent data generated in chapter 3 for the single and co-expression of *eGFP*, *mCherry* and *tagBFP*. The wavelength of lasers, wavelength of filters, channel names and PMT voltages are shown in Table 2-1.

Table 2-1: Optimised settings for the BD LSR Fortessa flow cytometer.

Table summarising the lasers selected, the emission filters for detection, the associated fluorophore channels and optimised PMT voltages for the detection and quantification of *eGFP*, *mCherry* and *tagBFP*. These settings were applied for both single and multi-fluorescent protein expression.

Laser Name	Excitation Wavelength (nm)	Emission Filter (nm)	Channel Name	Fluorophore Detected	PMT Voltages
Forward Scatter	N/A	N/A	FSC-A	N/A	10
Side Scatter	N/A	N/A	SSC-A	N/A	10
Violet	405	450/50	Comp-BV421-A	<i>tagBFP</i>	143
Blue	488	525/50	Comp-FITC-A	<i>eGFP</i>	190
Green	532	610/20	Comp-PE-Texas Red-A	<i>mCherry</i>	203

The sampling of fixed transfected cells from the 96 well flat bottom culture plate after 24 h was performed using the machine's autosampler. The settings for the autosampler were as follows – sample flow rate: 2.0 $\mu\text{l}/\text{sec}$; sample volume: 150 μl ; mixing volume: 100 μl ; mixing speed: 180 $\mu\text{l}/\text{sec}$; number of mixes: 2; wash volume: 800 μl . Various compensation levels were performed depending on the inter-assay control (singular expression of *eGFP*, *mCherry* and *tagBFP* on each plate) to account for spectral overlap, an example compensation matrix is shown in Table 2-2.

Table 2-2: Example compensation matrix generated by detection on the BD LSR Fortessa flow cytometer.

Table summarising an example compensation matrix generated from detecting intracellular *eGFP*, *mCherry* and *tagBFP* when using the BD LSR Fortessa flow cytometer.

Percentage (%)		Detection		
		BV421-A	FITC-A	PE-Texas Red-A
Emission Source	BV421-A	100	0.3	0
	FITC-A	0.5	100	0
	PE-Texas Red-A	0	0	100

The data was collected using BD FACSDiva™ software (BD Biosciences) and exported as FCS files. The FCS files were then processed by FlowJo® software (FlowJo LLC, USA) to quantify median fluorescent intensity (MFI) and frequency of detection.

2.4.2. Attune NxT flow cytometer

The Attune NxT flow cytometer (Invitrogen, ThermoFisher Scientific) was used for measuring single and co-expression of *eGFP*, *mCherry* and *tagBFP* in chapter 4. The optimised settings including wavelength of lasers, wavelength of filters, channel names and PMT voltages are shown in Table 2-3.

Table 2-3: Optimised settings for the Attune NxT flow cytometer.

Table summarising the lasers selected, the emission filters for detection, the associated fluorophore channels and optimised PMT voltages for the detection and quantification of *eGFP*, *mCherry* and *tagBFP*. These settings were applied for both single and multi-fluorescent protein expression.

Laser Name	Excitation Wavelength (nm)	Emission Filter (nm)	Channel Name	Fluorophore Detected	PMT Voltages
Forward Scatter	N/A	N/A	FSC-A	N/A	405
Side Scatter	N/A	N/A	SSC-A	N/A	347
Violet	405	440/50	VL1-A	<i>tagBFP</i>	165
Blue	488	530/30	BL1-A	<i>eGFP</i>	170
Yellow	561	620/15	YL2-A	<i>mCherry</i>	230

Live transfected cells were sampled from a 96 well culture plate after 24 h using the machine's autosampler. The autosampler settings were as follows: sample flow rate: 100 µl/min; sample volume: 150 µl; sample uptake volume: 80 µl; number of mixes: 2; number of washes: 2. Once again, compensation levels varied between sample plates and were determined on the inter-assay control. An example compensation matrix is shown in Table 2-4.

Table 2-4: Example compensation matrix generated by detection on the Attune NxT flow cytometer.

Table summarising an example compensation matrix generated from detecting intracellular *eGFP*, *mCherry* and *tagBFP* when using the Attune NxT flow cytometer.

Percentage (%)		Detection		
		VL1-A	BL-1A	YL2-A
Emission Source	VL1-A	100	0	0.027
	BL1-A	0.8	100	0.0229
	YL2-A	1.8	0.1482	100

The data was collected using the Attune NxT flow cytometer software (Invitrogen, Thermo Fisher Scientific) and exported as FCS files. The FCS files were then processed by FlowJo® software (FlowJo LLC, USA) to quantify MFI and frequency of detection.

2.4.3. Gating strategy for flow cytometry quantification

The gating strategy used to quantify singular fluorescent expression is shown in Figure 2-1. This began with isolating the viable cell population as shown in Figure 2-1A, followed by single cell events and excluding the doublets as shown in Figure 2-1B, and finally excluding autofluorescence of the parental cell as shown in Figure 2-1C. Positive transfectants expressing the recombinant fluorescent protein were measured as events within the fluorescent protein gate in the dot plot as shown in Figure 2-1D.

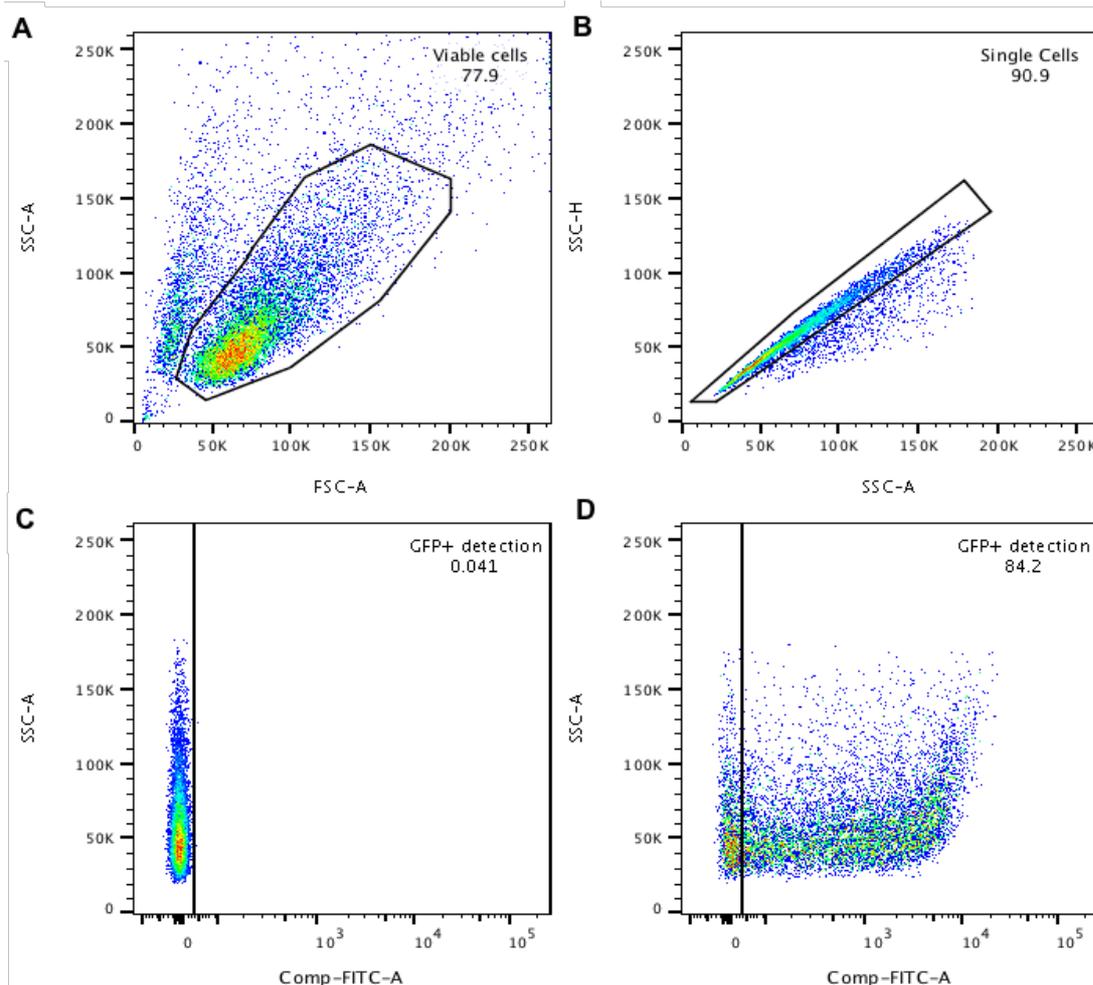


Figure 2-1: Gating strategy for singularly expressed recombinant fluorescent proteins.

Figure shows the method of gating used to isolate and analyse the transfected cell population which is expressing a single fluorescent protein. This example demonstrates identification of transfectants successfully expressing *eGFP*. The identical method was performed for transfectants successfully expressing *tagBFP* and *mCherry*. The population was identified by gating for viable cells (A), followed by gating for single cells/event and exclusion of doublets (B) and finally distinguishing between autofluorescence of the parental cell and positive detection of fluorescent protein using a mock transfected (C) and successful transfectants expressing the fluorescent protein (D).

The gating strategy for quantifying co-expression of multiple fluorescent proteins was similar to gating for singular fluorescent protein expression. The events were first gated for viable cells (Figure 2-2A) followed by single cells (Figure 2-2B). The autofluorescence of the parental cell was excluded for each fluorophore channel (Figure 2-2C). The events within the gate indicating for positive detection of *eGFP*, *tagBFP* and *mCherry* are shown in Figure 2-2D-F.

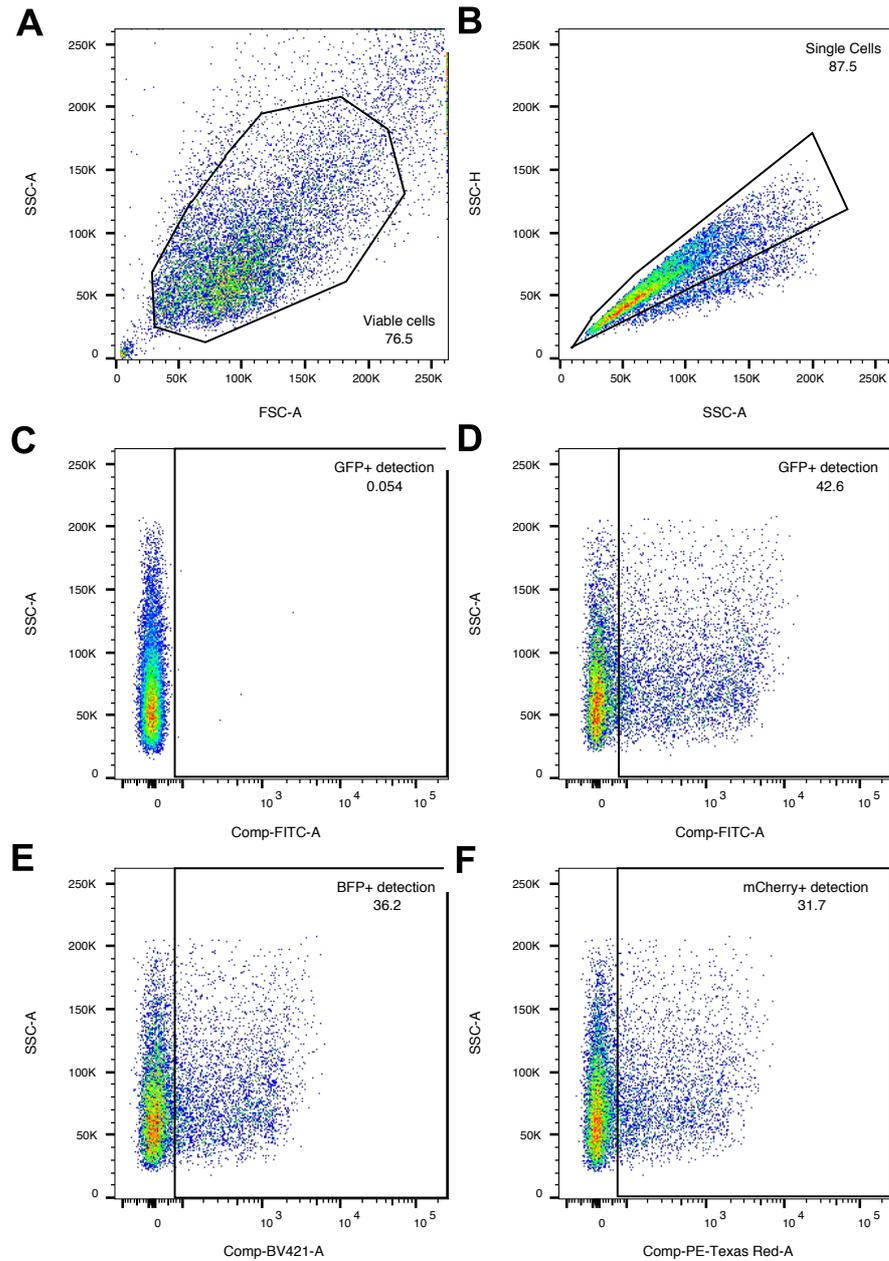


Figure 2-2: Gating strategy for expression of multiple recombinant fluorescent proteins.

Figure shows the method of gating used to isolate and analyse the transfected cell population which is expressing multiple fluorescent proteins. The population was identified by gating for viable cells (A), followed by gating for single cells/events and exclusion of doublets (B), autofluorescence of the parental cell is separated by using a mock transfected sample for the *eGFP*, *tagBFP* and *mCherry* fluorescent channels (in this figure an example for *eGFP* is shown) (C), and finally gating distinguishing co-expression of *eGFP* (D), *tagBFP* (E) and *mCherry* (F) from a MGEV.

In both the singular and multiple fluorescent protein expression, the absolute quantification was performed by measuring the median fluorescent intensity (MFI) of the positively identified events. The MFI was then multiplied by the percentage of the parent viable and single cell population. The unit of measure is called integrated MFI

(iMFI). This mode of analysis has been used in other aspects of flow cytometry based quantification as shown in literature (Darrah et al., 2007). The gating strategies used for the Attune NxT were similar and can be found in Appendix B.

2.5. Real-time quantitative PCR analysis of transfected CHO cells

Real-time quantitative PCR (qPCR) was performed on transiently transfected cells after 24 h of incubation in 96 well flat-bottom culture plates. The qPCR process involved RNA extraction of transfected cells, followed by reverse transcription of the RNA and finally quantification by qPCR. The specific methods for each step are described below:

2.5.1. Cell pellet storage and RNA extraction of transfected CHO cells

For long term storage, transfected cells were transferred from 96 well flat-bottom culture plates (Thermo Scientific, Thermo Fisher Scientific) to 1.5 ml tubes (Starlab) and centrifuged at 200g for 5 min. The supernatant was discarded, and the cell pellets were mixed with 200 μ l RNAlater stabilisation reagent (Qiagen). The cell pellets were stored at -80 °C until RNA extraction. Upon RNA extraction, the cell pellets were thawed and the RNAlater stabilisation reagent was removed by centrifugation at 5000g for 5 mins, followed by discarding the supernatant. Total RNA extraction was performed using a RNeasy mini kit (Qiagen) according to the manufacturer's protocol. The extracted RNA was quantified, and purity determined by measuring 260:230 nm and 260:280 nm absorbance ratios using the Nanodrop spectrophotometer 2000 (Thermo Scientific, Thermo Fisher Scientific).

2.5.2. Reverse transcription of extracted RNA samples

The extracted RNA was reverse transcribed using the Quantitect reverse transcription kit (Qiagen) according to manufacturer's protocol. A total of 800 ng of RNA was reverse transcribed after elimination of genomic DNA (using the gDNA wipeout buffer). Alongside the reverse transcribed RNA, there was a negative control (no reverse transcriptase added) performed as an inter-assay control during qPCR.

2.5.3. Quantitative PCR analysis

The complementary DNA (cDNA) generated from reverse transcription was diluted 3- or 10-fold with nuclease-free water (Qiagen) depending on performing a primer efficiency test or quantifying cDNA of transfected CHO cells. A qPCR reaction was set up using 2 µl cDNA, 2.5 µl primer mix (a mixture of a forward and reverse primer at a final concentration of 200 nM), 8 µl of nuclease-free water (Qiagen) and 12.5 µl of QuantiFast SYBR green PCR mastermix (Qiagen). The reactions per sample were set up in triplicate alongside two negative controls – absence of template and of reverse transcriptase in a MicroAmp® Fast Optical 96 well reaction plate with barcode (Applied Biosystems, Thermo Fisher Scientific). The cycle threshold (Ct) was quantified by using the 7500 fast real-time PCR system (Applied Biosystems, Thermo Fisher Scientific) when performing the following amplification method: 95°C for 5 min, followed by 40 cycles at 95 °C for 15 sec and 60 °C for 1 min. Furthermore, primer efficiency melting curve analysis was performed from 60 to 95 °C.

2.6. Quantification of secreted alkaline phosphatase by absorbance

SEAP was quantified from clarified supernatant collected after 72 h of culturing. The clarified supernatant was obtained by centrifugation of cultures at 200g for 5 min. The quantification was performed using the SensoLyte® pNPP SEAP Reporter Gene Assay Kit Colourimetric (AnaSpec Inc., California, USA). A 2-fold serial dilution standard curve was prepared using the kit standard (human placental alkaline phosphatase standard) and 1x assay buffer (diluted from 10x assay buffer with milliQ water) from 15.63 to 0.24 ng/ml. The supernatant samples were diluted 5-fold with the 1 x assay buffer. A 50 µl aliquot of diluted standard and supernatant was transferred into a clear 96 well flat bottom plate (Starlab). A negative control of 1x assay buffer was included on the assay plate. A 50 µl aliquot of the pNPP substrate was added to each well and incubated in the dark for 15 mins and mixed by orbital shaking for 30 sec. The absorbance was measured at 405 nm using the SpectraMax iD5 plate reader (Molecular Devices, Wokingham, UK). The raw absorbance values were adjusted for background signal by subtracting the negative control.

2.7. Quantification of *Cypridina* luciferase by luminescence

The clarified supernatant containing the *Cypridina* luciferase (*CL*) was collected by centrifugation of the culture at 200g for 5 min after 72 h of culturing. The *CL* was quantified using Pierce® *Cypridina* Luciferase Glow Assay kit (Thermo Scientific, Thermo Fisher Scientific). The clarified supernatant was diluted 100-fold with 1 x PBS (Gibco™, Thermo Fisher Scientific). A 20 µl aliquot of diluted clarified supernatant per sample and 20 µl of PBS (negative control) was transferred to a white 96 well flat bottom plate (Corning). A working solution was prepared by diluting the 100x vargulin (substrate) with the *Cypridina* Glow Assay Buffer, and 70 µl of the working solution was mixed with the samples in the white 96 well flat bottom plate. The plate was then included in the dark for 10 min and mixed by orbital shaking for 30 sec. The luminescence was measured using the SpectraMax iD5 plate reader. The raw luminescence values were adjusted for background signal by subtracting the negative control (PBS).

Chapter 3

3. Design and Characterisation of a Multi-Gene Expression Vector for CHO Cell Engineering

This chapter describes the design, verification and characterisation of a multi-gene expression vector (MGEV) specifically for recombinant gene expression in CHO cells. An extensive description is provided around the design space and desired criteria for a reproducible, functional and simple MGEV applicable in an industrial and academic environment. An in silico design process was performed to outline the genetic components required and an exemplar construction flowpath using Golden Gate assembly. The designed MGEV encoding the co-expression of eGFP, mCherry and tagBFP was tested in vitro from construction to transient expression. Following development of an optimised quantification assay using flow cytometry, the gene expression dynamics were characterised within the polycistronic cassette. The data yielded a positional effect with an average ratio of 1.00: 0.71: 0.77 and was reinforced by the population distribution analysis. It was hypothesised that the relative repression of downstream recombinant gene expression may be a by-product of transcriptional interference.

3.1. Introduction

As described in Section 1.6, MGEVs are DNA plasmids comprising of multiple recombinant genes assembled together. MGEVs have potential applications in numerous aspects of the biological sciences including gene therapy, biofuel production, cellular reprogramming and engineering, and biopharmaceutical production. For example, *S.cerevisiae* has been metabolically engineered to introduce a more effective biosynthesis pathway (found in *E.coli*) for the production of farnesyl pyrophosphate, an artemisinin (anti-malarial) precursor. As a result, this process has become more cost effective (Khalil and Collins, 2010).

Similar approaches using MGEVs could be applied to improving biopharmaceutical production. For example, potential targets have been identified such as the co-expression of *XBP-1* (Pybus et al., 2014) and *CypB* (Johari et al., 2015) alongside a recombinant difficult-to-express (DTE) monoclonal antibody (mAb) to enhance expression levels. Furthermore, introducing new phenotypic features such as α 2,6-sialylated glycans, a new glycosylation motif, (Fischer et al., 2015) could facilitate production of improved biopharmaceuticals. Currently, however, there are limited capabilities of achieving these levels of complex multi-gene engineering in mammalian cells.

The aim of developing MGEVs was to generate constructs which are robustly and effectively assembled and use a standardised template, whereby each expression cassette was uniform in layout and structure. The genetic parts which were involved in the formation of a MGEV were the recipient backbone vector and transcription units (TUs) which represent the recombinant genes. The first attempt of standardised DNA component assembly was using nucleic acid ordered assembly with directionality (NOMAD) that allowed for modular sequential cloning using the restriction enzyme *StyI* (Rebatchouk et al., 1996). DNA engineering and cloning has continued to progress and has facilitated a number of different approaches of assembling MGEVs and testing functionality in mammalian cells leading to various publications. For example, Kriz et al demonstrated an approach to assemble 8 fragments to form a MGEV using Cre/LoxP recombination. The method used a combination of acceptor and donor vectors to assemble 5 fluorescent protein genes into a single plasmid and was called MultiLabel. The MultiLabel system was shown to be functional both

transiently and stably in HEK293 and COS cells respectively (Kriz et al., 2010). An alternative method was published by Guye et al, whereby a two tier assembly system was developed for MGEV construction. The first tier used gateway cloning to assemble the TUs comprising of the promoter, recombinant gene and a destination vector. The second tier used Gibson assembly (as described in Figure 1-3) to assemble multiple TUs to form a MGEV. The assembly of the TU and MGEV was pre-designed and the order of the genes within the MGEV were designated by allocating specifically designed destination vectors prior to *in vitro* assembly. The approach demonstrated the assembly of a 7-TU MGEV which was 39 kb and was tested in HEK 293 in both a transient and stable platform by comparing expression of three fluorescent proteins using a combination of constitutive and antibiotic-inducing promoters (Guye et al., 2013).

Similarly, Weber et al used a tier system to assemble the TU and MGEV using Golden Gate assembly. Tier 1 was the cloning of individual genetic parts (e.g. promoters, untranslated regions (UTRs), recombinant genes, transcription terminators) into a destination vector. Tier 2 was the combination of different genetic parts together to form a TU. The TUs were cloned into unique destination vectors which facilitated directional cloning of multiple-TUs as the final tier. This system was called MoClo cloning and was shown to successfully assemble 11 TUs (Weber et al., 2011). Duportet et al used the MoClo cloning method to test a 7-TU MGEV in HEK 293 both for transient expression and targeted integrated driven stable expression by measuring fluorescent protein expression (Duportet et al., 2014).

The three examples describe approaches which were previously published, which successfully construct a functional MGEV. However, they employ varying methods which increased the complexity of assembling MGEVs either by using more than one cloning technique, requiring a large number of acceptor and/or donor vectors or using multiple restriction sites and linker systems. Furthermore, current literature has shown the design and functionality of MGEVs in human-derived cell lines such as HEK293 and COS cells. Limited data has been shown for the specific design and application of MGEVs for CHO cells and specifically for engineering relating to biopharmaceutical production.

This chapter explores the design and characterisation of a MGEV specifically designed for CHO cell engineering and biopharmaceutical production. The process began with identifying the optimal performance attributes for the MGEV system, the cloning strategy and a means to test the MGEV system. This was followed by designing linkers to control the order of DNA fragment cloning when employing Golden Gate assembly. Additionally, an *in silico* design of a model system was constructed highlighting all the different genetic parts required to assemble a MGEV in an approach to simulate MGEV process. Finally, the design was tested *in vitro* to investigate the efficiency of cloning and characterise the expression dynamics of multiple recombinant genes within a MGEV by using three fluorescent protein reporters.

3.2. Review of Design Attributes for a multi-gene expression vector system

To design a MGEV system specific for CHO cell engineering related to biopharmaceutical production, a list of desired performance attributes was identified. This was achieved by a combination of reviewing literature, collaborating with MedImmune and considering the current limitations within CHO cell engineering in an academic and industrial environment. This in turn provided the design space required to make a functional MGEV. The attributes identified for the MGEV were as follows:

- a) *Robust and accurate expression system* – is essential for biopharmaceutical production in industry and cell engineering in academia. This would allow for improved control of gene expression and consequently generate predictable recombinant cell lines. Hence, facilitating high fidelity and accuracy of CHO cell engineering (Brown et al., 2014). As a result, minimising product heterogeneity during production and simplifying the approval process by drug regulatory agencies (Li et al., 2010).
- b) *Simple cloning methodology* – would lead to a more efficient, less time-consuming and simpler process to successfully assemble a MGEV. This is ideal as it mitigated risk of error and does not require extensive molecular biology experience to generate MGEVs. Furthermore, a simple protocol allows for better transferability in an academic environment.
- c) *License-free* – The development of a new vector system to be applied commercially in industry, such as production of biopharmaceuticals, could have a risk of paying for licensing and royalties. Therefore, it is advantageous

to design a MGEV which would be license-free from any proprietary reagents and technology.

- d) *Functional in a transient and stable platform* – Inclusion of appropriate genetic elements in the MGEV to allow for transferability between transient and stable recombinant expression. Although, a unique mammalian episomal replication element is required to maintain DNA copies within the host cell for an extended period of transient expression, it is not required in stable expression. Accordingly, designing an efficient interchanging system to insert or excise either the origin of replication element or the selection marker gene allows for seamless functionality in both platforms.
- e) *Rapid assembly of multiple recombinant genes* – Identifying a cloning technique that required a short duration for efficient assembly of multiple recombinant genes in a MGEV. This is ideal for saving time during vector construction compared to familiar multi-step approaches. Furthermore, a rapid method is advantageous towards generating large number of MGEVs for high throughput (HT) screening of accessory genes to potentially engineer the CHO cell factory.
- f) *Cost effective* – The design of the MGEV must be cost effective. Additionally, this is especially important if large number of constructs were going to be generated for HT screening in both academia and industry. Therefore, minimising cost of reagent costs is ideal. Generally, *de novo* synthesis is costly, therefore approaches within design to minimise cost would be beneficial (refer to Appendix A for cost analysis for synthesis by various companies in 2016).
- g) *Engineering flexibility and future proofing* – is beneficial for research applications. Modifications could therefore be easily performed on the MGEV and incur minimal costs. An approach to achieve this is designing a ‘plug-and-play’ system within a standardised TU scaffold. This allows for the interchangeability of different genetic elements such as promoters, untranslated regions (UTRs), and coding DNA sequences (CDSes). The following approach led to future proofing the system by allowing the addition of newly designed genetic parts such as synthetic promoters or transcription terminators.

As mentioned in Section 1.6.3, Golden Gate assembly was deemed the favourable methodology. The method uses type II restriction endonucleases which recognise a non-palindromic restriction site and digest the phosphodiester bond outside of the recognition site. The unique feature of the digest is that it forms a 4 bp overhang on the 5' end of the sense and anti-sense strand as shown in Figure 3-1 (Engler et al., 2008).

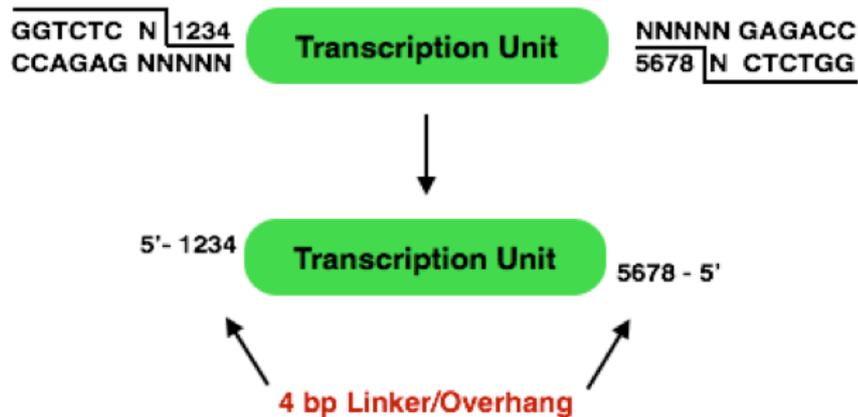


Figure 3-1: Golden gate assembly reaction.

Demonstrating the digestion reaction of type II restriction enzymes, specifically BsaI. The figure also shows how the orientation of the restriction site can be used to generate 5' 4 bp overhangs on either end of the TU and on opposing DNA strands.

Figure 3-1 also shows how changing the orientation of the restriction site can lead to flanking 4 bp overhangs on either side of the TU. A reverse orientation of the restriction site was designed for the recipient vector such that the overhangs were complementary to the TU. Furthermore, the flanking overhangs could be designed *in silico* to be unique and complementary between multiple TUs and the recipient vector. This allowed for control of TU assembly order and orientation when constructing a MGEV (Engler et al., 2009).

Golden Gate assembly is a unique cloning technique since the restriction digestion and ligation reactions can occur concurrently in a single tube. The catalysed digestion occurs outside of the recognition site and therefore, the recognition site is excised, leaving the complementary linkers to ligate together. As opposed to traditional digestion-ligation cloning, this method is significantly quicker and ideal for rapid and effective cloning. Furthermore, the linkers between TUs are significantly shorter compared to alternative cloning techniques (Gibson assembly and Gateway cloning) leading to shorter scar sequences and minimising the length of a MGEV. Although,

Gibson assembly can achieve scarless assembly, the linkers would have to be re-designed for each TU dependent on the genetic parts within it and hence introduce complexity compared to a standardised 4 bp linker flanking each TU.

The requirement for an accurate method of detecting expression of multiple genes was essential to effectively characterise the MGEV. After assessing various approaches, it was decided that co-expressing three fluorescent proteins and using a flow cytometer would be an effective means of quantifying recombinant gene expression in a MGEV. As shown in literature, fluorescent proteins have frequently been used as reporter genes to measure recombinant expression within a MGEV (Duportet et al., 2014; Kriz et al., 2010). This specific class of proteins are also non-complex since they do not require additional enzymes or cofactors to fold into a matured fluorophore (D. Craggs, 2009). Furthermore, they exhibit low toxicity and stress on the host cell during recombinant expression (Coralli et al., 2001; Viotti et al., 2011). The fluorescent proteins to be expressed were selected based on the following criteria – i) the fluorophores were required to be excited by different wavelength lasers; ii) the emission spectra were mutually exclusive and have minimal overlap to avoid compensation issues; iii) fluorophores frequently used as reporter genes for recombinant expression; iv) fluorophores which were compatible and detectable by the flow cytometer which was available. The outcome indicated that *eGFP*, *mCherry* and *tagBFP* were the most appropriate for mutually exclusive co-expression as shown in Figure 3-2.

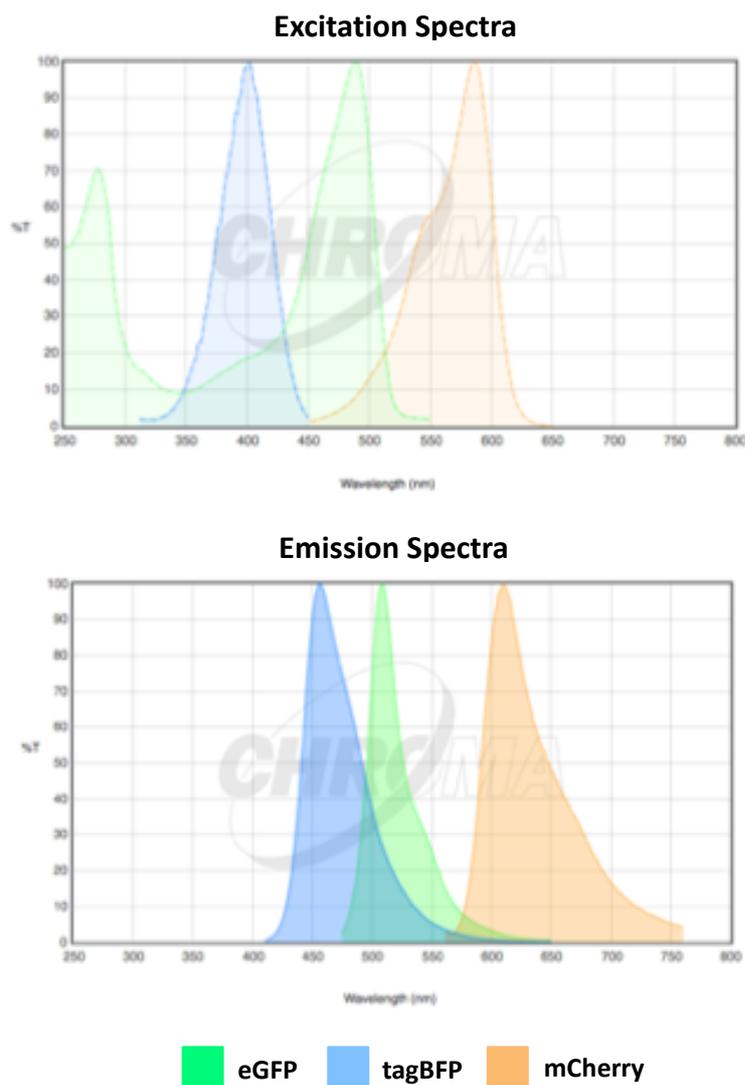


Figure 3-2: Excitation and emission spectra for *eGFP*, *mCherry* and *tagBFP*.

The spectra shows the excitation and emission profiles for *eGFP*, *mCherry* and *tagBFP*. The excitation graph showed the percentage of threshold of excitation expected when exposed to a particular wavelength laser. The emissions graph showed the percentage of threshold emitted when detected by a particular wavelength filter. These spectra assisted in optimising the flow cytometer settings to achieve optimal detection using the equipment available. The chromatograms were generated on spectra viewer within the Chroma Technology Group website (<https://www.chroma.com/spectra-viewer>).

As a result of identifying the optimal design attributes, cloning methodology and means to test the MGEV system, the next step was to *in silico* design and *in vitro* test the new system.

3.3. Results

3.3.1. Design of a multi-gene expression vector

This section describes the various design outputs for the MGEV system including the linkers for Golden Gate assembly, layout and composition of both the TU and vector backbone and the simulation of a MGEV construction process.

3.3.1.1. Designing 4 bp linkers for controlling gene assembly

A total of 256 unique 4 bp linkers can be used to assemble DNA fragments or TUs together using Golden Gate assembly. However, two criteria were included in the design of the linkers; a minimum of a 2 bp difference between linkers and/or different order of the 4 bp linkers, which led to a reduction in the homology between linkers as a precaution towards avoiding mismatch or incorrect assembly during Golden Gate-mediated construction (Engler and Marillonnet, 2011; Engler et al., 2009). Table 3-1 shows a series of linkers designed for a potential 8 TU MGEV, however, more complementary linkers can be designed to expand the number of TUs within the MGEV if required.

Table 3-1: Complementary 4 bp linkers for multi-gene expression vector cloning.

In silico designed 4 bp linkers and their respective complementary sequences that can be used to assemble 8 TUs to a plasmid backbone to form a MGEV.

Position of Linker	5' Linker	Complementary Linker
1	5'-GGAG-3'	3'-CCTC-5'
2	5'-TACT-3'	3'-ATGA-5'
3	5'-CCAT-3'	3'-GGTA-5'
4	5'-AATG-3'	3'-TTAC-5'
5	5'-AGGT-3'	3'-TCCA-5'
6	5'-TTCG-3'	3'-AAGC-5'
7	5'-GCTT-3'	3'-CGAA-5'
8	5'-GGTA-3'	3'-CCAT-5'
End	5'-CGCT-3'	3'-GCGA-5'

Furthermore, Figure 3-3 demonstrates how the linkers (as described in Table 3-2) can control the order and orientation of three TUs being cloned within a MGEV when using Golden Gate assembly.



Figure 3-3: Control of transcription unit assembly during multi-gene expression vector construction.

In silico pre-designed 4 bp overhangs can be used as complementary linkers between multiple TUs to control the order and orientation of the genes during MGEV construction.

3.3.1.2. *In silico* design of genetic components for assembly of a multi-gene expression vector

The *in silico* design of the MGEV was divided into two parts. One part was to design a standardised TU which allows for robustness, reproducibility, simplicity and engineering flexibility. The other part was to design a vector backbone that provides functionality in a transient and stable platform, and would be transferable towards existing molecular biology protocols established in academia and industry. This section will expand further on the designed TU and vector backbone.

3.3.1.2.1. Design of the transcription unit

The first step to designing a standardised TU for recombinant gene expression was identifying the genetic elements required. The essential elements required were the promoter, UTRs and CDS for successful expression. However, it was anticipated that a library of synthetic promoters would be employed to control transcription and titrate gene expression. Therefore, a feature of interchanging the proximal transcription factor regulatory elements (TFREs) was desirable while the core promoter remains intact. Furthermore, it was hypothesised that a genetic element such as a transcription terminator or insulator would be required to achieve predictable expression and control of multiple recombinant genes as individual modules (Liao et al., 2018; Proudfoot, 2016). Whereby, each TU would function independently and dictated by the transcriptional control of the synthetic promoter to allow for predictable stoichiometric expression of each recombinant gene (Torella et al., 2014b). As a result, the final list of genetic parts included a synthetic/viral proximal and core element of a promoter, 5' UTR (including Kozak sequence), CDS (including signal sequence peptide), polyadenylation (pA) element and a terminator/insulator element.

The next step was to decide how the TU would be constructed either by *de novo* synthesis or *in vitro* assembly using oligonucleotide fragments. *In vitro* construction of the TU would increase the scope of interchanging different elements to rapidly generate multiple MGEV variants for various engineering requirements. However, this would increase the complexity of cloning by either using two different type II restriction endonucleases or using two different cloning techniques, as shown by Weber et al and Guye et al respectively (Guye et al., 2013; Weber et al., 2011). Simplicity was an essential feature of the MGEV, hence *in silico* design and *de novo* synthesis was the preferred option.

Unfortunately, assessing the cost of *de novo* synthesis by third party gene synthesis companies highlighted that synthesis of large number of TUs for HT screening and engineering of CHO cells would be time consuming and relatively costly, even though the more cost effective design was selected (refer to Appendix A for example costs of synthesis in 2016 by different gene synthesis companies). Therefore, the TU required some form of engineering flexibility, such that repeated synthesis was not required. The solution was to design a ‘plug-and-play’ feature within the TU to facilitate post-synthesis modification. This was achieved by first highlighting that the synthetic/viral proximal element, the CDS, pA element and terminator/insulator element would most likely require interchanging for tunability of expression. Traditional restriction digestion-ligation cloning was used to facilitate the ‘plug-and-play’ feature due to its familiarity within industry and academia. However, there are a finite number of co-functional and compatible type II restriction endonucleases with unique recognition sites to be applied within a multi-TU MGEV. Ergo, a set of standard restriction sites was strategically positioned within a TU, as shown in Figure 3-4. These sites were curated to be compatible with each other. Additionally, each sticky end is unique to avoid any mismatch during ligation. As shown in Figure 3-4, the AgeI restriction site was positioned upstream of the Kozak sequence, hence separating it from the remainder of the 5' UTR element. This is because the Kozak sequence was required to be directly upstream of the CDS start codon to maintain translation initiation functionality. Therefore, for the purpose of this TU design, the Kozak sequence was not included within the 5' UTR element but associated as part of the CDS.

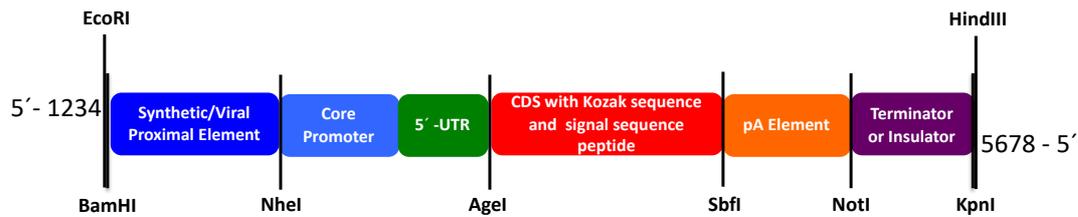


Figure 3-4: Standardised transcription unit layout.

The figure displays a standardised layout for a recombinant gene called a TU. It comprises of a synthetic/viral proximal element, core promoter, 5' UTR, CDS with a Kozak sequence and signal sequence peptide, pA element and a transcription terminator or insulator element. The restriction enzyme sites flanking each element allow for iterative modification and swapping of genetic components using restriction digestion-ligation cloning. Although the Kozak sequence is associated as a 5' UTR, the location of it must be directly upstream of the start codon, therefore the only viable location of a restriction enzyme site was upstream of the Kozak sequence, Therefore, for the purpose of this TU design, the Kozak sequence is associated with the CDS rather than traditionally labelled as a 5' UTR.

3.3.1.2.2. Design of recipient vector backbone

The recipient vector backbone was designed *in silico* with a set of specific design features for functionality in both industry and academia. The elements included within the vector backbone were a Golden Gate assembly site flanked with a BsaI restriction site, pBR322_origin (microbial origin of replication), an ampicillin resistance marker, an episomal origin of replication and a glutamine synthetase (*GS*) gene cassette, as shown in Figure 3-5.

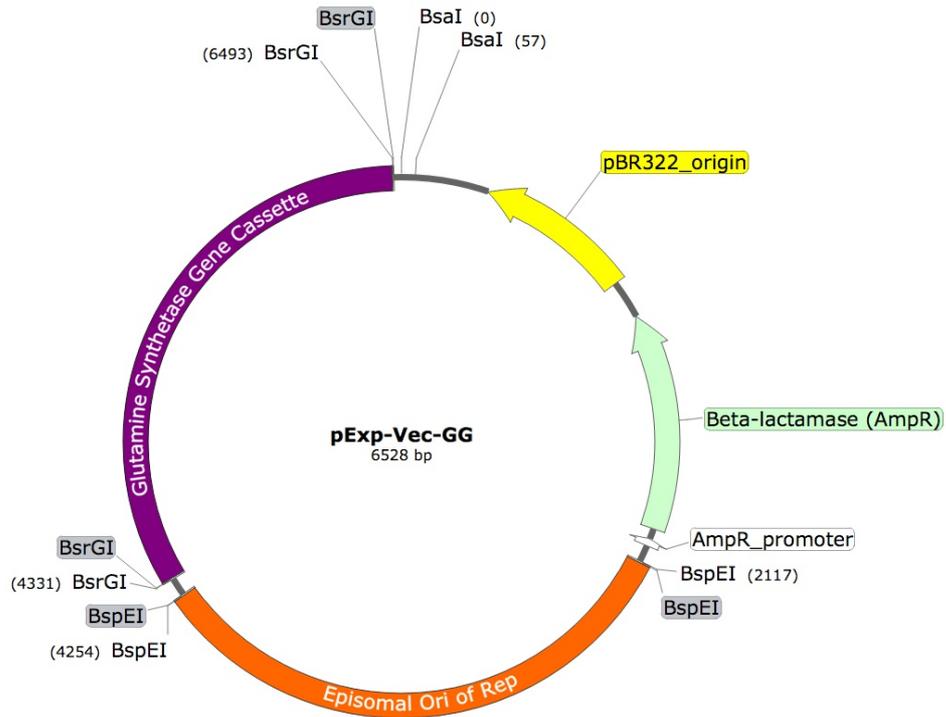


Figure 3-5: pExp-Vec-GG recipient vector backbone for multi-gene expression vector assembly. Vector map showing the lay out of the *in silico* designed backbone vector for MGEV assembly. The vector includes a Golden Gate assembly site flanked by two BsaI restriction sites, an origin of replication for microbial amplification (pBR322_origin), an ampicillin resistance marker (β -lactamase gene), episomal origin of replication for mammalian cells and glutamine synthetase (*GS*) gene cassette.

The recipient vector backbone provided the MGEV with functionality within a transient and stable platform by the presence of the episomal origin of replication element and the *GS* gene cassette respectively. The episomal origin of replication was essential for long-term transient expression and was compatible with MedImmune’s proprietary CHO transient host cell line. Whereas, the *GS* gene cassette was used as a selection marker for generating a stable cell line and was functional in MedImmune’s Medi-CHO host cell line. Both of these elements were flanked with a unique type II restriction endonuclease site, BsrGI and BspEI, to excise either the *GS* gene cassette or episomal origin of replication element respectively depending on the desired function of the MGEV.

The Golden Gate assembly site had two pre-designated linkers annotated as linker 1 (5'-GGAG-3') and linker E (5'-CGCT-3') which were constant. Therefore, the 5' of the first TU and the 3' of the final TU had complementary linkers to clone the polycistronic cassette into the recipient vector. The recipient vector also had a

microbial origin of replication to allow for amplification within transformed *E.coli*. Once coupled with the β -lactamase gene to provide resistance against ampicillin, this allowed for standard DNA plasmid amplification processes. The *GS* gene cassette, episomal origin of replication, microbial origin of replication and β -lactamase gene were all sourced from MedImmune's current vector system. This meant that all the elements were already performance tested and proven to be robust and functional for CHO-mediated recombinant expression both stably and transiently. Furthermore, the vector backbone would align with MedImmune's internal processes, easing the transition towards commercial applications and introducing an upgraded vector system for their industrial platform.

3.3.1.3. Overview of cloning and modification of a multi-gene expression vector

The construction of the MGEV using Golden Gate assembly and the DNA components described above involves a three-step process.

1. *In silico design* – Each TU comprised of six elements – a synthetic/viral proximal element, a core promoter, 5' UTR, CDS, a pA element and a terminator/insulator element. Each element was selected from MedImmune's current vector system (e.g. hCMV-MIE proximal and core promoter elements, 5' UTR and SV40 pA) which have been thoroughly performance tested in a commercial environment. The CDSes were codon optimised for CHO cell expression to perform efficient translation. The sequence of the TU was assembled along with the order and orientation of the TUs within the MGEV. This was achieved by allocating complementary 4 bp linkers embedded in the BsaI restriction site on the flanks of each TU. The finalised TU sequence was then *de novo* synthesised using a third-party gene synthesis provider.
2. *MGEV assembly* – used Golden Gate assembly (as described in Chapter 2) to construct a polycistronic cassette comprising of multiple TUs and cloning into the recipient backbone vector using the complementary linkers. The method utilised NEB® Golden Gate assembly mix which comprised of an optimised mixture of BsaI enzyme and T4 DNA ligase, as well as an optimised buffer to facilitate the enzymatic reactions. Recipient vector (75 ng) was mixed with 2:1 molar ratio of the synthesised TUs along with the buffer and enzymes. For up

to 4 fragments, the mixture was incubated at 37 °C for 1 h followed by 5 min at 55 °C in a thermocycler. For 5 to 10 fragments, the incubation was cycled between 37 °C and 16 °C for 1 min at each temperature for 30 cycles followed by 5 min at 55 °C. For 11 fragments or more, the incubation method was the same apart from the incubation at 37 °C was extended to 5 min.

3. *Interchanging genetic elements* – The genetic elements which were interchangeable through the ‘plug-and-play’ feature were the synthetic/viral proximal element, CDS, pA element and terminator/insulator element. Each of these elements was flanked with a pair of unique type II restriction endonuclease sites which are co-functional. Therefore, these elements could be excised and replaced with alternative fragments flanked by the same enzyme pair by conventional restriction digestion-ligation cloning. The modification of the TUs were performed as singular units prior to MGEV construction.

3.3.2. Testing the cloning technique for assembling multi-gene expression vectors

To comprehensively test the Golden Gate assembly method and characterise the functionality of the *in silico* designed MGEV system, a three TU system was designed and synthesised to co-express three fluorescent proteins. As described in section 3.3.1.3, the three TUs were assembled computationally and comprised of a human cytomegalovirus-major intermediate early (hCMV-MIE) promoter, 5' UTR and simian virus 40 (SV40) pA element and one of three fluorescent protein CDSes (*eGFP*, *mCherry* and *tagBFP*) as presented in Figure 3-6.

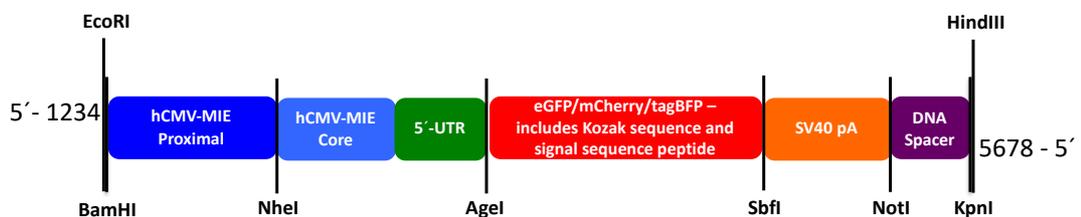


Figure 3-6: Genetic elements and layout of transcription unit encoding for fluorescent proteins. The schematic depicts the genetic elements selected and the layout of the TU encoding for three fluorescent proteins – *eGFP*, *mCherry* and *tagBFP*. Therefore, three different TUs (for each fluorescent protein) were designed and *de novo* synthesised using the same human cytomegalovirus-major intermediate early (hCMV-MIE) proximal and core element, a MedImmune proprietary 5' UTR, simian virus (SV) 40 polyadenylation (pA) element and an inactive DNA spacer.

The TU order within a polycistronic cassette was assigned, and in turn a pair of complementary linkers were designated to control the order of cloning, as shown in Figure 3-7.



Figure 3-7: Assembly order of three transcription units for multi-gene expression vector testing. Figure showing the *in silico* designed order of assembly of three TUs expressing the fluorescent proteins *eGFP*, *tagBFP* and *mCherry* and the respective complementary linkers used to achieve the cloning.

The *in silico* designed TUs and recipient vector were synthesised using GeneArt's gene synthesis service after assessing the cost of *de novo* synthesis services provided by various different suppliers (refer to Appendix A). Upon receipt of the TUs and pExp-Vec-GG vector (Figure 3-5), the genetic components were tested *in vitro* to investigate the successful assembly of a MGEV where TUs were assembled in the order described in Figure 3-7. This was performed using NEB's Golden Gate assembly kit as described in Chapter 2. Purified DNA of 12 selected colonies were screened using AgeI-HF enzyme followed by gel electrophoresis-mediated visual inspection. The result of the restriction digest colony screen can be seen in Figure 3-8A.

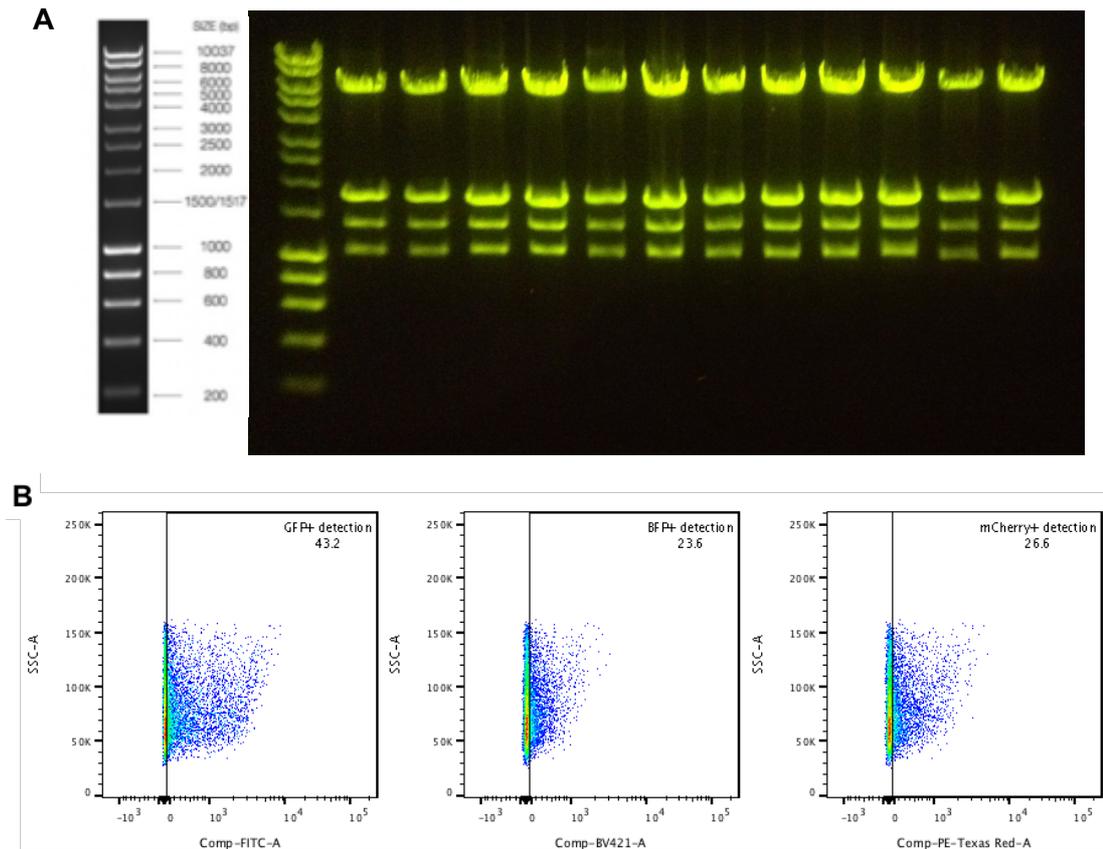


Figure 3-8: Successful assembly and expression of three transcription units and pExp-Vec-GG backbone plasmid.

A- 1% agarose gel image showing digested DNA from 12 positive transformed colonies with the MGEV comprising of three TUs encoding for *eGFP*, *tagBFP* and *mCherry* and the pExp-Vec-GG backbone vector. DNA was restriction digested using AgeI-HF and determined by size of the DNA fragments. **B-** Dot plots from flow cytometry detection showing positive detection of *tagBFP*, *eGFP* and *mCherry* at 405 nm (violet), 488 nm (blue) and 532 nm (green) wavelengths. Therefore, an orthogonal method to verify successful assembly and functionality of the MGEV cloning method. The MGEV tested was constructed with the TUs in the following order – TU1-eGFP, TU2-tagBFP and TU3-mCherry and referred to as pMGEV-GBC.

The agarose gel image in Figure 3-8A showed DNA bands fluorescently stained and visualised under UV light. The expected bands of a correctly formed MGEV would be at 5844, 1716, 1698, 1314 and 1020 bp when digested with AgeI-HF. All the bands observed in the gel were the right size when compared against the DNA ladder. However, only one band was observed at around 1700 bp which was due to insufficient separation between the 1716 and 1698 bp fragments. This was an expected limitation from using a 1% agarose gel. However, the increased intensity of the band at ~1700 bp would suggest two overlapping fragments of DNA is present. Therefore, these data suggested all 12 colonies contained the correctly cloned MGEV and indicated that the Golden Gate assembly protocol was an effective cloning technique. However, to verify that the TUs had assembled in the correct order, the plasmid was sequenced. One of

the successfully screened clones was sequenced to show correct order of assembly as dictated by the pre-designed linkers (the sequence alignment for pMGEV-GBC can be found in Appendix A).

An alternative means of verifying the successful assembly and functionality of the *in silico* designed MGEV was by detecting fluorescent protein expression using a flow cytometer. This was achieved by transiently transfecting the MGEV in the MedI-CHO host cell using the HT transfection protocol (described in Chapter 2). The expression of the fluorescent proteins was detected after 24 h of culturing. This was achieved by using the 405 nm (violet), 488 nm (blue) and 532 nm (green) lasers to excite any intracellularly expressed *tagBFP*, *eGFP* and *mCherry* respectively. However, these settings were unoptimised and the result was used as a ‘yes-no’ detection system and not as a quantification. As shown in Figure 3-8B, positive events were detected within the established gates (based on mock transfected cells) for each of the dot plots. 43.2%, 23.6% and 26.6% of the population had successfully expressed *eGFP*, *tagBFP* and *mCherry* respectively. Interestingly, a difference in frequency of positive fluorescent expression detected was observed for each fluorescent protein even though all three genes were driven by the same promoter hence identical transcriptional power. A hypothesis for the variation could be caused by the order of the TUs in the MGEV leading to a positional effect where repression of downstream gene expression (*tagBFP* and *mCherry*) may be occurring.

3.3.3. Developing an optimised detection method to quantify expression of three fluorescent proteins

As described in Section 3.2, *eGFP*, *mCherry* and *tagBFP* were selected as fluorescent proteins that could be suitably co-expressed and quantified by the flow cytometer available. However, to achieve accurate quantifiable data to characterise the protein expression dynamics within the MGEV, the detection method had to be optimised as described below.

3.3.3.1. Optimising the parameters of the flow cytometer for fluorescent protein expression

The flow cytometer (LSR Fortessa) and the chosen fluorophores (*eGFP*, *mCherry* and *tagBFP*) were tested to investigate if this approach would be a reliable, sensitive and accurate form of quantification. This was achieved by transiently expressing the fluorescent proteins as single gene vectors (SGVs) and as a MGEV in the MedI-CHO host cell line for 24 h using the HT transfection method. The first step in optimising the detection assay was to establish the optimal voltages and gating strategy to accurately quantify all three fluorescent proteins as described in Chapter 2.

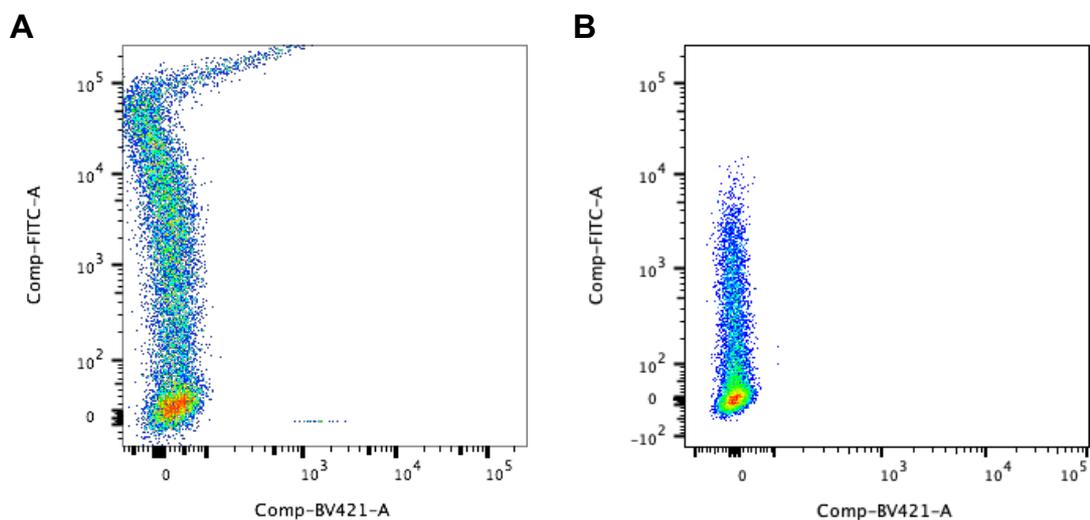


Figure 3-9: Removing false positive *tagBFP* signal during *eGFP* detection.

Dot plots portraying successful transient expression of *eGFP* within MedI-CHO cells after 24h. **(A)** Plot shows high *eGFP* detection with fluorescence intensity $>10^5$ also leads to detection within the 450/50 nm filter as shown by the right-hand kink within the plot. The 450/50 nm filter is employed for *tagBFP* detection; hence the kink represents false detection of *tagBFP*. **(B)** The lowering of photomultiplier tube voltages of the 405 and 488 nm laser to reduce fluorescence intensity to $<10^4$ and applying compensation removed the kink. Therefore, eliminating the false detection of *tagBFP* during *eGFP* transient expression. In both plots Comp-BV421-A and Comp-FITC-A represent fluorescence intensity within the 450/50 nm and 525/50 nm filter for *tagBFP* and *eGFP* detection respectively.

Whilst developing the quantification strategy for both singular and multiple fluorescent protein expression, an unexpected observation of the *eGFP* fluorophore was made. As the fluorescence intensity of *eGFP* was higher ($>10^4$ units), it led to detection in the 450/50 nm filter, as shown in Figure 3-9A. This led to a false positive detection of *tagBFP*. Confirmation of this was done by only expressing the *eGFP* in MedI-CHO. Initial approaches to use compensation between the two channels did not sufficiently remove the false positive signal (data not shown). After reducing the photomultiplier tube (PMT) voltages of both the 405 nm and 488 nm lasers and

coupled with compensation, the false detection of *tagBFP* fluorescence was removed (Figure 3-9B).

3.3.3.2. Assessing the dynamic range of fluorescent protein quantification by flow cytometry

In order to determine the sensitivity and accuracy of using a flow cytometer to quantify fluorescent protein expression, the dynamic range of detection needed to be defined. This was achieved by varying the gene dose of each fluorescent protein (*eGFP*, *mCherry* and *tagBFP*) when regulated by a hCMV-MIE promoter and expressed transiently. Plasmid DNA (100 to 1000 ng) encoding for either *eGFP*, *mCherry* or *tagBFP* was transfected in MedI-CHO using the HT transient transfection protocol and cultured for 24 h. The fluorescence for *eGFP*, *mCherry* and *tagBFP* within the transfected cells were then measured using the optimised flow cytometer settings as shown in Table 2-1 and gating strategy described in Figure 2-1. The iMFI was calculated and plotted against the quantity of DNA transfected and shown in Figure 3-10.

For each of the fluorescent protein curves of titrated DNA against iMFI, a linear range of detection was identified between 400 to 1000 ng as shown in Figure 3-10, where the r^2 for *eGFP*, *mCherry* and *tagBFP* was 0.9839, 0.9598 and 0.9866 respectively. Conversely, minimal detection was observed below 300 ng. This indicated that the dynamic range for accurate quantification of recombinant fluorescent protein expression was between 400 to 1000 ng. Therefore, transfecting DNA at 600 ng would allow for observing both increases and decreases of recombinant fluorescent protein expression.

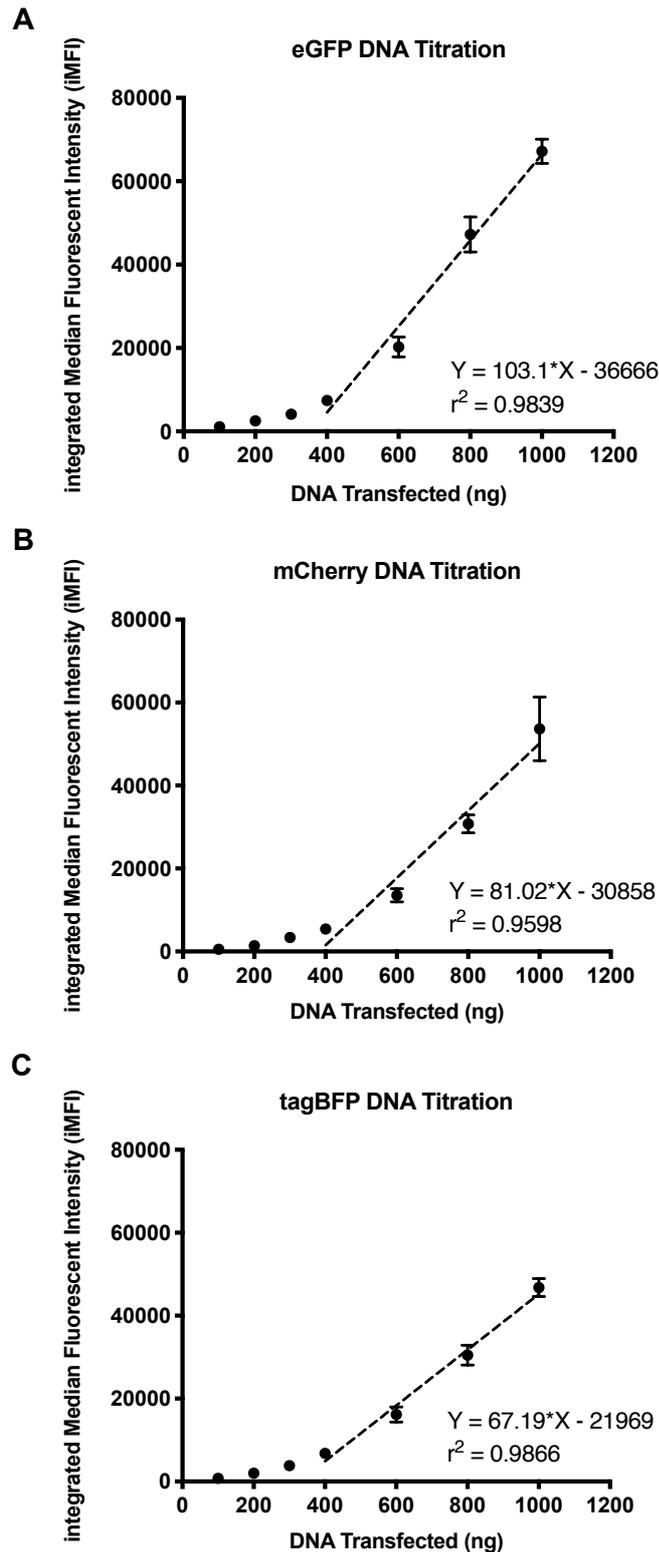


Figure 3-10: Dynamic range of *eGFP*, *mCherry* and *tagBFP* quantification by flow cytometry. Graphs showing relationship between iMFI and quantity of DNA transfected in Medi-CHO over 24h transient expression. Error bars represent the standard error of the mean (SEM) of biological triplicates. The data shows the dynamic range and sensitivity of fluorescent detection after 24h of transient expression of *eGFP* (A), *mCherry* (B) and *tagBFP* (C). The dynamic range of detection for all three fluorescent proteins was between 400 and 1000 ng as shown by dotted line representing linear regression analysis where linearity was determined by an $r^2 \geq 0.9598$, and limited sensitivity observed below 300 ng.

As observed in Figure 3-11, there was a variation in fluorescent expression between *eGFP*, *mCherry* and *tagBFP* when transfected at identical DNA loads and quantified by iMFI after 24 h.

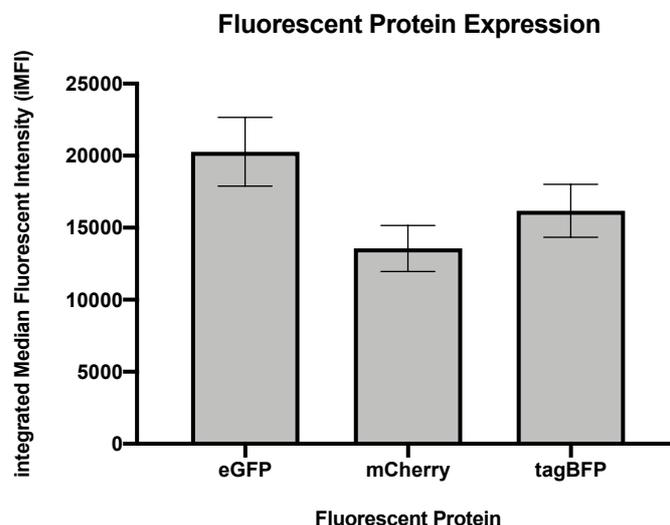


Figure 3-11: Differential level of fluorescent protein detection in Medi-CHO.

Bar chart demonstrates the differential levels of *eGFP*, *mCherry* and *tagBFP* when transfected identically in Medi-CHO. All three proteins were transiently transfected using 600 ng of their respective plasmid and quantified by iMFI after 24 h. As shown, the mean iMFI for each protein was different, however no significant difference was observed by a one-way ANOVA statistical test with a Tukey correction.

The mean iMFI in Figure 3-11 for *eGFP*, *mCherry* and *tagBFP* was 20268.92, 13561.72 and 16176.18 respectively. This indicates the fluorescent protein detection was variable since all three proteins were transfected at identical DNA quantities (600 ng). However, the difference in iMFI was not statistically different when analysed by a one-way ANOVA with a Tukey correction, Therefore, the difference in detection was variable but not significant, and could be caused by differences in protein maturation or quantum yield (Shaner et al., 2005). For example, quantum yield can be influenced by the innate nature of the protein or by the flow cytometer parameters (such as the combination of the excitation wavelength of the laser and the PMT voltage levels) leading to differences in fluorescent intensity.

3.3.4. Identifying an optimal DNA load of multi-gene expression vectors to avoid cytotoxicity

The quantity of DNA transfected into mammalian cells can cause cytotoxicity and impeded on growth and recovery of the cells post-transfection. A typical MGEV was hypothesised to be larger than the average plasmid for transfection and therefore a higher absolute mass of DNA is introduced per cell. The increased quantity could increase cytotoxicity when using the originally developed HT transient transfection protocol. The impact on MGEV-mediated cytotoxicity was assessed by varying the quantity of MGEV transfected at 400 ng, 600 ng and 800 ng of two variants along with a water control (Mock) using the HT transient transfection protocol in the MedI-CHO cell line. The two variants tested were MGEV-GBC and MGEV-CBG which encode for *eGFP*, *mCherry* and *tagBFP*. However, the TU order encoding the reporters within the MGEV were different where MGEV-GBC was TU1-eGFP, TU2-tagBFP and TU3-mCherry, whereas, MGEV-CBG was TU1-mCherry, TU2-tagBFP and TU3-eGFP. Within the MGEV variants, each TU was composed of the same peripheral genetic elements (hCMV-MIE promoter, 5' UTR and SV40 pA). The transfected cells were cultured for 24 h and the viable cell concentration and percentage viability were quantified using the ViCell XR automated cell counter.

Growth Characteristics of Transiently Expressed MGEV Variants

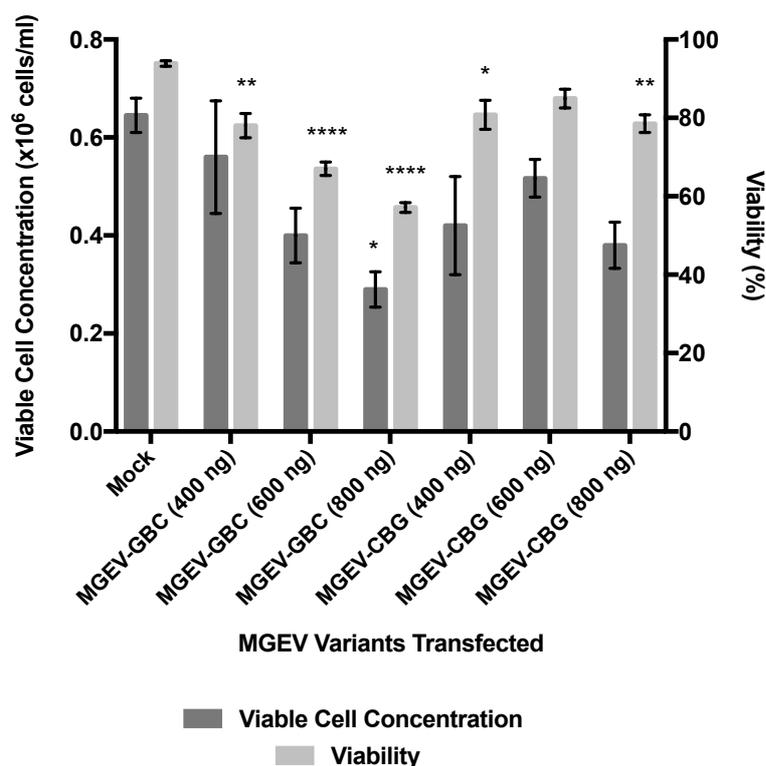


Figure 3-12: Growth characteristics of transiently expressed multi-gene expression vector variants.

Bar chart showing the effect on viable cell concentration and viability when varying the quantity (400, 600 and 800 ng) of two MGEV variants (MGEV-GBC and MGEV CBG) transfected in Medi-CHO after 24 h. The error bars represent SEM based on technical replicates of n=3 for MGEV variants and n=2 for mock transfected cells. The ‘*’ represent statistical significance using a Dunnett one-way ANOVA test comparing against mock cells where ‘*’, ‘****’ and ‘*****’ represents $p < 0.0332$, $p < 0.0021$ and $p < 0.0001$ respectively.

Figure 3-12 shows the viable cell concentration and percentage viability of transfected cells 24 h post-electroporation. The titration of MGEV DNA of 400 ng, 600 ng and 800 ng shows lower average viable cell concentration for both variants against mock cells. However, significant difference in viable cell concentration was only observed between the MGEV-GBC (800 ng) condition and the mock when using a Dunnett one-way ANOVA test. Similarly, the average viability of the cells transfected with MGEV were lower than the mock and the difference was significant when compared against the mock apart from the MGEV-CBG (600ng) condition. The data shows that viability of transfected cells is inversely proportional to the quantity of MGEV transfected for the MGEV-GBC variant, however the trend is less apparent for the MGEV-CBG variant. Although the trend is not consistent between two variants, the reduction in viable cell concentration and viability is affected by transfection of MGEV. In

conclusion, transfecting cells with 600 ng of MGEV has a sufficient trade-off between transfecting sufficient copies of the MGEV and minimising potential cytotoxic effects.

3.3.5. Investigating the gene expression dynamics within a multi-gene expression vector

The design of MGEVs and use of insulator elements between TUs as a fundamental feature for independent functionality of TUs has been explored in literature (Torella et al., 2014b). This indicated that a series of TUs assembled in tandem to form a polycistronic cassette can affect relative gene expression levels. Therefore, to investigate the gene expression dynamics within a MGEV, multiple variants were constructed by varying the position of the three fluorescent proteins (*eGFP*, *mCherry* and *tagBFP*), while maintaining the peripheral genetic elements. For example, all three TUs were driven by a hCMV-MIE promoter, a 5' UTR and a SV40 pA element. The fluorescent protein CDSes were swapped between the positions within the polycistronic cassette by using the 'plug-and-play' feature within the TU. Whereby, using restriction digestion-ligation cloning, the CDSes were combined with TU-1, TU-2 and TU-3 representing the order of the genes as shown in Figure 3-7. Table 3-2 summarises the MGEV variants constructed and the order of the fluorescent protein CDSes within the polycistronic cassette.

Table 3-2: Multi-gene expression vector variants constructed by varying fluorescent protein gene order.

Table showing the three MGEV variants constructed and the different order of *eGFP*, *mCherry* and *tagBFP* genes within the polycistronic cassette. These plasmids were used to investigate the variation in gene expression levels depending on the gene order in the MGEV.

MGEV Variant	Fluorescent Protein Gene Order		
	Position 1	Position 2	Position 3
MGEV-GCB	<i>eGFP</i>	<i>mCherry</i>	<i>tagBFP</i>
MGEV-BGC	<i>tagBFP</i>	<i>eGFP</i>	<i>mCherry</i>
MGEV-CBG	<i>mCherry</i>	<i>tagBFP</i>	<i>eGFP</i>

The aim of the MGEV variants was to measure the relative gene expression of the fluorescent proteins and investigate if there was a position-dependent difference in expression within a polycistronic cassette. However, differences in fluorescent protein detection was observed as described in Figure 3-11, hence inter-comparison of *eGFP*, *mCherry* and *tagBFP* would not allow accurate quantification of position-dependent effects. Therefore, changing the positions of the fluorescent protein within the polycistronic cassette allows direct comparison of each reporter in position 1, 2 and 3

across the MGEV variants and accounts for the variation in fluorescent protein detection. The expression dynamics were measured by transiently expressing the MGEV variants (MGEV-GCB, MGEV-BCG and MGEV-CBG) alongside co-expression of three SGVs encoding for the fluorescent proteins at a 1: 1: 1 ratio. The SGVs are the TUs within a basic vector backbone comprising of a β -lactamase gene and a microbial origin of replication. These TUs encoding for the fluorescent proteins are the building blocks within the MGEV and hence contain the same genetic elements (hCMV-MIE promoter, 5' UTR, fluorescent protein CDS and SV40 pA) as the polycistronic cassette within the MGEV variants. The MGEV variants and SGVs were quantified by using the HT transient transfection protocol with an optimised DNA load of 600 ng (as identified by the linear range for fluorescent protein quantification by flow cytometry) in the MedI-CHO host cell line. The transfected cells were cultured for 24 h and the fluorescence was quantified using the optimised flow cytometry settings. The iMFI was calculated using the optimised gating strategy for multi-fluorescent protein expression as described in Figure 2-2. The quantified fluorescence was plotted in a bar chart comparing gene loci within the polycistronic cassette for each fluorescent protein independently, as shown in Figure 3-13.

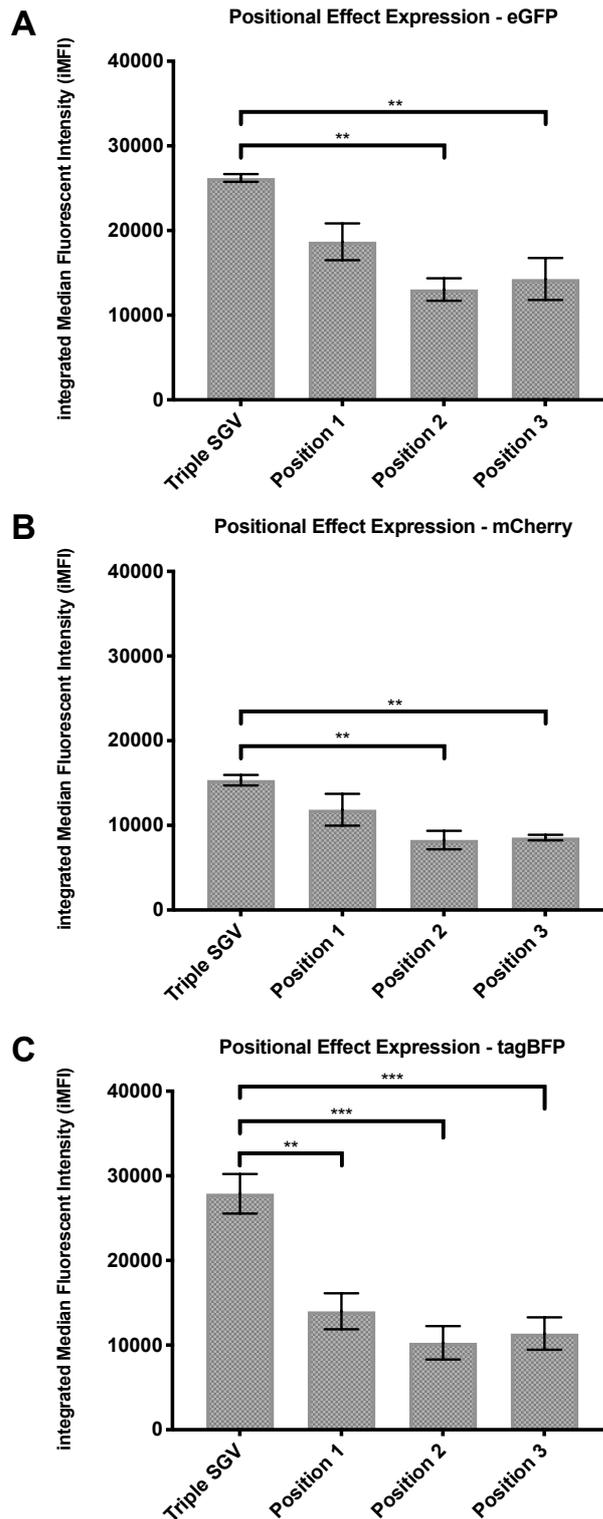


Figure 3-13: Recombinant gene expression dynamics within a multi-gene expression vector.

Panel of bar charts comparing the transient expression of three fluorescent proteins (*eGFP*, *mCherry* and *tagBFP*) by iMFI. The error bars represent the SEM for technical triplicates. The data compares the fluorescent expression of *eGFP* (A), *mCherry* (B) and *tagBFP* (C) in positions 1, 2 and 3 of the polycistronic cassette across the different MGEV variants (refer to table 3-3). The co-expression of three plasmids encoding for *eGFP*, *mCherry* and *tagBFP* was measured alongside the MGEVs and referred to as Triple SGV. A one-way ANOVA statistical test with a Tukey correction was performed and significance differences were stipulated by “***” and “****” representing a $p < 0.0021$ and $p < 0.0002$.

The results show that the average iMFI of *eGFP*, *mCherry* and *tagBFP* within the MGEV was lower than the triple SGV co-transfected variants. Although, 600 ng of MGEV and SGV (200 ng of each plasmid encoding for *eGFP*, *mCherry* and *tagBFP*) were transfected, a reduction in iMFI ranging from 22.7% to 49.8% was observed. Furthermore, statistical analysis by a one-way ANOVA test with a Tukey correction demonstrated significant difference in iMFI between the triple SGV, and position 2 and 3 for both *eGFP* and *mCherry*. Whereas, a significant difference was observed between triple SGV and all three positions for *tagBFP*, as displayed in Figure 3-12. Therefore, the decrease in fluorescent protein expression between three SGVs co-transfected and a single MGEV indicates a difference in gene expression dynamics within a MGEV.

A second observation was the reduction in average iMFI in position 2 and 3 compared to position 1 within the polycistronic cassette and a significant difference from the average iMFI of the triple SGV variant for all three fluorescent proteins, as shown in Figure 3-13. The trend in relative expression levels in relation to gene loci within the MGEV was position 1 > position 3 > position 2. The positional effect was quantified for each fluorescent protein by comparing the average iMFI in position 2 and 3 against position 1, as shown in Table 3-3.

Table 3-3: Relative fold change in fluorescent protein expression dependent on gene position within the polycistronic cassette.

Table showing the relative fold change in gene expression for each fluorescent protein across the MGEV variants. The fold change was normalised against average iMFI in position 1 to quantify the reduction in the downstream positions for each protein. An average fold change and the variation in fold change in the form of standard deviation was calculated across the three fluorescent proteins in both position 2 and 3. Additionally, an unpaired t test was performed on the iMFI of *eGFP*, *mCherry* and *tagBFP* between position 2 and 3 against position 1 where the difference was not statistically significant.

Fluorescent Protein	Position within Polycistronic Cassette		
	1	2	3
<i>eGFP</i>	1.00	0.70	0.76
<i>tagBFP</i>	1.00	0.73	0.81
<i>mCherry</i>	1.00	0.70	0.72
Average	-	0.71	0.77
Standard Deviation	-	0.02	0.04
Unpaired t test of iMFI in position 2 and 3 against position 1 for each fluorescent protein		not significant	not significant

The average relative fold change within the polycistronic cassette was 1.00: 0.71: 0.77, indicating a 29% and 23% reduction in relative gene expression in position 2 and 3 respectively. Furthermore, a 2% and 4% variation between *eGFP*, *mCherry* and

tagBFP in position 2 and 3 was observed indicating consistent reduction in gene expression in the respective positions. However, no significant difference was observed by an unpaired t test of the iMFI values over three technical replicates between position 2 and 3 against position 1 for each fluorescent protein. Therefore, a reduction in fluorescent protein expression can be observed in position 2 and 3 but the difference was not significant over the variation of the assay, hence this data alone was not conclusive.

As a result, a further investigation into the positional effect was performed by analysing the cell population distribution over the 10000 events recorded for each MGEV by the flow cytometer. The analysis was performed by identifying the subsection of the population that positively expressing all three fluorescent proteins by using an exclusion-based gating strategy. The fluorescent values for *eGFP*, *mCherry* and *tagBFP* of a single event/cell were extracted from FlowJo (analytical software to process flow cytometry data) in a tabular format. The individual fluorescent values per cell per MGEV variant were processed in collaboration with Dr Joe Cartwright and a Studio R script he developed to analyse the positional effect distribution within the population. The script normalises the difference in fluorescent brightness observed between *eGFP*, *mCherry* and *tagBFP* using historical single expression data. This allowed for direct comparison of fluorescent protein within the MGEV. The normalised fluorescent values were converted into a ratio per cell. The ratios were categorised within bins and the frequency of the ratio per fluorescent protein within the cell population was plotted. The data analysis was performed for the MGEV-GCB, MGEV-BGC and MGEV-CBG variants, and the frequency distributions are shown in Figure 3-14.

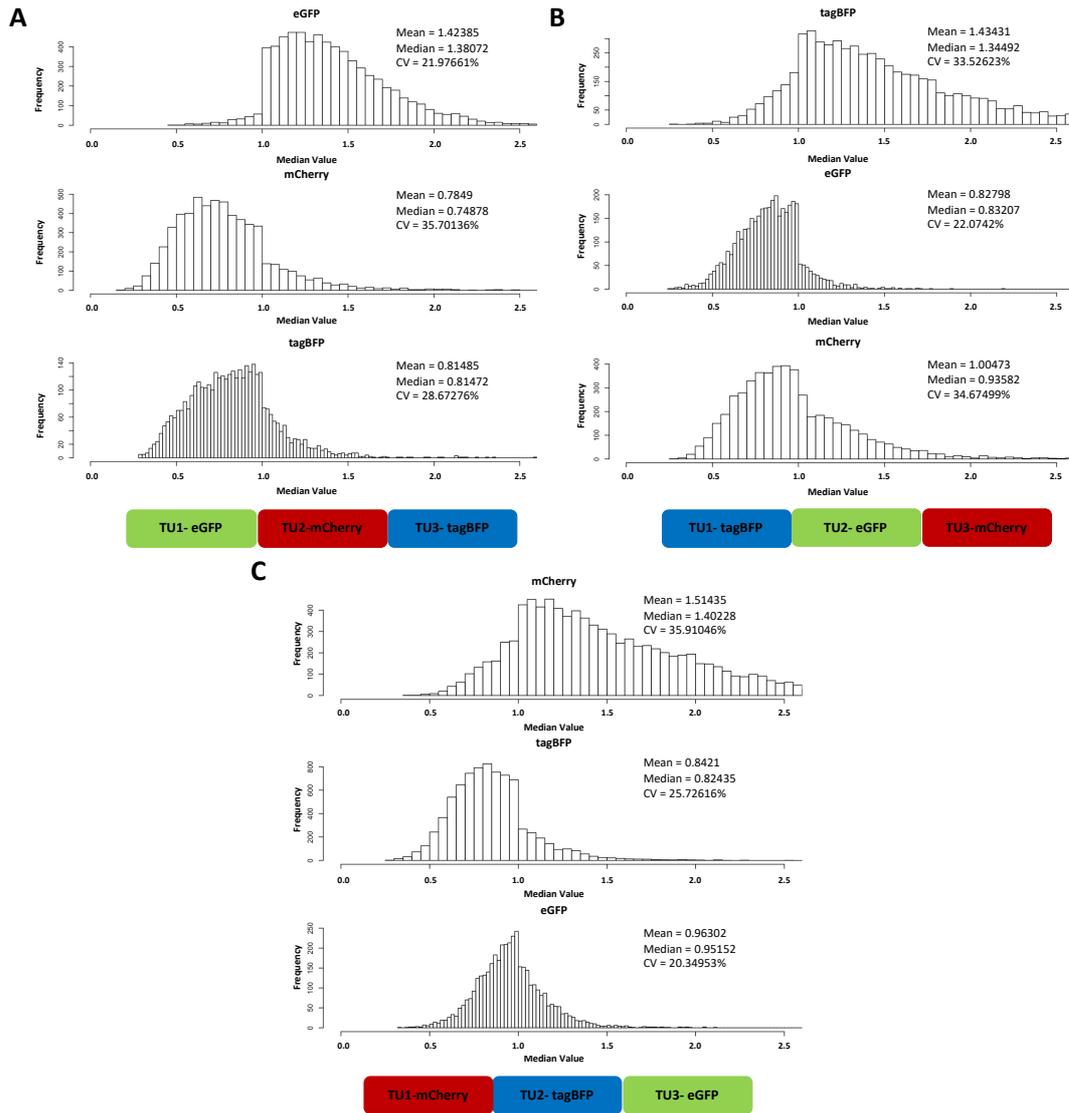


Figure 3-14: Single cell distribution of positional effect for three multi-gene expression vector variants.

Distribution of quantified positional effect using normalised fluorescent intensity values within a polycistronic cassette for MGEV-GCB (A), MGEV-BGC (B) and MGEV-CBG (C) within the transfected cell population. The fluorescent values were converted into ratios and categorised into bins. The frequency of a categorised ratio for each fluorescent protein was plotted and organised in order of the polycistronic cassette. The schematic below represents the order of the polycistronic cassette where the green, red and blue blocks represent the *eGFP*, *mCherry* and *tagBFP* TUs within the MGEV. Each frequency plot states the calculated mean and median ratio along with the co-variances.

The frequency distribution showed that for all three MGEV variants the fluorescent protein expressed within the cell population in position 1 was higher than the expected ratio of 1, whereas the distribution was lower than the expected ratio of 1 for position 2 and 3. The mean ratio for the fluorescent protein in position 1 was the highest at 1.42, 1.43 and 1.51 for MGEV-GCB, MGEV-BGC and MGEV-CBG variants respectively. The mean ratio in position 2 was the lowest at 0.78, 0.83 and 0.84 for MGEV-GCB, MGEV-BGC and MGEV-CBG variants respectively. The mean ratio in

position 3 was lower than position 1 but higher than position 2 at 0.81, 1.00 and 0.96 for MGEV-GCB, MGEV-BGC and MGEV-CBG variants respectively. The data reinforced the conclusion from the overall population analysis that relative expression trend within the polycistronic cassette was position 1 > position 3 > position 2. Additionally, the frequency distribution analysis indicates that the positional effect may be similar between MGEV variants, whereby the percentage variation of the mean expression ratio between variants in position 1, 2 and 3 was 4.9%, 3.2% and 10.0% respectively.

Table 3-4: Mean expression ratios within the polycistronic cassette of the multi-gene expression vector variants.

The table states the mean expression ratio within the polycistronic cassette across the transfected population expressing the different MGEV variants. The ratios indicate the relative fold change of fluorescent protein expression within the entire population.

MGEV Variant	Mean Expression Ratio in the Polycistronic Cassette		
	Position 1	Position 2	Position 3
MGEV-GCB	1.42	0.78	0.81
MGEV-BGC	1.43	0.83	1.00
MGEV-CBG	1.51	0.84	0.96

The hypothesis for the reduced levels of gene expression in the MGEV when compared against the SGV, and the positional effect within the polycistronic cassette may have been due to transcriptional interference. Transcriptional interference suppresses transcription of a neighbouring gene as a consequence of an initial transcription event (Palmer et al., 2011; Sherwin et al., 2005). There were a number of mechanisms which contributed towards transcriptional interference. The potential mechanisms which contributed towards relative suppression in gene expression downstream maybe a combination of RNA polymerase II (RNA pol II) occlusion, transcription factor (TF) or pre-initiation complex (PIC) dislodgement and TF competition (Palmer et al., 2011).

3.4. Discussion

This chapter describes the design process followed by the *in vitro* testing and characterisation of a MGEV specifically designed for CHO cell engineering. The aim of the work was to develop a MGEV construction protocol and understand the gene expression dynamics of multiple recombinant genes within a single plasmid.

The first step in designing a MGEV system for CHO cell engineering required identifying the design space and the desired features within the vector system. The design attributes were outlined in collaboration with MedImmune to develop a MGEV applicable in both the biopharmaceutical industry and academia. The list of features for the MGEV were as follows – a robust and reproducible expression system; a simple cloning methodology; license-free system; functional in a transient and stable platform; rapid assembly of multiple recombinant genes; cost effective; engineering flexibility and future proofing.

The following step was the *in silico* design process which began with deciding on an effective cloning technique. After reviewing the literature, Golden Gate assembly was identified to be the optimal method when compared against Gibson assembly and gateway cloning for its simplicity (a desired design criteria). The technique used a familiar process of restriction digestion-ligation cloning, however, the unique feature of the type II endonuclease, in this case BsaI, digested the phosphodiester backbone outside of the recognition site. Therefore, allowing for digestion and ligation to occur concurrently in a single tube. Furthermore, the by-product of digestion leads to 4-bp overhangs/sticky ends which could be designed *in silico* to be complementary and control order and orientation of multiple TUs during MGEV assembly.

The reproducibility feature within a MGEV was addressed by designing a standardised TU, whereby the structure of each recombinant gene followed a combination of defined genetic elements (e.g. synthetic/viral proximal element, core promoter, 5'UTR, CDS with Kozak sequence and signal sequence peptide, 3' pA element and terminator/insulator element). The hypothesis was that standardisation could promote consistency and minimise variation between TUs by omitting unannotated or unknown DNA sequences within the TU, in an aim to facilitate reproducibility and predictability of gene expression within a MGEV.

The TU structure also included a number of strategically positioned restriction enzyme recognition sites which were co-functional between genetic elements, allowing for an iterative 'plug-and-play' system. This feature allowed for engineering flexibility and future proofing by ease of manipulating and modifying the TU. The 'plug-and-play'

feature also minimised cost since TUs did not require repeated *de novo* synthesis and alternatively, shorter genetic fragments can be synthesised and cloned into the TU.

A recipient backbone vector was required to house the multiple TUs and comprised of additional elements to include other design features. The recipient vector included an episomal origin of replication element and a *GS* gene cassette for CHO cell line selection, allowing for functionality in transient and stable platforms respectively. Both of these elements can be removed or interchanged allowing for flexibility dependent on desired functionality.

The *in silico* designed MGEV was specifically designed to assemble three standardised TUs comprising of a hCMV-MIE promoter, 5' UTR, 3' SV40 pA UTR and encoding for either *eGFP*, *mCherry* or *tagBFP*, and the recipient vector. The genetic components were *de novo* synthesised for *in vitro* testing. The first step was to evaluate the Golden Gate assembly cloning technique using an optimised kit from NEB. 100% success rate was observed by restriction digestion colony screen of 12 transformed colonies. One of the plasmids was transfected and measured by flow cytometry as an orthogonal means of verifying successful cloning and preliminary functionality and DNA sequence verified to confirm correct construction of the MGEV (order of TUs within the polycistronic cassette). Data indicated the design and cloning technique was functional, with further MGEV variants constructed at a later date (data not shown), deeming the method reproducible and robust. The restriction digestion, ligation and transformation process for MGEV construction would take 3.5 h using Golden Gate assembly. Whereas, traditional cloning approaches would take multiple days to assemble the same TUs fragments in a sequential manner, indicating the cloning methodology was substantially quicker. Moreover, the single step process of Golden Gate assembly is simpler than the cumbersome traditional restriction digestion-ligation cloning methodology.

The 'plug-and-play' feature was also successfully designed by iteratively cloning the *eGFP*, *mCherry* and *tagBFP* CDSes within TU-1, 2 and 3 using the *AgeI* and *SbfI* restriction sites and further reiterated in chapter 4 and 5 by cloning different synthetic promoters and transcription terminator variants. Therefore, the TU was adaptable for different future engineering and experimental approaches, as well as, saving on

additional *de novo* gene synthesis costs. The MGEV is also license-free as all the parts were bespoke designs generated within this project, therefore, commercially viable for MedImmune's biopharmaceutical production platform. Finally, the episomal origin of replication and *GS* cassette within the MGEV were not specifically tested for long-term transient and stable gene expression. However, these element were supplied from MedImmune's well established platform and current vector framework. Therefore, the episomal origin of replication and *GS* cassette are predicted to perform as expected. Furthermore, the microbial origin of replication and a β -lactamase gene for ampicillin-mediated amplification in *E.coli* were also adapted from MedImmune's current vector system and have eased transferability by adopting existing molecular biology protocols within industry and academia.

As mentioned previously, *eGFP*, *mCherry* and *tagBFP* were the fluorophores selected as reporter genes to characterise gene expression trends within the MGEV. This was decided by assessment of literature and spectras. During optimisation of the flow cytometer settings for accurate quantification, an unexpected observation of false positive signal of *tagBFP* was detected during high fluorescent intensity ($>10^4$ units) detection of *eGFP* when expressed singularly in Medi-CHO. As a result, the PMT voltages for the two lasers (405 nm and 488 nm) used to excite the fluorophores were reduced and compensation was applied to eliminate the false signal. A consequence to this approach was the potential increase of false negative signal by compressing the population within a narrow window of fluorescence leading to low fluorescence overlapping with the host cells autofluorescence. This was deemed an acceptable compromise as a false positive signal would have a larger impact on the accuracy of co-expressing fluorophores as a function of gene expression. The quantification assay was further optimised by identifying the dynamic range of accurate fluorescent intensity. Identification was done by performing a DNA titration of each fluorophore by transient expression in Medi-CHO. The data indicated the linear range of detection was between 400 and 1000 ng of transfected DNA and 600 ng of transfected DNA would be ideal to observe both an increase and decrease in gene expression. Similarly, 600 ng of transfected DNA in Medi-CHO was identified as an optimal compromise for DNA cytotoxicity post-electroporation when compared against 0 ng, 400 ng and 800 ng (original DNA load for HT transient transfection protocol) of DNA.

The characterisation of the gene expression dynamics within a MGEV was investigated by quantifying the iMFI of *eGFP*, *mCherry* and *tagBFP*. This was achieved by constructing multiple variants whereby the loci of the three fluorophore CDSes were moved within the polycistronic cassette. Observations demonstrated that the overall gene expression of the fluorophores were relatively lower than the co-expression of three SGVs encoding for *eGFP*, *mCherry* and *tagBFP*. The suppression could be caused by RNA pol II occlusion by the transcription of the upstream *GS* gene (which could not be quantified by flow cytometry), as a form of transcriptional interference (Palmer et al., 2011) or by negative supercoiling in the wake of RNA pol II transcription elongation leading to inhibition of upstream transcription (Curtin et al., 2008).

Further analysis of the relative gene expression within the polycistronic cassette of *eGFP*, *mCherry* and *tagBFP* was calculated by comparing the same respective fluorophore across the different MGEV variants. This form of analysis was to discount for the difference in fluorescent brightness between the fluorophores. A trend in relative gene expression within the polycistronic cassette showed highest expression was observed in position 1 and lowest in position 2 with a relative reduction of 29%. The relative order of gene expression was position 1 > position 3 > position 2. When comparing the relative fold change against position 1 the average ratio between the three fluorophores in position 1, 2 and 3 was 1.00: 0.71: 0.77 respectively. The expression trend within the MGEV was further characterised by investigating the distribution of detection within the population. This was achieved by normalising the fluorescent brightness discrepancy to allow direct comparison of fluorophores within the MGEV. The distribution reinforced the overall population analysis, whereby position 1 had the highest relative gene expression level and position 2 had the lowest relative gene expression level with minor increase in position 3. Therefore, across three MGEV variants the same trend was observed.

Published literature has indicated that recombinant genes in tandem can have a suppressive effect on the gene downstream due to transcriptional interference. There are multiple mechanisms identified in literature and it was hypothesised that a combination of RNA pol II occlusion, TF or PIC dislodgement, and TF competition may have contributed towards the effect. RNA pol II occlusion involves the pausing

of the polymerase in proximity of the downstream promoter where PIC formation could be affected by steric hinderance. TF or PIC dislodgement is where the elongating RNA pol II crosses into the downstream promoter forcing dissociation of the enhancer TF or PIC. Finally, TF competition can be that three units of hCMV-MIE promoters per MGEV copy driving fluorescent protein expression, and limited levels of enhancer TFs within the MedI-CHO host could hamper gene expression levels (Palmer et al., 2011; Shearwin et al., 2005). However, within a 3-gene system, the slight elevation in the final position could be a by-product of a higher level of transcriptional interference in position 2. Whereby the promoter region of the third TU had less interference. For example, due to lower levels of RNA pol II recruitment in position 2, there were lower levels of occlusion in position 3, therefore slightly reversing the suppressive effect.

The observations of suppressed gene expression within the MGEV and positional effects were determined over technical replicates and a significant difference was only observed between fluorescent protein expression between the three SGVs co-expressed and the MGEV. Therefore, more replicates are required to assist in empirically quantify gene suppression and positional effects. However, other literature on multi-gene plasmid designs have also identified variable gene expression from TUs assembled in a tandem series influenced by transcriptional interference and the requirement of insulators to avert these effects (Torella et al., 2014b). Therefore, the observations in gene expression trends within this experiment can be used to infer overall conclusions in gene expression performance within the MGEV.

The data demonstrated that there may be a level of consistent suppression by transcriptional interference occurring when the same relatively high strength promoter was used. However, the question arises whether the transcriptional interference was constant and independent of promoter strength. Additionally, could it be accounted for, such that predictable and quantifiable stoichiometric ratios using a library of synthetic promoters could be achieved. Alternatively, is a specific transcription terminator or insulator element essential to eliminate the interference and achieve modularity within a MGEV? These discussion points are addressed and extensively explored in Chapters 4

Chapter 4

4. Application of Synthetic Promoters in a Multi-Gene Expression Vector to Stoichiometrically Control Gene Expression in CHO cells

This chapter investigates the application of synthetic promoters with varying transcriptional activity in a multi-gene expression vector (MGEV) to control expression stoichiometry of multiple recombinant genes simultaneously in CHO cells. This was achieved by adapting a validated synthetic promoter library for compatibility with the MGEV system and verifying the functionality using a fluorescent protein reporting system in the MedI-CHO host cell line by transient expression and showed titratable gene expression. Further characterisation of the combinatorial functionality of a subset of synthetic promoters highlighted that promoter interaction was observed and inferred to be a consequence of promoter squelching combined with metabolic burden. The positional effects within a MGEV were investigated using different synthetic promoters which indicated that transcriptional interference is dependent on transcriptional strength. A library of 27 MGEVs using a low, medium and high strength synthetic promoter spanning every combination within a three-gene system was quantified by qPCR to further discern transcriptional repression trends within the polycistronic cassette. For example, expression within the MGEV was substantially repressed compared to single gene plasmid expression and the synthetic promoters exhibited variable transcription from their expected activity. Finally, the library also yielded a set of empirically derived gene expression stoichiometric ratios which could be applied to future engineering applications.

4.1. Introduction

Multi-gene expression vectors (MGEVs) can be applied in various areas of biotechnology including biofuel production, synthetic biology, gene therapy, and cellular reprogramming (Guye et al., 2013; Weber et al., 2011). Currently, various publications and reviews have demonstrated these applications using synthetic circuitry. A synthetic circuit is similar to a MGEV, that multiple genes are expressed simultaneously. However, a synthetic circuit comprises of a gene network made up of multiple single gene plasmids using inducible promoters and synthetic transcription factors (TFs) to work synergistically. A synthetic circuit applies aspects of electronic circuitry including switches and oscillators to control gene networks in a biological system (Ruder et al., 2011).

For example, mammalian cells have been reprogrammed using gene circuits to arrest cells at G1 phase of the cell cycle by regulating cyclin-dependent kinase inhibitor p27 to generate a mixed population of proliferating and non-proliferating cells. Therefore, mimicking a cell-based cancer model for the development of cancer therapies. Moreover, a complex logic gate gene switch embedded within a synthetic gene circuit was designed to target and induce cancer cell death. This was achieved by using two tumour-specific promoters, when co-induced leads to a signal cascade converting a prodrug to ganciclovir causing target-specific cancer cell death (Nissim and Bar-Ziv, 2010). Alternatively, bacteria such as *Salmonella spp* have been engineered to target cancerous cells by using chemotactic receptors detecting aspartate and invading cancerous cells to deliver a drug component (Weber and Fussenegger, 2012). Other applications of successful mammalian cell engineering using synthetic circuits was demonstrated by Ye et al, who engineered HEK293 cells to express a glucagon-like peptide variant using a synthetic circuit signal cascade induced by an optogenetic transcription device. These engineered cells were encapsulated and transplanted in type 2 diabetic mice to improve blood glucose homeostasis after exposure to blue light, therefore identifying a potential cell therapy for treatment of type 2 diabetes (Ye et al., 2011). Therefore, the examples described above demonstrate how complex synthetic circuitry can benefit various biomedical applications.

Synthetic circuits and MGEVs also have beneficial applications towards drug molecule production. Publications have described using circuits to engineer and

reconstruct the mevalonate-dependent biosynthetic pathway from yeast cells into *E.coli* to reduce the production cost of an artemisinin precursor for anti-malarial treatment (Khalil and Collins, 2010). Similarly, successful metabolic engineering of the isoprenoid pathway in *E.coli* was achieved using a modular approach to alter the gene network stoichiometry using promoters and gene copy number, in turn improving production of Taxol (Ajikumar et al., 2010). However, there has been limited development of MGEVs and synthetic circuits for biopharmaceutical production in CHO cells. MGEVs would be advantageous since they would facilitate the co-expression of a therapeutic protein alongside an accessory gene to improve recombinant protein production. For instance, recent research has shown the co-expression of *XBP-1* alongside a difficult-to-express (DTE) monoclonal antibody (mAb) (Pybus et al., 2014) or *CypB* alongside a Fc-fusion protein (Johari et al., 2015) have both enhanced protein expression.

Currently, majority of multi-gene engineering tools that are available are fully synthetic circuits. A disadvantage of synthetic circuits is the requirement of recombinantly expressing synthetic TFs which may have an additional metabolic burden on the host cell. This is non-ideal in biopharmaceutical production and gene therapy, where significant metabolic burden is already endured by the host cell for the recombinant expression of a therapeutic protein/target. Therefore, the additional burden of concurrently expressing synthetic TFs would negatively impact on the host cell. Furthermore, only synthetic circuits have demonstrated precise control of induced or repressed gene expression. For example, a synthetic gene network was developed and coupled with a transcriptionally controlled repressor element that concurrently switches off a RNAi component. This demonstrated >99% repression in CHO and HEK cells. The system also demonstrated tunability by dose response of the inducing element leading to high, medium and low repression (Deans et al., 2007; Ruder et al., 2011). However, no multi-gene system has demonstrated more precise tunability and quantifiable gene expression, hence warranting the study described in this chapter.

Titration of gene expression within a MGEV is achievable by using synthetic promoters. Synthetic promoters are fully designed and *de novo* synthesised DNA elements which regulate gene expression. The structure of a synthetic promoter comprises of transcription factor regulatory elements (TFREs) and a minimal core element as

described in section 1.7.1. The main type of synthetic promoters are designed as part of a full synthetic system whereby the TFREs are designed to interact with a synthetic TF (Kemmer et al., 2010). An alternative design is a constitutive synthetic promoter using the host cells TF repertoire to regulate gene expression. A number of publications have developed this derivative of promoters for various bacterial (Seghezzi et al., 2011; Yim et al., 2013), yeast (Blazeck et al., 2012; Stadlmayr et al., 2010), and mammalian cells (Ferreira et al., 2011) including CHO cells (Brown et al., 2014).

Two approaches have been established for the design of constitutive synthetic promoters. One approach is the randomised assembly of synthetic oligonucleotide repeats of *cis*-regulatory elements which are defined and demonstrate a range of transcriptional strength (Brown et al., 2014). The second approach involves an *in silico* profiling of TF abundance using transcriptomic data of the host cell, followed by identifying mutually exclusive TFREs and finally determining the relative transcriptional activity of the TF-TFRE interaction (Brown et al., 2017).

In this project, a synthetic promoter library developed by Dr. Adam J. Brown using an *in silico* design approach, as described in the Brown et al publication, was employed for titrating gene expression. This library of synthetic promoters were designed for controlling gene expression at varying levels within the MedI-CHO cell background both during transient and stable gene expression. The promoter activity was quantified by measuring secreted alkaline phosphatase (*SEAP*) expression (Brown et al., 2017). The library comprises of 14 fully characterised promoters exhibiting gene expression strength of approximately 5%, 10%, 20% 40%, 60%, 80% and 100%. Therefore, the step-wise range of transcriptional strength has the potential to titrate expression of multiple recombinant genes and derive a stoichiometric ratio. Furthermore, the comprehensive testing both by transient and stable expression of the promoter library concludes the robustness and predictability of each promoter. However, minimal work has been performed on understanding the functionality and predictability of the promoter library working in synergy when multiple synthetic promoters with varying transcriptional activity are paired together.

There are many advantages of titrating gene expression at defined stoichiometric ratios within a MGEV. For example, the synthetic promoters provide a defined relative transcriptional activity (RTA) and therefore are predictable. This can be used as a tool to further characterise the positional effects within a polycistronic cassette at different transcriptional strengths (as discussed in Chapter 3 – Figure 3-13 & 3-14). Alternatively, investigating whether synthetic promoters can be applied to compensate for transcriptional interference within a MGEV. Therefore, providing extensive understanding of the genetic context of MGEVs and their functionality. Furthermore, at present, minimal data is available that has quantified the impact of transcriptional interference on post-transcription (mRNA copies) and protein synthesis. This is particularly important as most CHO cell engineering strategies require the co-expression of molecular chaperones or transactivators to increase biopharmaceutical production, therefore, the beneficial effects are observed at a protein level. As a result, the development of methodologies to measure gene expression both at a transcriptional level and protein level would be advantageous in quantifying gene expression titration and dynamics.

The titratability of multiple recombinant genes has advantages in facilitating enhanced biopharmaceutical production. For instance, research has shown achieving an optimal light (*LC*) and heavy chain (*HC*) gene expression ratio can significantly enhance mAb titres by 8-fold (Pybus et al., 2014). Another example is the co-expression of chaperone proteins or TFs involved in protein folding, secretion and glycosylation which can improve performance or introduce new phenotypes within the CHO cell factory. Recent data generated by Dr. Joe Cartwright and Dr. Claire Arnall has shown that varying the gene dose by DNA copies of TFs like XBP-1s alongside a DTE mAb can significantly increase mAb titres (Cartwright et al., 2020). Therefore, to achieve this in a stably expressing cell line, predictable titrated gene expression by synthetic promoters regulating *LC*, *HC* and *XBP-1s* gene expression at a desired stoichiometric ratio could further enhance mAb production.

Currently, publications have indicated the presence of transcriptional interference within gene circuits by the requirement and testing of insulator elements (Liao et al., 2018; Torella et al., 2014b). However, no quantification of transcriptional interference mediated positional effects within MGEVs, and synthetic circuitry has been

performed. Furthermore, no papers have demonstrated quantification of relative stoichiometric ratios of multiple recombinant genes within a MGEV. Therefore, based on these premises, the hypothesis of this study was that constitutive synthetic promoters of varying transcriptional activity within a MGEV can achieve predictable multi-gene stoichiometry by controlling expression of multiple transgenes simultaneously. Moreover, to using differing promoter strength can elucidate gene expression dynamics within a polycistronic cassette.

This was achieved by first adapting the library of synthetic promoters to be compatible with the designed transcription unit (TU) and MGEV system. The modified synthetic promoter library was tested to validate transcriptional activity within the new expression cassette and system. Two methodologies were developed and assessed to quantify differential levels of multiple reporter genes at a transcriptional and protein level using qPCR and flow cytometry respectively. The qPCR method was shown to be an accurate means to quantify differential gene expression, whereas flow cytometry was shown to be a semi-quantitative tool for gene expression trends. A low, medium and high transcriptionally active synthetic promoter was identified, and further characterisation was performed to understand and quantify the self and cross-interaction between the synthetic promoter variants. The positional effect within the polycistronic cassette of a MGEV was quantified and indicated to be promoter strength dependent. A library of 27 different MGEVs with varying combinations of a low, medium and high transcriptionally active synthetic promoter within a polycistronic cassette were screened. This identified specific transcriptional repression trends such as expression within the MGEV was substantially repressed compared to its single gene plasmid counterparts and the synthetic promoters exhibited variable transcription activity from their expected performance. Finally, a set of empirically derived gene expression stoichiometric ratios for each MGEV variant was calculated for potential future multi-gene engineering applications in CHO cells.

4.2. Results

4.2.1. Adapting and cloning of a synthetic promoter library into the transcription unit for multi-gene expression vector construction

The synthetic promoter library developed by Dr. Adam J. Brown was modified to be compatible with the MGEV system designed in Chapter 3. This was achieved by PCR, modifying the synthetic proximal element to introduce a BamHI and NheI restriction site on the 5' and 3' of the sequence respectively. This allowed the interchangeability of the synthetic proximal element using the 'plug-and-play' feature within the TU. Therefore, the element was cloned upstream of the human cytomegalovirus -major intermediate-early (hCMV-MIE) core promoter as shown in Figure 4-1. Each TU comprised of a hCMV-MIE core element to standardise the promoter structure such that the synthetic proximal element dictated transcriptional activity. Moreover, the same 5' UTR and SV40 pA (adopted from MedImmune's commercial vector system since they were previously performance tested both in industry and academia) were used within the TU (as shown in Figure 4-1), and each fluorescent protein CDS was codon optimised for CHO cell expression. Therefore, standardising the rate of transcription termination and translation between TUs. As a result, the protein expression and mRNA copies observed would be directly correlated to the transcriptional activity of the synthetic promoter employed for each fluorescent protein.

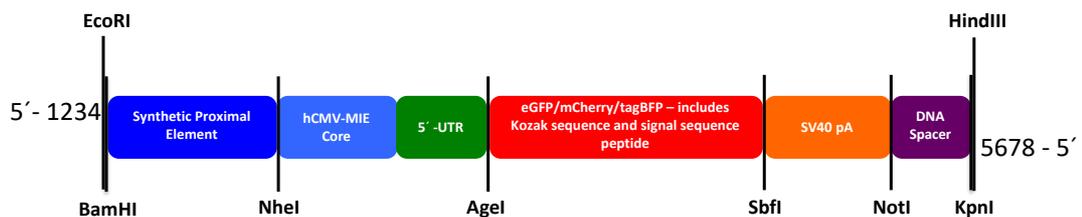


Figure 4-1: Transcription unit composition for assessing synthetic promoter activity.

The schematic depicts the genetic elements selected and the layout of the TU encoding for three fluorescent proteins – *eGFP*, *mCherry* and *tagBFP*. The synthetic proximal element was cloned upstream of the hCMV-MIE core using BamHI and NheI. The same hCMV-MIE core element, a MedImmune proprietary 5' UTR, SV40 pA element and an inactive DNA spacer were used consistently for each TU. Therefore, the synthetic proximal element was the only fragment to be interchanged within the TU to measure the synthetic promoter variant activity both within single gene vectors and a MGEV.

The PCR was performed using a series of forward and reverse primers (refer to Appendix B) to amplify the synthetic proximal element over 35 cycles using various

optimal annealing temperatures based on the primer pair properties. The PCR amplicons were verified by DNA fragment size using gel electrophoresis. The results of the gel electrophoresis are shown in Figure 4-2.

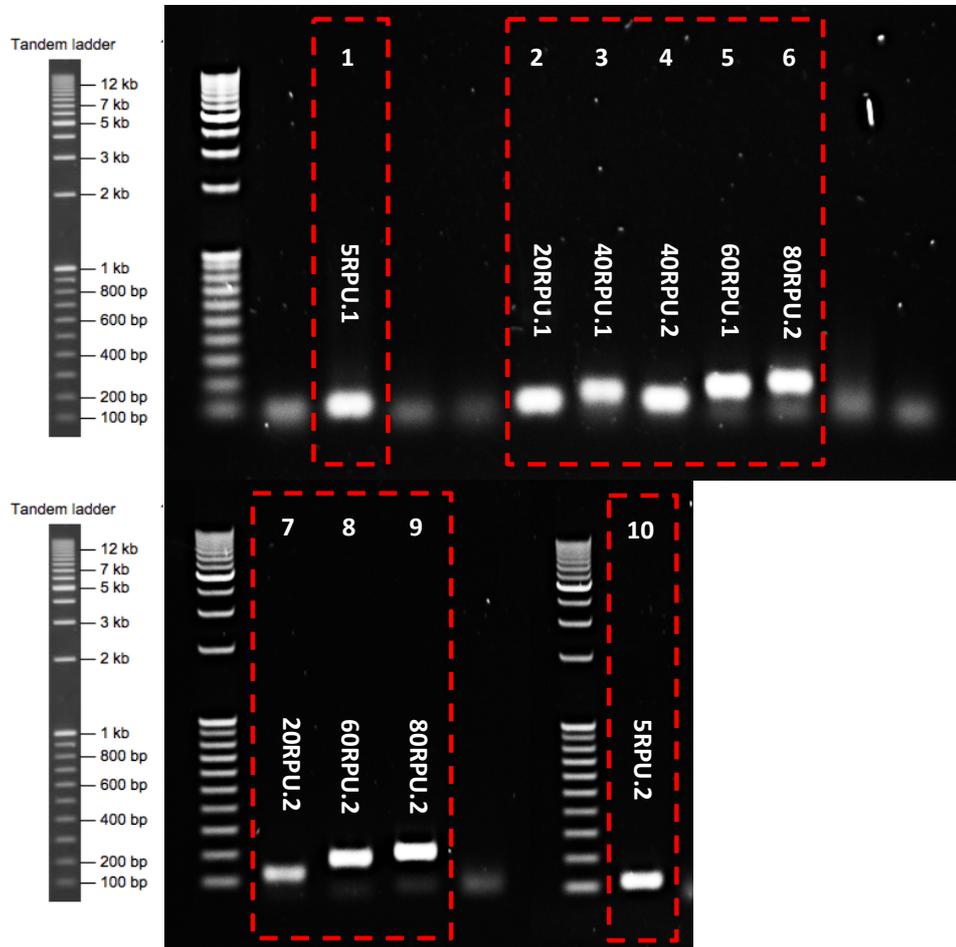


Figure 4-2: PCR modification of synthetic proximal enhancer element.

A 1% agarose gel showing the successful PCR amplification of modified synthetic proximal elements. The bands highlighted within the red box represent the proximal elements successfully modified and amplified, as deduced by the size of the band against the reference ladder. Lane 1 to 10 represent the proximal region of 5RPU.1, 20RPU.1, 40RPU.1, 40RPU.2, 60RPU.1, 80RPU.2, 20RPU.2, 60RPU.2, 80RPU.2 and 5RPU.2 synthetic promoters respectively. These promoters were derived from the library developed by Brown et al.

Figure 4-2 shows the successful modification and amplification of the 5RPU.1, 5RPU.2, 20RPU.1, 20RPU.2, 40RPU.1, 40RPU.2, 60RPU.1, 60RPU.2, 80RPU.1 and 80RPU.2 proximal elements, as highlighted by the red boxes within the figure. This was deduced based on expected size of each amplicon as shown in Table 4-1 against the tandem DNA ladder.

Table 4-1: Expected Amplicon size for each synthetic proximal element.

Table listing the synthetic proximal element PCR modified and amplified with the expected amplicon size to verify successful modification.

Synthetic Proximal Element Name	Expected Amplicon Size (bp)
5RPU.1	102
5RPU.2	101
20RPU.1	106
20RPU.2	118
40RPU.1	146
40RPU.2	125
60RPU.1	183
60RPU.2	176
80RPU.1	201
80RPU.2	201

The 10RPU.1, 10RPU.2, 100RPU.1 and 100RPU.2 were unsuccessfully modified and amplified by the lack of DNA fragments detected. A number of repeat PCRs were performed, varying the annealing temperatures and using longer, more specific primers that yielded no success (data not shown). It was decided not to *de novo* synthesise these promoter variants as a sufficient range of transcriptional activity was represented by the proximal elements successfully amplified as determined by published *in vitro* expression data (Brown et al., 2017). Moreover, the cost of *de novo* synthesis was determined to outweigh the information potentially gained for this specific study. As a result, these promoter variants were no longer pursued for the rest of the study.

The successfully modified synthetic proximal elements were purified using Qiagen's PCR clean up kit. These elements were then cloned into TU-1, 2 and 3 encoding for *eGFP*, *mCherry* and *tagBFP* by restriction digestion-ligation cloning using BamHI and NheI endonucleases. Transformed *E.coli* colonies were screened by restriction digest colony screen; whereby purified plasmid DNA was digested using BamHI and NheI endonucleases and verified by gel electrophoresis. The results of the restriction digestion colony screen are shown Figure 4-3.

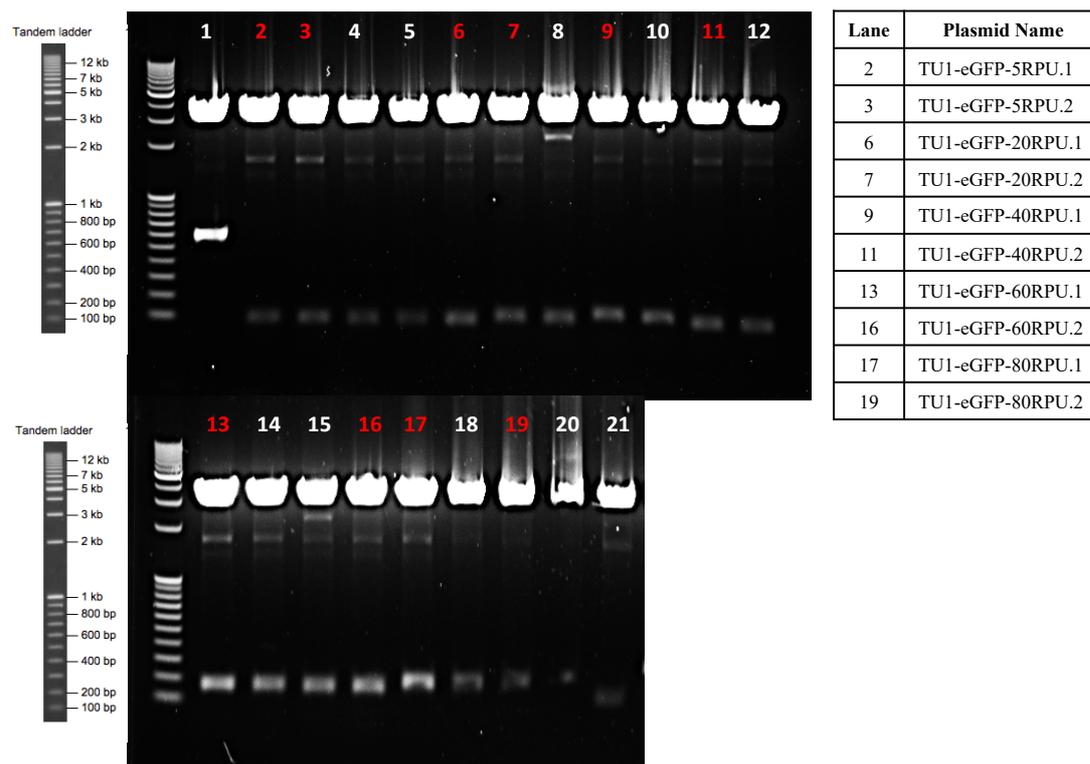


Figure 4-3: Cloning of synthetic promoter proximal elements into the transcription unit.

A 1% agarose gel showing the results of a restriction digest-based colony screen using BamHI and NheI endonucleases. The image indicates the successful cloning of the PCR modified synthetic proximal element into TU-1 encoding for *eGFP*. The colony numbers highlighted in red were selected and progressed to DNA amplification. The table summarises the lanes highlighted in red representing the synthetic proximal variant cloned into the TU1-eGFP plasmid.

Figure 4-3 shows the successfully cloning of the synthetic promoter proximal element into TU-1 encoding for *eGFP*. This was deduced by the DNA fragment sizes observed as a result of the restriction digestion reaction. The size of the smaller fragment represented the length of the synthetic promoter proximal element. Additionally, the faint band observed at < 2 kb in Figure 4-3 was likely to be supercoiled undigested plasmid DNA. The numbers highlighted in red in Figure 4-3 were the positively identified colonies and progressed forward to amplification. An identical screen was performed for the cloning of the elements into TU-2 and TU-3 (data not shown).

4.2.2. Characterising the modified synthetic promoter library by fluorescent protein expression

The synthetic promoter library was modified by changing the core element to be compatible with the MGEV system. Furthermore, the original library was tested and developed in a secreted alkaline phosphatase (*SEAP*) reporter. Therefore, the modified promoters were verified by measuring fluorescent protein expression of *eGFP*,

mCherry and *tagBFP*. The first step was to optimise the flow cytometer settings as a different flow cytometer was used for this study compared to Chapter 3. Therefore, the same process of optimisation was performed as described in section 3.3.3.1 for the Attune NxT flow cytometer. An additional observation was made where positive detection was observed within the VL1-A channel when only DNA encoding *mCherry* was being recombinantly expressed. This would have led to false positive detection of *tagBFP* whenever *mCherry* was expressed. This false detection was removed by a combination of reducing the PMT voltage of the 561 nm (yellow) laser and applying compensation. The optimised settings were finalised and applied for all further experiments and are summarised in Table 4-2.

Table 4-2: Optimised settings for the Attune NxT flow cytometer.

The table specifies the lasers and emission filters used along with their designated channel name in reference to dot plots to measure the expression of *eGFP*, *mCherry* and *tagBFP*. The PMT voltages were optimised and the final voltages were used for all further flow cytometry experiments in this chapter.

Laser Name	Excitation Wavelength (nm)	Emission Filter (nm)	Channel Name	Fluorophore Detected	PMT Voltages
Forward Scatter	N/A	N/A	FSC-A	N/A	405
Side Scatter	N/A	N/A	SSC-A	N/A	347
Violet	405	440/50	VL1-A	<i>tagBFP</i>	165
Blue	488	530/30	BL1-A	<i>eGFP</i>	170
Yellow	561	620/15	YL2-A	<i>mCherry</i>	230

The synthetic promoters were further curated for compatibility with the MGEV system by screening for undesired restriction endonuclease sites. It was identified that some variants of the 10 promoters had a *BsaI* restriction site present within the proximal region, which would hinder Golden Gate-mediated MGEV construction. This occurrence was because the synthetic promoter library was designed using consensus TFRE block sequences and the combination of TFRE sequences employed led to the emergence of the *BsaI* site. Due to the promoters being sequence specific, removing the *BsaI* site could impact on cognate TF binding affinity and negatively impacting the promoter's transcriptional activity. Therefore, the synthetic promoters with the *BsaI* site present were omitted from the study.

The 5 synthetic promoters deemed compatible with the MGEV system were 5RPU, 20RPU, 40RPU, 60RPU and 80RPU. The expression characteristics of these promoters were quantified by expression of *eGFP*, *mCherry* and *tagBFP* by flow cytometry. This was achieved by transient expression of the fluorescent proteins regulated by the synthetic promoter library in Medi-CHO after 24h using the HT transient transfection protocol. The fluorescent intensity as a function of gene expression was measured using the Attune NxT according to the settings stated in Table 4-2. The cell population expressing the fluorescent proteins were identified by gating for viable cells, followed by single cells and exclusion of doublets. Cellular autofluorescence was excluded by using mock transfected (with d.H₂O) cells in the VL-1, BL-1 and YL-2 channels. The fluorescent protein expression was quantified by measuring the integrated median fluorescent intensity (iMFI) as shown in following equation:

$$iMFI = MFI \times \text{frequency of parental population (\%)}$$

Where *MFI* = median fluorescent intensity of fluorescent protein detection and *frequency of parental population* = percentage of parental population showing positive fluorescent protein detection. This approach is used to measure the total detection of fluorescence (Darrah et al., 2007). The iMFI for each fluorescent protein driven by different synthetic promoters were plotted in a bar chart as shown in Figure 4-4.

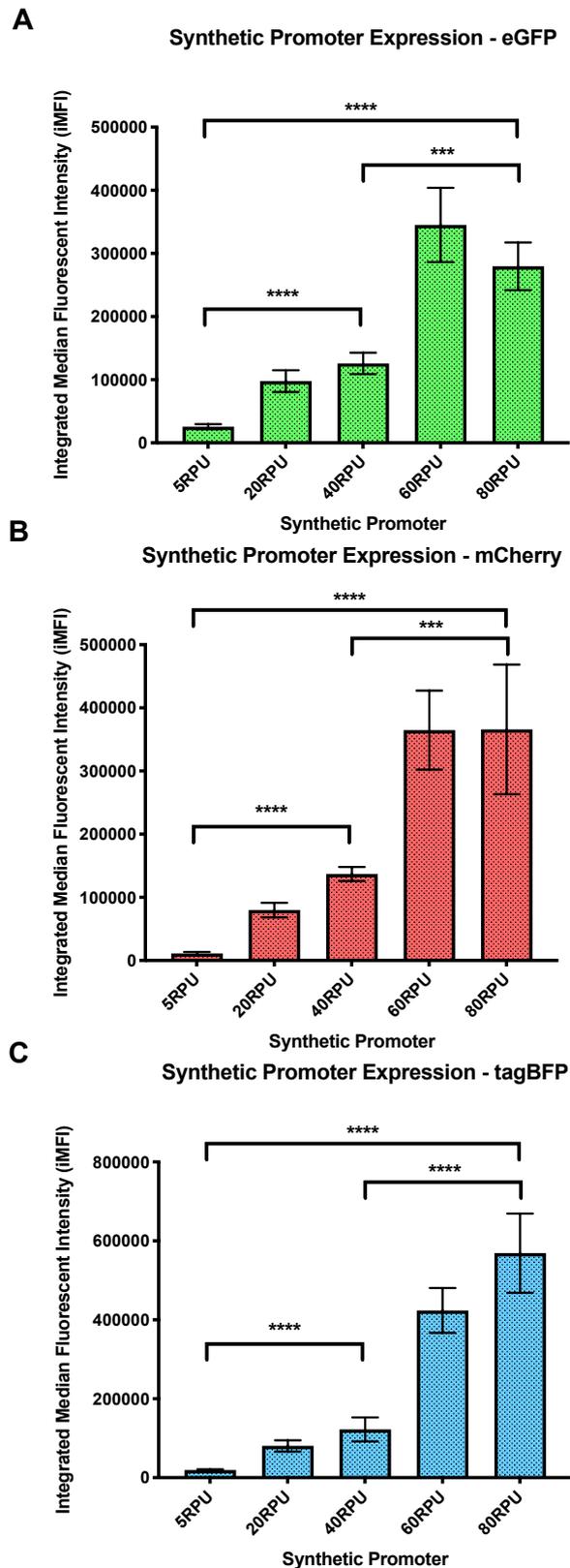


Figure 4-4: Quantifying synthetic promoter performance within a transcription unit.

The performance of the modified synthetic promoters were tested by measuring iMFI of *eGFP* (A), *mCherry* (B) and *tagBFP* (C). The error bars represent standard error of the mean (SEM) of biological triplicates. A multiple t test statistical analysis was performed and identified a significant difference between the 5RPU, 40RPU and 80RPU promoters for each fluorescent protein. The ‘***’ and ‘****’ represents a $p < 0.002$ and $p < 0.0001$ respectively.

The synthetic promoters show a gradient of expression for all three fluorescent proteins. The lowest expression for all three fluorescent proteins was regulated by the 5RPU promoter, whereas the highest expression was regulated by the 80RPU promoter apart from *eGFP*. The average fold change from 5RPU to the 80RPU for *eGFP*, *mCherry* and *tagBFP* was 11.23, 31.82 and 28.88 respectively. The entire range of fold changes are stated in Table 4-3.

Table 4-3: Relative fold change in fluorescent protein expression compared against 5RPU.

The average fold changes of three biological replicates when compared against the lowest transcriptionally active synthetic promoter- 5RPU, for *eGFP*, *mCherry* and *tagBFP*.

Synthetic Promoter	<i>eGFP</i> Expression Fold Change to 5RPU	<i>mCherry</i> Expression Fold Change to 5RPU	<i>tagBFP</i> Expression Fold Change to 5RPU
5RPU	1.00	1.00	1.00
20RPU	3.90	7.30	4.12
40RPU	5.19	12.92	6.21
60RPU	14.19	32.99	21.64
80RPU	11.23	31.82	28.88

Apart from *eGFP*, the fold changes between the weakest and strongest promoters were as expected. However, a minimal difference between the 60RPU and 80RPU unit was observed across all three proteins. Furthermore, the highest fold change was observed with the 60RPU for *eGFP* and *mCherry* (although marginally for *mCherry* expression). The unexpected expression profile of *eGFP* maybe a by-product of intracellular accumulation leading to proteolytic effects or aggregation (Krasowska et al., 2010). A step-wise gradient of titrated fluorescent protein expression was observed, but the separation was less distinct than expected. The multiple t test statistical analysis (including a Benjamini and Hochberg correction) showed a significant different between the 5RPU, 40RPU and 80RPU promoters, as shown in Figure 4-4. Therefore, the 5RPU, 40RPU and 80RPU promoters were identified as a low, medium and high transcriptionally active synthetic promoter respectively.

4.2.3. Developing and assessing assays for characterising positional effects and quantifying titrated gene expression

Real-time quantitative PCR (qPCR) and flow cytometry were the two techniques identified to measure positional effects and titration of gene expression within a MGEV co-expressing three reporter genes. These methodologies would quantify the expression of fluorescent proteins both at a transcriptional and protein level. In turn, to comprehensively characterise the functionality and titratability of the MGEV as an engineering tool. Both approaches were assessed for accuracy and sensitivity in an aim to determine the suitability of the assay. This was tested by co-transfecting a mixture of three separate single gene vectors (SGVs) encoding *eGFP*, *mCherry* and *tagBFP* using the same promoter variants at equimolar ratios, as shown in Figure 4-5.

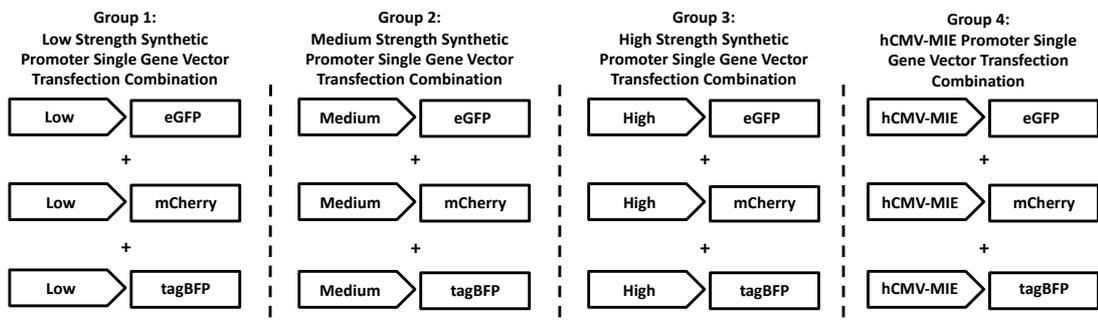


Figure 4-5: Schematic summarising the transfection conditions to analyse the dynamic range of detection of qPCR and flow cytometry.

Experimental outline depicting the equimolar co-transfection of three separate single gene vectors (SGVs) encoding for *eGFP*, *mCherry* and *tagBFP* while utilising the same synthetic promoter variant or hCMV-MIE promoter. A low, medium and high strength synthetic promoter and a hCMV-MIE promoter combination were each transfected at total DNA loads ranging from 100 to 800 ng. The DNA quantities transfected, and the promoter variants used simulate the range of expression potentially observed during quantify MGEV performance.

The quantity of DNA transfected ranged from 100 to 800 ng, to test the full range of potential expression observed when three reporter genes are co-expressed within the host cell. This was to simulate the multi-gene expression performance potentially observed within a MGEV. Therefore, the dynamic range of detection by qPCR and flow cytometry was assessed using these transfections conditions. The following section will expand on testing of qPCR and flow cytometry as a means of quantification.

4.2.3.1. Development of a qPCR assay to quantify transcriptional activity

A method for performing qPCR specifically for CHO cell expression was previously developed and published by Dr. Adam J. Brown which included an optimised protocol and identification of reference genes (Brown et al., 2018). The same method was used for the quantification of *eGFP*, *mCherry* and *tagBFP* mRNA after designing efficient primers.

4.2.3.1.1. Designing and testing primer efficiency for qPCR

The initial step was to design primers to target and amplify amplicons representing mRNA transcripts of *eGFP*, *mCherry*, *tagBFP* and *glutamine synthetase (GS)* genes. The primers were designed using the web-tool Primer-BLAST by National Centre for Biotechnology Information (NCBI) which recommend a series of primer pairs based on the provided coding DNA sequence (CDS), all of which did not interact with the CHO genome. Apart from *GS*, where Primer-BLAST could not suggest a primer pair that could differentiate between the recombinant and endogenous *GS* gene. Table 4-4 shows the final list of primers selected for testing and quantifying transcription.

Table 4-4: Primers designed and used for real-time quantitative PCR of *glutamine synthetase*, *eGFP*, *mCherry* and *tagBFP* genes.

The forward (FW) and reverse (RV) primers described below were designed, tested and used for the quantification of mRNA copies of transiently expressed *GS*, *eGFP*, *mCherry* and *tagBFP* in MedI-CHO.

Targeted Gene	Primer Name	Primer Sequence (5' - 3')	Melting Point (°C)	GC Content (%)	Amplicon Size (bp)
<i>Glutamine Synthetase</i>	GS-FW-1	TTTCTGCTGGTGTCCG CCAAT	57.5	50.0	153
	GS-RV-1	CATTGAGAAGGCATG TGCGG	56.9	55.0	
<i>eGFP</i>	eGFP-FW-5	ACAAGACCAGAGCCG AAGTG	57.2	55.0	157
	eGFP-RV-5	TTCTGCTTGTCGGCCATG AT	57.1	50.0	
<i>mCherry</i>	mCherry-FW-2	CCAGTTTATGTACGGCTC CAA	54.8	47.6	106
	mCherry-RV-2	GTTTCATCACTCTCTCCCA CTTG	55.1	50.0	
<i>tagBFP</i>	tagBFP-FW-5	CACCTCCTTTCTGTACGG CT	56.8	55.0	305
	tagBFP-RV-5	CCATGTCGTTTCTGCCTT CC	56.3	55.0	

The efficiencies of the primer sets were tested to verify their functionality. This was achieved by first generating cDNA derived from 24h transient expression of a MGEV with a simian virus 40 (SV40) promoter regulating *GS* expression and a hCMV-MIE promoter expressing *eGFP*, *mCherry* and *tagBFP*. The stock cDNA was diluted 10-fold followed by a 3-fold serial dilution for testing the *GS* primers, whereas the stock cDNA was diluted 3-fold followed by a 4-fold serial dilution for testing the *eGFP*, *mCherry* and *tagBFP* primers. The serial dilution was performed ranging from 4- 6 points to generate a linear curve. The cycle threshold (Ct) was plotted against the log of cDNA copies as shown in Figure 4-6.

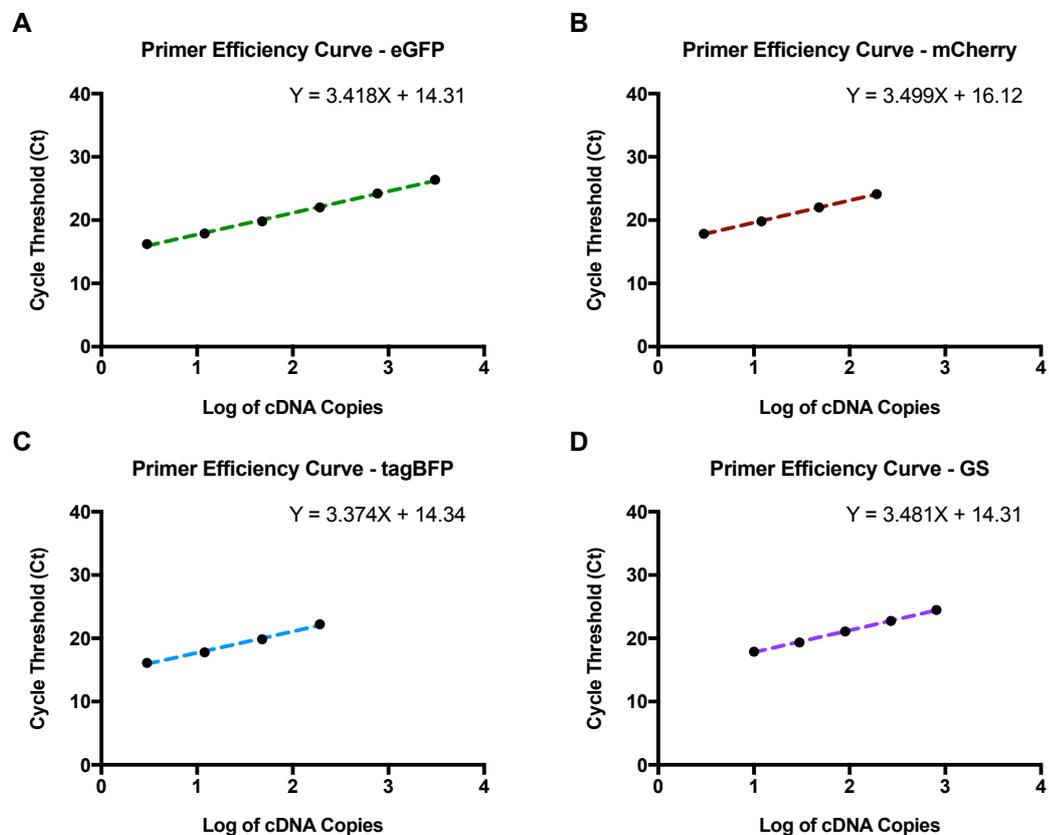


Figure 4-6: Testing primer efficiency for quantitative PCR.

The designed primer pairs for *eGFP*, *mCherry*, *tagBFP* and *GS* were tested to determine their efficiency for qPCR. This was achieved initially diluting cDNA from transiently expressing MedI-CHO in various serial dilution for the fluorescent proteins and *GS* over 4-6 points. The cycle threshold (Ct) was plotted against log of cDNA copies for *eGFP* (A), *mCherry* (B), *tagBFP* (C) and *GS* (D). A linear fit curve was calculated to determine the gradient and in turn the primer efficiency.

The data indicated the primers were successfully annealing to the cDNA encoding for *eGFP*, *mCherry*, *tagBFP* and *GS*. Furthermore, the primers displayed no cross reactivity between fluorescent proteins with either a very high or undetermined Ct

value (data not shown). The *GS* primer was suggested to bind to both recombinant and endogenous transcripts of *GS*, therefore a mock transfected cell was required to normalise for endogenous expression and accurately quantify recombinant *GS* expression. The relationship between Ct and log of cDNA copies was linear for all primer sets with an $r^2 > 0.99$. The primer efficiency was determined by calculating the gradient of the trend line and using ThermoFisher Scientific's qPCR primer efficiency calculator web tool. The efficiency percentages were 96.1%, 93.1%, 97.9% and 93.8% for *eGFP*, *mCherry*, *tagBFP* and *GS* respectively, and were deemed sufficient for quantification.

4.2.3.1.2. Determining the dynamic range and accuracy of transcriptional activity quantification by plasmid DNA titration

The dynamic range of detection by qPCR was investigated to assess if qPCR is a viable technique for quantification. This was determined by co-expressing *eGFP*, *mCherry* and *tagBFP* regulated by a low, medium, high strength synthetic promoter or a hCMV-MIE promoter at an equimolar ratio in Medi-CHO as shown in Figure 4-5. The quantity of DNA transfected ranged from 100 to 800 ng, therefore, testing the full range of potential expression in order to quantify accurate gene expression stoichiometric ratios in MGEVs. After 24h, the cDNA samples were prepared by performing RNA extraction, removing genomic DNA and reverse transcribing the RNA from transfected cells. The qPCR was performed on the samples using the designed primers as described in section 4.2.3.1.1 and the Ct was measured by the qPCR machine. The *mmadhc* and *fkbp1a* reference genes were also quantified to normalise for any variation in cDNA concentration and accurately measure differences in transcriptional activity. The expression fold change was calculated using a double delta Ct analysis by designating 100 ng of transfected DNA as the control. The fold change was plotted against the quantity of DNA transfected as shown in Figure 4-7.

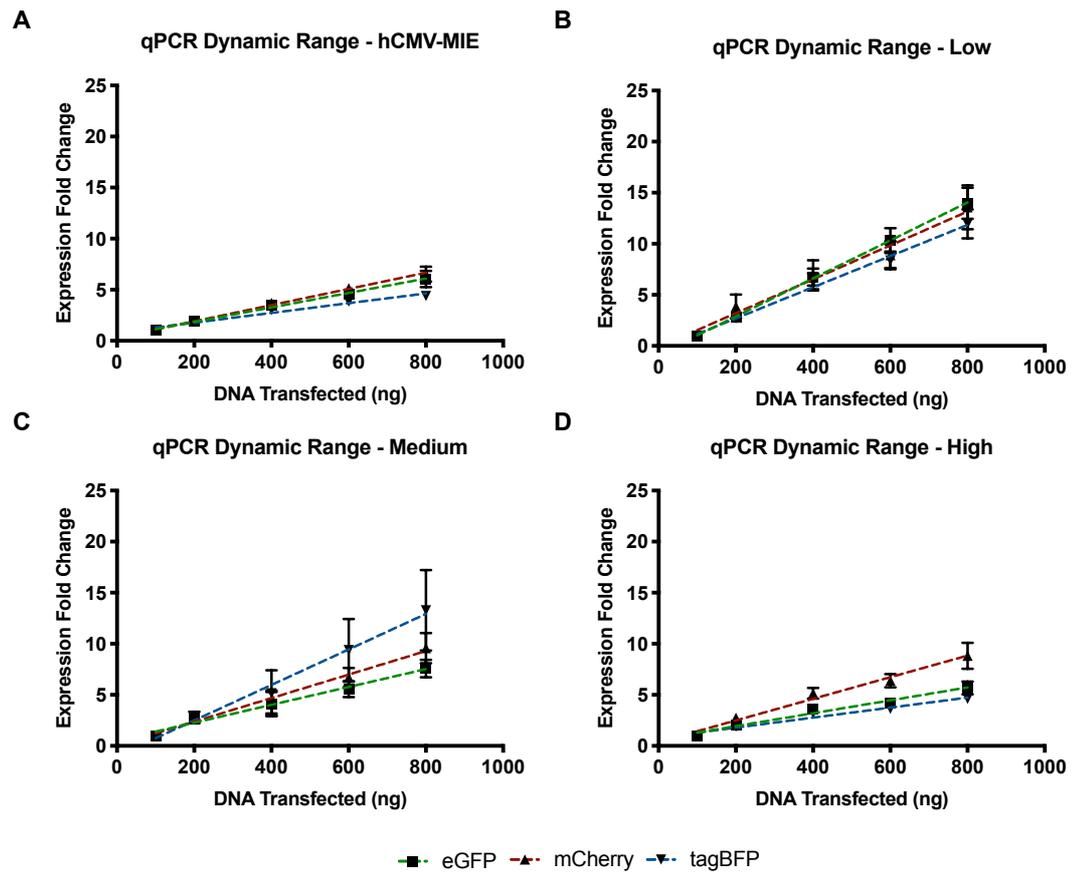


Figure 4-7: Investigating the dynamic range of qPCR-mediated detection by titrating plasmid DNA.

The dynamic range of qPCR detection was measured by titrating 100 to 800 ng of total plasmid DNA which comprised of equimolar *eGFP*, *mCherry* and *tagBFP* plasmids driven by either a hCMV-MIE (A) promoter, or a low (B), medium (C) or high (D) strength synthetic promoter. The relative fold change was plotted against the quantity of DNA to measure the linear range. The green, red and blue dotted lines represent the line of best-fit calculated by linear regression of *eGFP*, *mCherry* and *tagBFP* respectively and the error bars represent the SEM of biological triplicates.

Figure 4-7 shows the expression fold change against the DNA transfected was similar for all three fluorescent proteins when regulated by the same strength promoter. Furthermore, the relationship is linear for all three fluorescent proteins with the r^2 ranging from 0.96 to 0.99. However, the fold change between 100 ng and 800 ng of transfected DNA varies depending on the transcriptional activity of the promoter as shown in Table 4-5.

Table 4-5: Average expression fold change by qPCR between 100 and 800 ng of transfected DNA. The average expression fold changes were quantified by qPCR for *eGFP*, *mCherry* and *tagBFP* expression when comparing 100 ng and 800 ng of transfected DNA.

Promoter Variant	Average expression fold change between 100 and 800 ng of transfected DNA		
	<i>eGFP</i>	<i>mCherry</i>	<i>tagBFP</i>
Low	14.0	13.6	11.9
Medium	7.6	9.6	13.3
High	5.7	8.8	4.6
hCMV-MIE	6.1	6.5	4.4

The data shows a decrease in fold change as the transcriptional activity of the promoter was increased as shown by low, medium and high strength synthetic promoters. The higher fold change between 100 and 800 ng of transfected DNA when regulated by a low strength promoter could be caused by the poorer detection at the lowest DNA copies, therefore skewing the relative fold changes. Moreover, the different average expression fold changes between fluorescent reporter suggests variable relative levels of transcripts were detected for the same DNA load transfected. This could also be a consequence of variable mRNA dynamics influenced by the CDS of *eGFP*, *mCherry* and *tagBFP* as the 5' and 3' UTRs were constant within the TU (Ross, 1995; Wang et al., 2017).

The goal of the assay was to accurately quantify titrated levels of gene expression at a specific transcriptional strength. Therefore, investigating the difference in relative mRNA copies of *eGFP*, *mCherry* and *tagBFP* when regulated by a low, medium and high strength synthetic promoter, and a hCMV-MIE promoter were compared. This was achieved by performing a double delta Ct analysis relative to the high strength promoter and adjusted for difference in cDNA by normalising with two housekeeping genes (*mmadhc* and *fkbp1a*). Therefore, the relative promoter activity was determined by the difference in Ct values when compared against the high strength synthetic promoter for each fluorescent reporter at every transfected DNA load. The double delta Ct fold change values were converted to percentages to represent relative promoter activity. Figure 4-8 shows the relative promoter activity for each fluorescent protein when co-expressed across the different DNA loads from Figure 4-7.

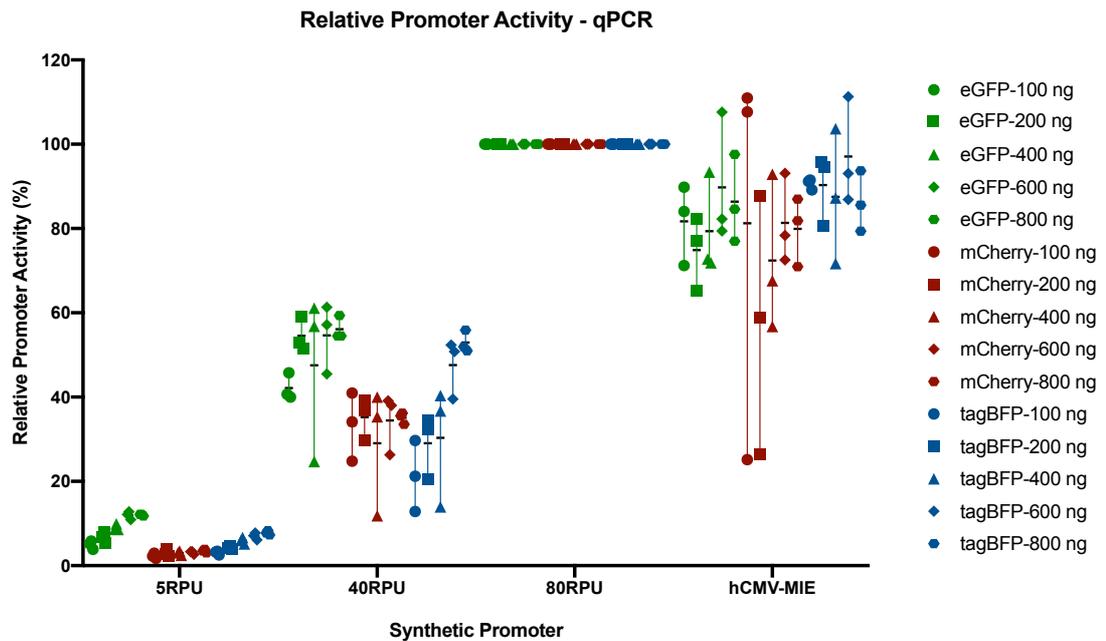


Figure 4-8: Relative promoter activity of *eGFP*, *mCherry* and *tagBFP* by qPCR analysis.

Graph showing the relative promoter activity when compared against the high strength promoter by double delta Ct analysis. The green, red and blue points represent *eGFP*, *mCherry* and *tagBFP* expression. The circle, square, triangle, diamond and hexagon shape points represent transfected DNA loads of 100, 200, 400, 600 and 800 ng respectively. The horizontal black line and vertical coloured lines represents the mean and range of relative promoter activity of biological triplicates at each condition.

The data shows that the difference in promoter strength was observed between the low, medium and high strength synthetic promoters when co-expressing *eGFP*, *mCherry* and *tagBFP* at different quantities of transfected DNA (refer to Figure 4-5). The largest variation in transcription was observed by hCMV-MIE with a standard deviation of 11.1%, 25.2% and 9.6% for *eGFP*, *mCherry* and *tagBFP* respectively. The medium strength synthetic promoter showed the next largest variation in transcription (most variation of the synthetic promoter variants) with a standard deviation of 10.1%, 7.7% and 14.5% for *eGFP*, *mCherry* and *tagBFP* respectively. The larger variability in hCMV-MIE transcription compared to the synthetic promoters could be a consequence of the less-defined nature and functionality of the viral promoter (Brown et al., 2014). Interestingly, the relative promoter activity of the low synthetic promoter increased when the DNA load transfected was titrated from 100 to 800 ng by 7.1%, and 4.7% for *eGFP* and *tagBFP* respectively, whereas, a minimal difference (1.2%) was observed for *mCherry*. This variation could be a consequence of poorer detection of transcripts at lower expression levels resulting from fewer DNA copies transfected. Moreover, no variation was observed by the high strength synthetic promoter since it was normalised to 100%. This was a by-product of using the double delta Ct analysis

methodology where differences in Ct values are measured relative to the high strength promoter for each replicate individually (as each replicate was quantified separately) to yield relative promoter activity of the low and medium strength synthetic promoters and hCMV-MIE promoter. The mean relative promoter activity across all DNA loads transfected was calculated and collated in Table 4-6.

Table 4-6: Average relative promoter activity from qPCR analysis.

The table shows the average relative promoter activity for a low, medium and high strength synthetic promoter, and a hCMV-MIE promoter for *eGFP*, *mCherry* and *tagBFP* when quantified by qPCR and calculated by double delta Ct analysis against the high strength synthetic promoter.

Promoter Variant	Average Relative Promoter Activity per Fluorescent Protein (\pm Standard Deviation) (%)		
	<i>eGFP</i>	<i>mCherry</i>	<i>tagBFP</i>
Low	8.94 (\pm 3.01)	2.97 (\pm 0.63)	5.61 (\pm 1.88)
Medium	51.01 (\pm 10.09)	33.43 (\pm 7.66)	36.23 (\pm 14.46)
High	100.00 (\pm 0.00)	100.00 (\pm 0.00)	100.00 (\pm 3.01)
hCMV-MIE	82.41 (\pm 11.14)	74.54 (\pm 25.19)	90.36 (\pm 9.61)

The expected trend for relative promoter activity was approximately 5%, 40% and 100% for the low, medium and high strength synthetic promoters respectively. As shown in Table 4-6, the relative promoter activity for *tagBFP* transcription was closest to predicted with *mCherry* similar as well. However, *eGFP* showed higher levels of relative promoter activity for the low and medium promoter when compared against the high strength promoter. This could be due to sequence specific mRNA dynamics and post-transcriptional activity affecting mRNA levels (Lewis et al., 2017). Overall, qPCR has a good dynamic range for quantifying transcriptional activity as a function of gene expression. Furthermore, a defined difference in low, medium and high level of expression can be quantified by using synthetic promoters. In conclusion, qPCR is a viable method of measuring and quantifying titration in gene expression.

4.2.3.2. Determining the dynamic range and accuracy of fluorescent protein expression by DNA titration

Similar to qPCR, the dynamic range of detection of the flow cytometer was investigated to assess if it is a viable technique for quantifying gene expression. This was determined by co-expressing *eGFP*, *mCherry* and *tagBFP* at an equimolar ratio when regulated by a low, medium, high strength synthetic promoter or a hCMV-MIE promoter as shown in Figure 4-5. The equimolar reporter DNA was transfected in MedI-CHO at a range of 100 to 800 ng. In turn, spanning the full range of potential expression to quantify gene expression stoichiometric ratios within a MGEV. After 24h, the cells were measured on the Attune NxT flow cytometer according the optimised settings described in Table 4-2. The transfected cells were gated for viable cells, followed by single cells excluding doublets and accounted for cellular autofluorescence under VL-1, BL-1 and YL-2 channels using mock transfected cells. Therefore, identifying cells positive for expressing *eGFP*, *mCherry* and *tagBFP*.

The range of iMFI for each fluorescent protein is different with higher fluorescence observed by *tagBFP* compared to *eGFP* and *mCherry*. This is a by-product of differing fluorescent brightness that is dependent on the excitation of the protein, the emission filter and the PMT voltage strength. Furthermore, there is a substantial difference in iMFI detection between a low and a high transcriptionally active synthetic promoter. At 800 ng, the iMFI quantified of the low strength promoter ranges between 5675.2 and 13303.7, whereas a high strength promoter showed fluorescence between 146736.6 and 308640.1. The fluorescence of the medium strength synthetic promoter and the hCMV-MIE promoter were in-between the two ranges as shown in Appendix B. The expression fold change relative to iMFI detected at 100 ng of transfected DNA was calculated and plotted against the quantity of DNA transfected as shown in Figure 4-9.

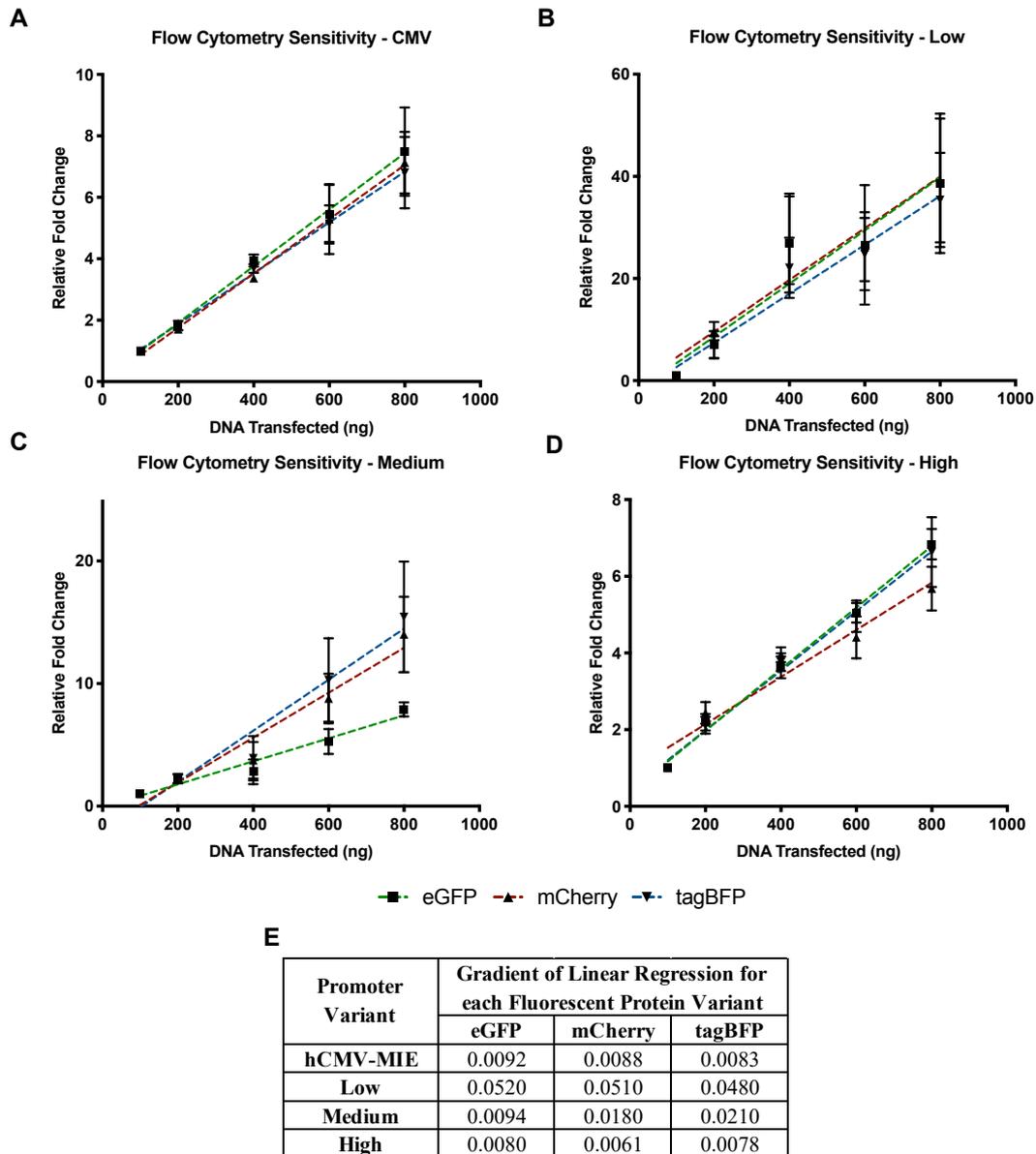


Figure 4-9: Determining dynamic range and sensitivity of flow cytometer-mediated detection by titrating plasmid DNA.

The dynamic range of flow cytometer-mediated detection was measured by titrating 100 to 800 ng of total plasmid DNA which comprised of equimolar *eGFP*, *mCherry* and *tagBFP* plasmids driven by either a hCMV-MIE (A) promoter, or a low (B), medium (C) or high (D) strength synthetic promoter. The fold change in iMFI was derived by normalising against the iMFI detected at 100 ng for each fluorescent protein and plotted against the quantity of DNA transfected to measure the linear range. The green, red and blue dotted lines represent the line of best-fit calculated by linear regression of *eGFP*, *mCherry* and *tagBFP* respectively and the error bars represent the SEM of biological triplicates. The table (E) states the gradients of each line of best-fit derived from linear regression to assess the relationship between DNA transfected and fluorescent reporter expression quantified by flow cytometry.

The dynamic range was determined by the linearity of the curves based on the r^2 values, which ranges from 0.91 to 0.99. As shown in Figure 4-9, the poorest linearity is exhibited by the low strength promoter with the average r^2 of 0.93 for all three reporters, whereas the high strength promoter displayed an average r^2 of 0.98. The

medium strength promoter was in-between with the average r^2 of 0.96. This indicates that the fluorescent protein expression at different gene doses regulated by varying levels of transcriptional strength is within the dynamic range. However, a higher relative fold change was observed by the low and medium strength promoters when compared to the hCMV-MIE and high strength promoters. The expected increase should be 8-fold based on the quantity of DNA transfected (assuming 100% transfection efficiency). Therefore, poor detection of fluorescence at low DNA copies transfected would skew the relative fold change and deviate from the expected. The gradients of each line of best-fit derived from linear regression were analysed to depict the relationship between DNA transfected and relative fluorescent protein detection to discern accuracy of quantification (Figure 4-9E). For accurate relative quantification, the gradient expected was 0.01, whereas the average gradient observed deviated by 500% for the low strength synthetic promoter-mediated expression. Similarly, the observed gradient deviated by 61.3% from the expected when utilising the medium strength synthetic promoter. Conversely, smaller deviations were observed for the hCMV-MIE (12.3%) and high strength synthetic promoter (27%). This indicates poor sensitivity of fluorescent detection was observed at low levels of gene expression dictated by weaker strength promoters.

The goal of this assay was to quantify titrated levels of gene expression as dictated by fluorescent intensity of the reporters. Therefore, investigating the difference in iMFI of *eGFP*, *mCherry* and *tagBFP* when regulated by a low, medium and high strength synthetic promoter, and a hCMV-MIE promoter were compared. This was achieved by calculating the percentage difference against the average iMFI of the respective fluorescent proteins regulated by a high strength promoter and labelled as relative promoter activity. Figure 4-10 shows the relative promoter activity for each fluorescent protein when co-expressed across the different DNA loads from Figure 4-9.

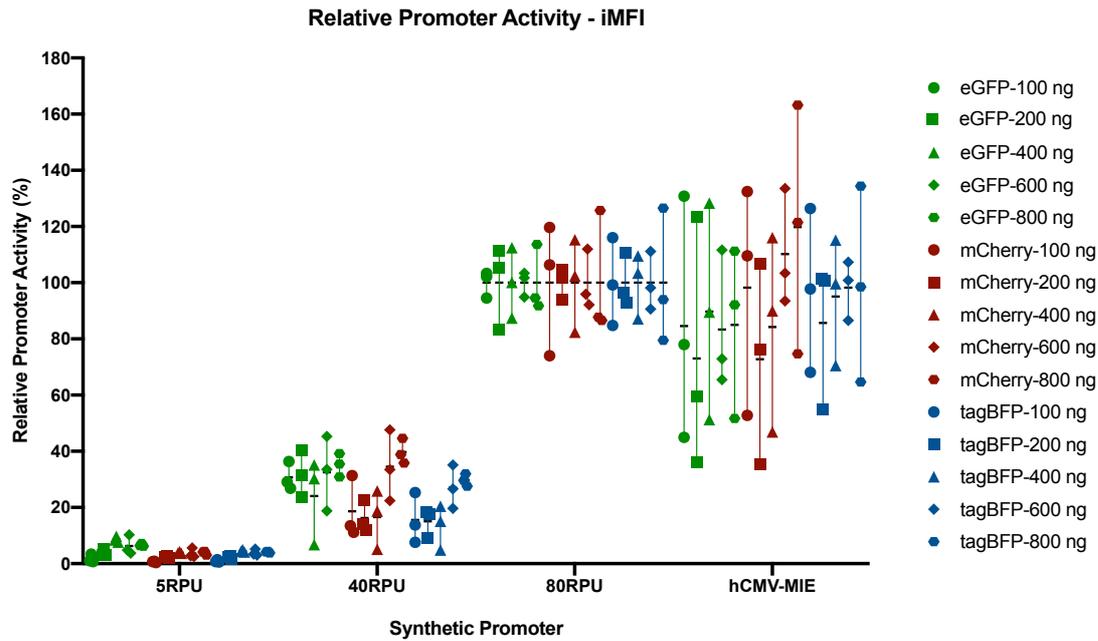


Figure 4-10: Relative promoter activity of *eGFP*, *mCherry* and *tagBFP* by flow cytometry analysis. Graph showing the relative promoter activity when compared against the average iMFI for high strength promoter expression of the respective fluorescent proteins. The green, red and blue symbols represent *eGFP*, *mCherry* and *tagBFP* expression. The circle, square, triangle, diamond and hexagon shape points represent transfected DNA loads of 100, 200, 400, 600 and 800 ng respectively. The horizontal black line and vertical coloured lines represents the mean and range of relative promoter activity of biological triplicates at each condition.

The data shows that a difference in promoter strength is observed between the low, medium and high strength synthetic promoters when co-expressing *eGFP*, *mCherry* and *tagBFP* at different transfected DNA quantities (as described in Figure 4-5). The largest variation in protein expression was observed by hCMV-MIE with a standard deviation of 32.0%, 35.3% and 22.6% for *eGFP*, *mCherry* and *tagBFP* respectively. This observation was consistent with the qPCR data (Figure 4-8) and indicates the viral promoter has more variable characteristics. The mean relative promoter activity was calculated and collated in Table 4-7.

Table 4-7: Average relative promoter activity from flow cytometry analysis.

The table shows the average relative promoter activity for a low, medium and high strength synthetic promoter, and a hCMV-MIE promoter for *eGFP*, *mCherry* and *tagBFP*. This was achieved by calculating percentage difference against the average iMFI for each fluorescent protein regulated by a high strength synthetic promoter.

Promoter Variant	Average Relative Promoter Activity per Fluorescent Protein (\pm Standard Deviation) (%)		
	<i>eGFP</i>	<i>mCherry</i>	<i>tagBFP</i>
Low	5.39 (\pm 2.94)	2.89 (\pm 1.49)	3.14 (\pm 1.50)
Medium	30.89 (\pm 9.43)	25.15 (\pm 13.04)	20.20 (\pm 9.12)
High	100.00 (\pm 8.92)	100.00 (\pm 14.45)	100.00 (\pm 12.86)
hCMV-MIE	83.13 (\pm 32.00)	97.04 (\pm 35.35)	95.13 (\pm 22.61)

Similar to qPCR analysis, the expected trend for relative promoter activity was approximately 5%, 40% and 100% for the low, medium and high strength synthetic promoters respectively. As shown in Table 4-7, the relative promoter activity for the low and medium strength promoters were lower compared to the qPCR data. Furthermore, the relative promoter activity percentages were lower than the expected values apart from the low strength promoter driving *eGFP* expression. Therefore, the data indicated that poorer accuracy is observed for quantification of *eGFP*, *mCherry* and *tagBFP* gene expression when regulated by a low or medium strength synthetic promoter different transfected DNA quantities (when co-transfected as separate SGVs at equimolar ratios as described in Figure 4-5).

Overall, the flow cytometer has a sufficient dynamic range to quantify *eGFP*, *mCherry* and *tagBFP* expression when regulated by the different promoter variants. However, the sensitivity of detection at low DNA copies was poor for the low and medium strength promoters. Moreover, the difference in observed and expected relative promoter activity was larger, reiterating poorer accuracy of detection.

4.2.3.3. Comparing gene expression quantification accuracy of qPCR and flowcytometry

To compare the quantitative accuracy of qPCR and flow cytometry, the relative fold change of expression (derived by comparing against expression at 100 ng of transfected DNA) detected by qPCR and flow cytometry of *eGFP*, *mCherry* and *tagBFP* during co-expression (as stated in Figure 4-5) was compared by linear regression for each promoter variant as shown in Figure 4-11.

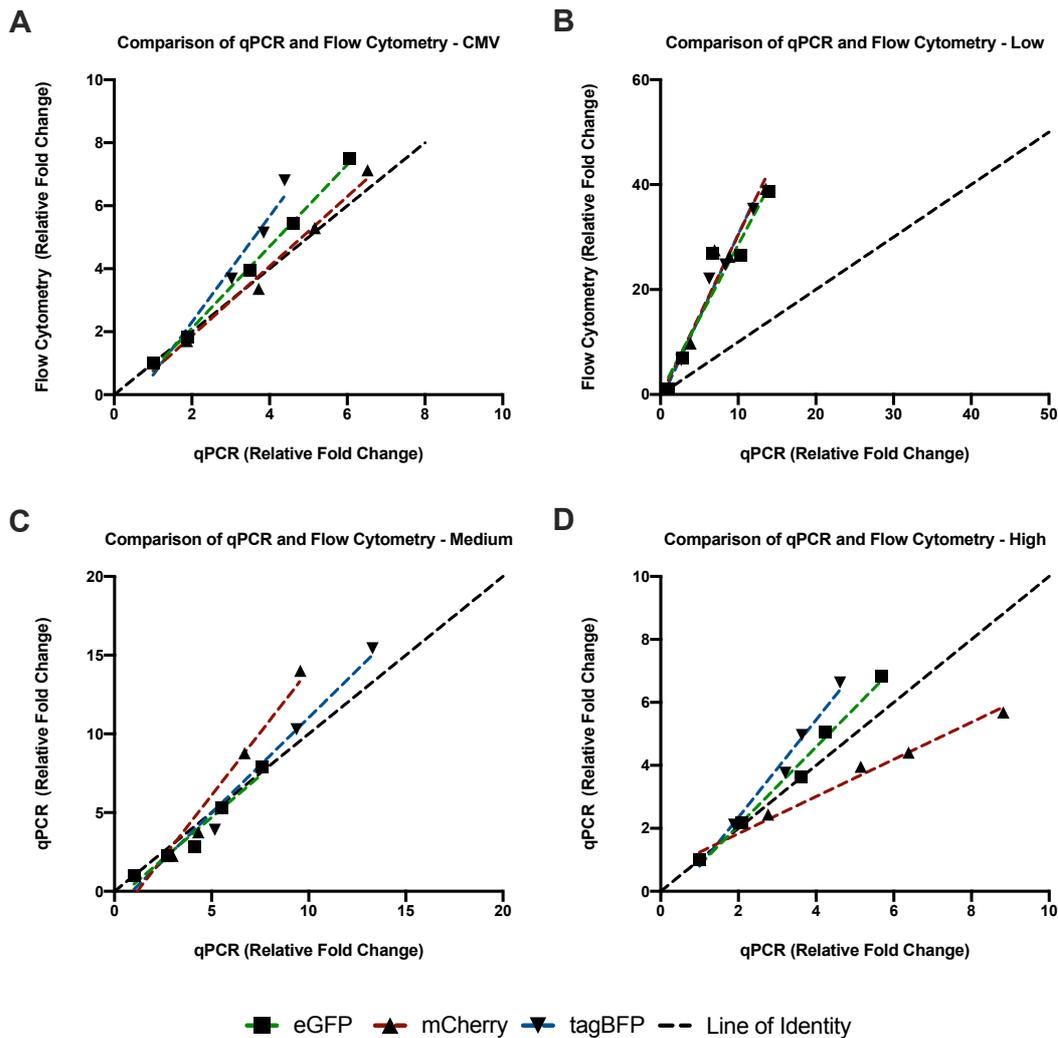


Figure 4-11: Comparing relative fold change of qPCR and flow cytometry-mediated detection. The fold change in expression of *eGFP*, *mCherry* and *tagBFP* was normalised to expression detected at 100 ng of transfected DNA by qPCR and flow cytometry for the respective reporter proteins using different promoter strengths. The relative fold changes derived by qPCR and flow cytometry for each protein were plotted against each other and linear regression analysis was performed to compare detection by both methodologies. The green, red, and blue dotted lines represent the line of best-fit for *eGFP*, *mCherry* and *tagBFP* expression, whereas the black dotted line represents the line of identity ($y = x$) as reference for direct correlation.

The line of identity ($y = x$) in Figure 4-11 represents perfect correlation in detection between qPCR and flow cytometry and the expected fold change in expression when titrated by transfected DNA loads (100 to 800 ng). The low strength synthetic promoter-mediated expression was substantially skewed from the expected fold change by the deviation of the gradient (3-fold) from the line of identity. Therefore, the comparison reiterates that the flow cytometer has poor fluorescent detection sensitivity at low expression levels for all three reporter proteins compared to detection by qPCR. The other substantially deviated slope from the line of identity was *mCherry* driven by a high strength synthetic promoter, where a 1.6-fold variation was observed.

This indicates that there is a potential limitation of fluorescent detection of *mCherry* at high expression levels, as shown by the lower iMFI expression fold changes than the expected (particularly at higher transfected DNA loads). In conclusion, the qPCR generally quantified gene expression levels more accurately compared to the flow cytometer. Therefore, the flow cytometer can provide semi-quantitative analysis of fluorescent protein gene expression, whereas the qPCR provides more accurate absolute quantification of gene expression when dictated by promoters of different transcriptional activity.

4.2.4. Characterising combinatorial functionality of synthetic promoters

The synthetic promoter library has been validated to show predictable expression at a range of transcriptional strengths by testing each promoter variant individually by *SEAP* expression (Brown et al., 2017). However, no characterisation had been performed to test the co-functionality of the multiple synthetic promoters concurrently. Therefore, the aim was to investigate functionality of synthetic promoter combinations, in order to determine if predictable gene expression stoichiometric ratios can be achieved in a MGEV. The co-functionality was measured by transiently expressing *eGFP*, *mCherry* and *tagBFP* using the HT transient transfection protocol. To investigate co-functionality of the same promoter variant when controlling multiple genes, the low, medium and high strength synthetic promoter's expression activity was tested by utilising the same promoter species to control expression of three fluorescent protein reporter at equimolar ratios simultaneously. Similarly, synergistic functionality of different synthetic promoter variants controlling expression of multiple genes simultaneously was determined using the same approach.

Promoter self-interaction investigates expression activity when the same synthetic promoter variant is used to control multiple genes simultaneously. This was measured by increasing the number of promoter units in increments of 200 ng of a separate SGV encoding for one of three fluorescent proteins eventually leading to a final quantity of 600 ng transfected DNA comprising of three different plasmid variants as shown in Figure 4-12A. For example, 600 ng of transfected DNA comprised of 200 ng of pLow-eGFP, pLow-mCherry and pLow-tagBFP each regulated by the low strength synthetic promoter. The promoter self-interaction was quantified as relative fold change against singular expression of each fluorescent protein at 200 ng of transfected DNA in Medi-

CHO (and non-coding plasmid DNA was supplemented to normalise the quantity of transfected DNA to 600 ng). The results of the promoter self-interaction were plotted as a bar chart as shown in Figure 4-12B.

A

Promoter Units (represented as separate single gene vectors either singularly or co-transfected)					
Promoter Strength	1			2	3
Low	eGFP	mCherry	tagBFP	eGFP + mCherry	eGFP + mCherry + tagBFP
Medium	eGFP	mCherry	tagBFP	eGFP + mCherry	eGFP + mCherry + tagBFP
High	eGFP	mCherry	tagBFP	eGFP + mCherry	eGFP + mCherry + tagBFP

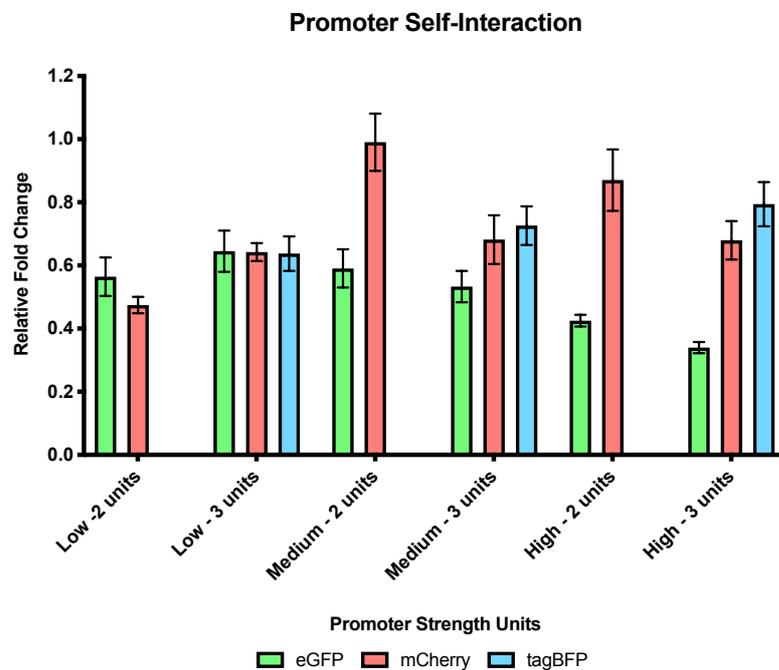
B

Figure 4-12: Measuring synthetic promoter self-interaction by flow cytometry.

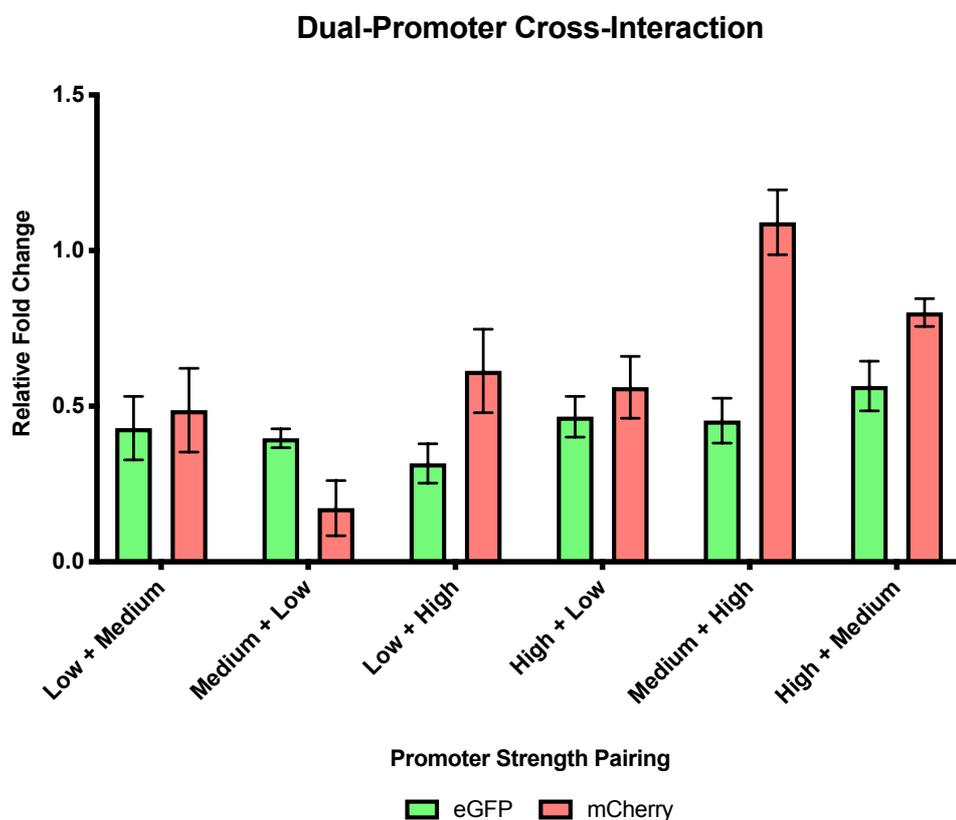
(A) A table summarising the experimental outline for quantifying synthetic promoter activity when using the same promoter variant to controlling expression of multiple fluorescent reporter genes simultaneously. Each row represents the synthetic promoter variant tested and each column represents the number of promoter units (represented as 200 ng intervals of a separate SGV encoding for either *eGFP*, *mCherry* or *tagBFP*). Every cell within the table represents a transfection condition, where one promoter unit represents either 200 ng of either *eGFP*, *mCherry* or *tagBFP* expressed utilising a synthetic promoter variant. When two or three promoter units are tested, then equimolar ratios of two or three separate plasmids encoding for the fluorescent reporters are transfected simultaneously. This was measured using transient expression in Medi-CHO cells. (B) The low, medium and high strength synthetic promoters' self-functionality was tested by quantifying the iMFI of *eGFP*, *mCherry* and *tagBFP* according to the combinations described in (A). A relative fold change was calculated against the single unit expression of each fluorescent protein at the same gene copies. Each bar represents the average relative fold change and the error bars represent SEM of biological triplicates.

The screen of doubling and tripling the same synthetic promoter variant units, as shown in Figure 4-12B, showed an overall reduction in fluorescent protein expression by $\leq 59\%$. The average fold changes in expression using two low strength synthetic promoter units expressing *eGFP* and *mCherry* was 0.56 and 0.47 respectively. Whereas, three units of the same promoter showed average relative fold changes of 0.65, 0.64 and 0.64 for *eGFP*, *mCherry* and *tagBFP* respectively. The marginal increase in fold change observed (average of 13%) between using two and three units suggests minimal difference in promoter interaction is observed when increasing the low strength synthetic promoter units from two to three.

The average relative fold change of *eGFP* and *mCherry* utilising two units of the medium strength synthetic promoter concurrently was 0.59 and 0.99 respectively. However, when three units of the medium strength promoter were used simultaneously, the average relative fold changes of *eGFP*, *mCherry* and *tagBFP* were 0.53, 0.68 and 0.73 respectively. A minimal difference in fold change was observed for *eGFP* but a 31% reduction was measured for *mCherry*. Similarly, two units of the high strength synthetic promoter employed to express *eGFP* and *mCherry* showed a relative fold change of 0.43 and 0.87 respectively. Whilst, three units of the high strength promoter dictating *eGFP*, *mCherry* and *tagBFP* expression demonstrated an average fold change of 0.34, 0.68 and 0.79 respectively. Therefore, the variable reduction in average fold change observed between two and three units of the medium and high strength synthetic promoters suggests an additive repression of recombinant gene expression suggesting some promoter self-interference.

Overall, it is hypothesised the decrease in gene expression when increasing promoter units to express multiple recombinants simultaneously may be caused by promoter squelching. Promoter squelching is defined as the repression in promoter activity caused by either competition of TFs and cofactors or interaction of different TFs, both of which can hinder transcription initiation (Cahill et al., 1994; Huliák et al., 2012). However, the repression could also be a consequence of metabolic burden of expressing more than one fluorescent protein concurrently.

After determining and quantifying the self-interaction, the same process was performed to investigate the behaviour of pairing synthetic promoters with different transcriptional activity to control two genes simultaneously. This was performed by measuring fold change of equimolar transiently co-expressed *eGFP* and *mCherry* for different promoter pairs relative to single promoter variant expression for the respective reporters at identical DNA copies (with supplemented non-coding plasmid DNA to maintain constant DNA load during transfection). The relative fold change for each promoter pair was plotted in a bar chart as shown in Figure 4-13.



Synthetic Promoter Combination	Plasmid Combination	
Low + Medium	pLow-eGFP	pMedium-mCherry
Medium + Low	pMedium-eGFP	pLow-mCherry
Low + High	pLow-eGFP	pHigh-mCherry
High + Low	pHigh-eGFP	pLow-mCherry
Medium + High	pMedium-eGFP	pHigh-mCherry
High + Medium	pHigh-eGFP	pMedium-mCherry

Figure 4-13: Measuring dual synthetic promoter cross-interaction by flow cytometry.

The low, medium and high strength synthetic promoters were investigated to measure cross-interaction by pairing each variant together. This was achieved by co-expressing *eGFP* and *mCherry* with different strength synthetic promoter pairs at an equimolar ratio. The relative fold change was calculated against the single unit expression of each fluorescent protein at the same gene copies. The error bars represent SEM of biological triplicates. The table below summarises the plasmid combination and specifies the synthetic promoter variant utilised to control expression of either *eGFP* or *mCherry* as separate SGVs.

A reduction in fluorescent protein expression was observed for every promoter pair by various amounts with the highest reduction observed at 82.7%. The low strength promoter expression activity was reduced between 0.17- and 0.43- fold and the medium strength promoter expression activity was reduced between 0.40- and 0.49- fold when co-expressed. The low and high promoter expression activity pairing demonstrated fold changes between 0.32- and 0.56- fold and 0.47- and 0.61- fold respectively. Finally, the medium and high expression activity pairing displayed fold changes between 0.45- and 0.80- fold, and 0.57- to 1.10- fold change respectively. The most reduction in expression was observed between the low and medium promoter pair, followed by the low and high promoter pair, with the least reduction between the medium and high promoter pair. The data indicates the presence of cross-interaction coupled with metabolic burden (for expressing two fluorescent proteins simultaneously) between promoter variants.

The aim of the synthetic promoters tested was to be integrated into a MGEV to titrate gene expression of three recombinant genes. Therefore, verifying the performance of the promoter variants working in conjunction would assist in accurately predicting expression. This was achieved by mixing and matching the promoters with low, medium and high transcriptional activity co-expressing *eGFP*, *mCherry* and *tagBFP*, and measuring the transient expression in MedI-CHO. The relative fold change was calculated against single fluorescent protein expression using the respective promoter variants at identical gene copies and plotted in a bar chart as shown in Figure 4-14.

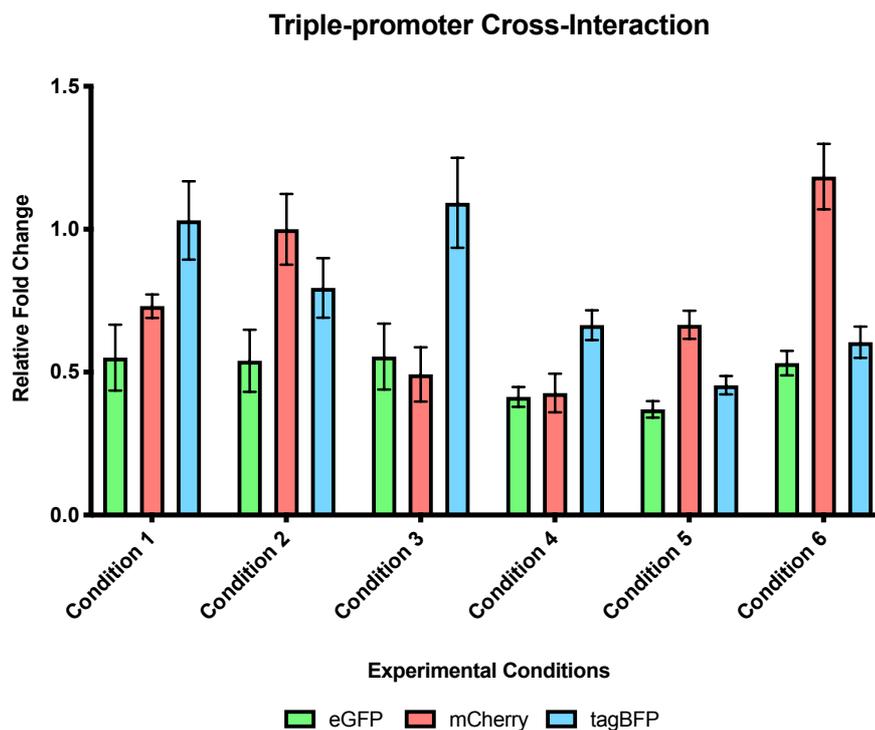


Figure 4-14: Measuring triple synthetic promoter cross-interaction by flow cytometry.

The low, medium and high strength synthetic promoters were investigated to measure cross-interaction when combining different variants together. This was achieved by co-transfecting different strength synthetic promoter combinations controlling simultaneous expression of *eGFP*, *mCherry* and *tagBFP* on separate SGVs at an equimolar ratio within Medi-CHO. The relative fold change was calculated against the single unit expression of each fluorescent protein at identical gene copies. The error bars represent SEM of biological triplicates. The table summarises the combination of fluorescent proteins and synthetic promoters used in each condition.

The promoter cross-interaction during simultaneous expression of the synthetic promoters with different transcriptional strengths demonstrated a larger reduction in expression of the low and medium strength promoters. The high strength promoter showed minimal effect in expression of *mCherry* and *tagBFP* but a more substantial impact on *eGFP* expression. The average fold change of expression across all three fluorescent protein regulated by the same strength synthetic promoter from conditions 1 to 6 were calculated. The expression by the low strength promoter was 0.51-fold, the medium strength promoter was 0.66-fold and the high strength promoter was 0.85-fold, reinforcing the initial observation. The *eGFP* expression regulated by the high strength promoter was lower than *mCherry* and *tagBFP* suggesting a deviation in

expression. This could be a consequence of the potential undesired behaviour of *eGFP* during intracellular accumulation observed previously in section 4.2.2. The data suggests the effect of cross-interaction and metabolic burden is different when three different promoter variants are expressing recombinant genes within the cell simultaneously. Furthermore, the variation in expression fold change during self- and cross-interaction analysis could also be a consequence of lower accuracy by flow cytometry-based quantification, as described in section 4.2.3.2 and 4.2.3.3. Therefore, the observed trend is valid but an orthogonal means to accurately quantify promoter interactions and squelching is required.

4.2.5. Normalising for variable mRNA dynamics of *eGFP*, *mCherry* and *tagBFP* to quantify transcriptional activity

In an aim to quantify the transcriptional activity of the synthetic promoters, a direct comparison of *eGFP*, *mCherry* and *tagBFP* expression both within the context of a MGEV or separate SGVs would be required. To investigate the comparability of the reporters, a transient co-transfection of *eGFP*, *mCherry* and *tagBFP* on separate SGVs utilising the same synthetic promoter variant (low, medium and high strength promoters, and hCMV-MIE promoter) at equimolar ratios over a range of DNA loads (100 to 800 ng) was performed in Medi-CHO, as shown in Figure 4-5. After 24h, samples were taken for RNA extraction and reverse transcribed to quantify cDNA as a function of mRNA copies by qPCR. The absolute quantification of *eGFP*, *mCherry* and *tagBFP* mRNA copies was achieved by generating a 7-point, 5-fold serial dilution DNA copy standard curve using linearised and purified SGVs encoding for *eGFP*, *mCherry* and *tagBFP* ranging from 1×10^8 to 6.4×10^3 copies. The DNA copies were plotted against observed Ct values detected by qPCR (data shown in the Figure B-5). The mRNA copies of *eGFP*, *mCherry* and *tagBFP* utilising the same strength synthetic promoter (Groups 1-4 in reference to Figure 4-5) across the DNA loads were quantified using an asymmetric sigmoidal regression. The mRNA copy of *eGFP*, *mCherry* and *tagBFP* when co-expressed as separate SGVs at a total DNA load of 600 ng was compared and shown in Figure 4-15

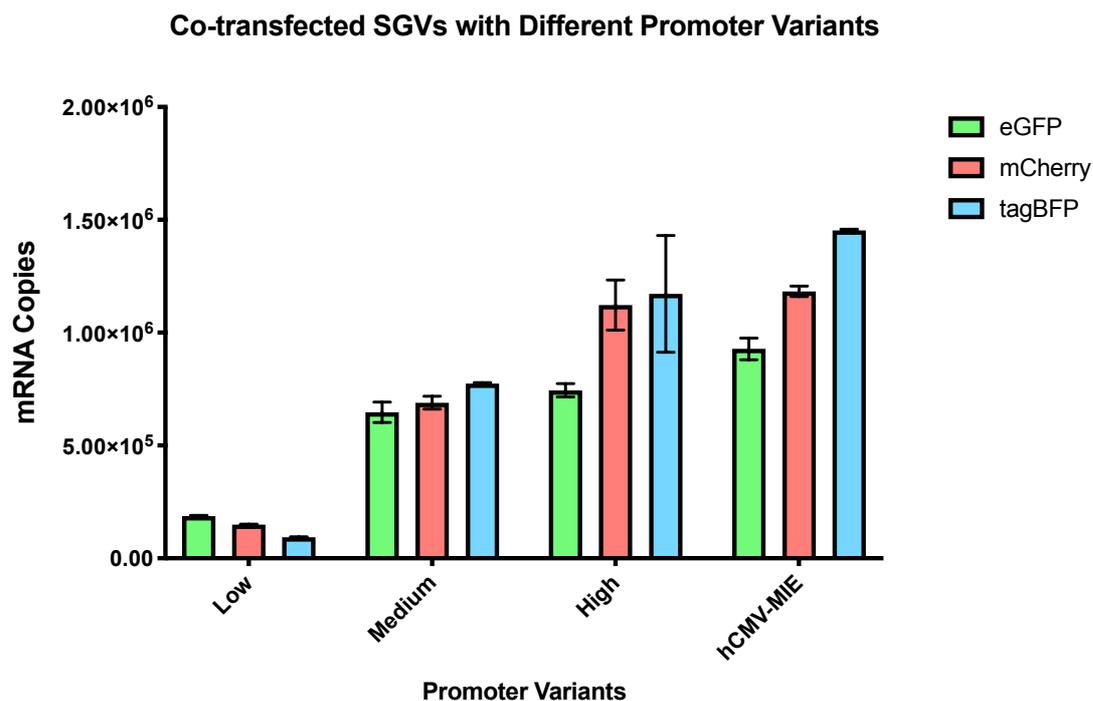


Figure 4-15: Differential mRNA dynamics *eGFP*, *mCherry* and *tagBFP* at different transcriptional activities.

Bar chart showing the quantified mRNA copies of *eGFP*, *mCherry* and *tagBFP* when regulated by the same promoter variant and co-expressed at an equimolar ratio in MedI-CHO equating to a total DNA load of 600 ng. The bars and error bars represent the average mRNA copies and SEM of biological triplicates.

The data shows that the mRNA copies of the fluorescent reporters vary when regulated by different strength synthetic promoters. For example, when *eGFP*, *mCherry* and *tagBFP* are dictated by the low strength synthetic promoter, mRNA copies detected were *eGFP* > *mCherry* > *tagBFP* with a 50.2% difference between *eGFP* and *tagBFP* mRNA copies. Conversely, the medium strength synthetic promoter-mediated expression of the fluorescent reporters demonstrated mRNA copies of *tagBFP* > *mCherry* > *eGFP* with a marginal difference of 16.5% between *eGFP* and *tagBFP* mRNA copies. The same trend in mRNA copies was observed when expression of the fluorescent reporters were dictated by the high strength synthetic promoter and hCMV-MIE promoter. However, the difference in mRNA copies were larger (36.5% and 36.2% for the high strength synthetic promoter and hCMV-MIE promoter respectively) when comparing *eGFP* and *tagBFP* mRNA copies. Although, promoter interference was detected (as shown in Figure 4-12), the mRNA copy variation between reporters was unexpected. This was because squelching would potentially repress overall promoter activity but not differentiate between fluorescent reporters, moreover, the TUs were standardised with an identical hCMV-MIE core element and

UTRs to standardise post-transcription rates (Barrett et al., 2012; Johari et al., 2019). Therefore, the varying mRNA copies between fluorescent reporters may be a consequence of differing mRNA dynamics, such as mRNA half-life and mRNA secondary structure formation (Mauger et al., 2019; Ross, 1995; Wang et al., 2017).

As a result, to directly compare fluorescent reporter expression, the mRNA copies measured for each promoter strength (low, medium, high and hCMV-MIE) at every transfected DNA load (Figure 4-5) were normalised relative to the low strength synthetic promoter and combined arithmetically. This generated an external calibration line for each fluorescent reporter deriving a relationship between mRNA copies and relative transcriptional activity (RTA) as displayed in Figure 4-16.

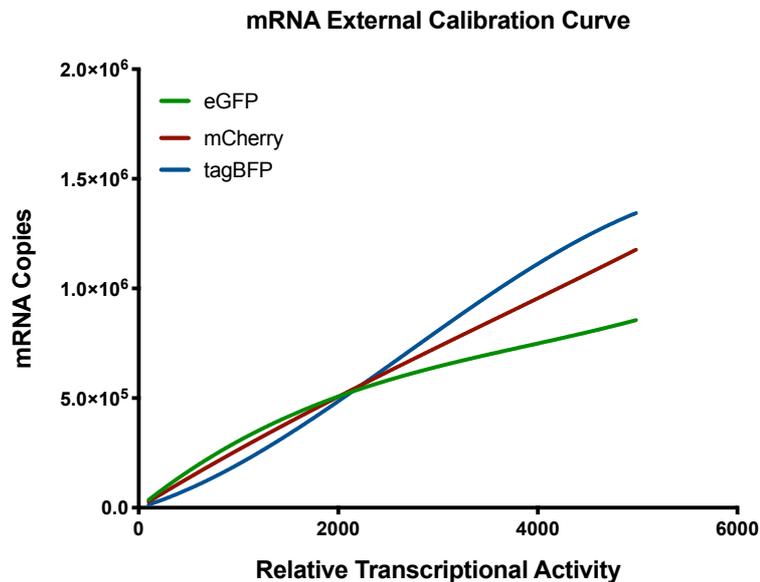


Figure 4-16: mRNA external calibration curves.

The external calibration curves constructed using the mRNA copies measured from different promoter strengths (low, medium, high and hCMV-MIE) across the range of transfected DNA loads (100 to 800 ng) for each fluorescent reporter (*eGFP*, *mCherry* and *tagBFP*). The curves represent the relationship of each fluorescent reporter mRNA copies to relative transcriptional activity (RTA) by accounting for differences in mRNA dynamics. A third-order polynomial regression curve was fitted for *eGFP* (green), *mCherry* (red) and *tagBFP* (blue).

A third-order polynomial regression curve was generated with an r^2 of 0.976, 0.988 and 0.964 for *eGFP*, *mCherry* and *tagBFP* respectively. The external calibration curves assume that the promoter-mediated transcriptional activity between reporters would be constant if normalised for mRNA dynamics. Therefore, these calibration curves were used to normalise mRNA copies of *eGFP*, *mCherry* and *tagBFP* for the

following experiments and facilitate direct comparison of the different fluorescent reporters both within SGVs and a MGEV context.

4.2.6. Characterising transcriptional strength-dependent positional effect within a multi-gene expression vector

Preliminary testing of the MGEV co-expressing *eGFP*, *mCherry* and *tagBFP* regulated by a hCMV-MIE promoter indicated a positional effect within the polycistronic cassette. A further investigation was explored in this section, to determine if the positional effect is transcriptional strength dependent or independent. This was achieved by constructing MGEVs expressing *eGFP*, *mCherry* and *tagBFP* in position 1, 2 and 3 in the polycistronic cassette. The fluorescent protein expression was regulated by the same strength promoter within the MGEV. The transcriptional activity was varied by using a low, medium and high activity synthetic promoter or a hCMV-MIE promoter as shown in Table 4-8.

Table 4-8: Multi-gene expression vector variants constructed to characterise positional effects. The table summarises the design and construction of MGEVs co-expressing *eGFP*, *mCherry* and *tagBFP*, including the loci of the fluorescent protein CDSes and the promoter variants utilised to dictate expression within the polycistronic cassette.

MGEV Variant	Promoter and fluorescent protein gene loci within the polycistronic cassette		
	Position 1 – <i>eGFP</i>	Position 2 - <i>mCherry</i>	Position 3 - <i>tagBFP</i>
MGEV-GCB-Low	Low	Low	Low
MGEV-GCB-Medium	Medium	Medium	Medium
MGEV-GCB-High	High	High	High
MGEV-GCB-CMV	hCMV-MIE	hCMV-MIE	hCMV-MIE

The experiment was performed by transfecting the MGEV variants alongside their SGV counterparts at similar gene copies as an equivalent control in Medi-CHO. After 24h, the absolute mRNA copies were quantified for each reporter using a DNA copy standard curve and normalised to RTA using an external calibration curve (as described in Section 4.2.5). The positional effect was quantified by comparing the RTA of the SGV controls against the respective MGEV variants. In Figure 4-17, the RTA of the SGV control, whereby no positional effect should be observed, was compared with the RTA of the MGEV variants in position 1, 2 and 3 of the polycistronic cassette.

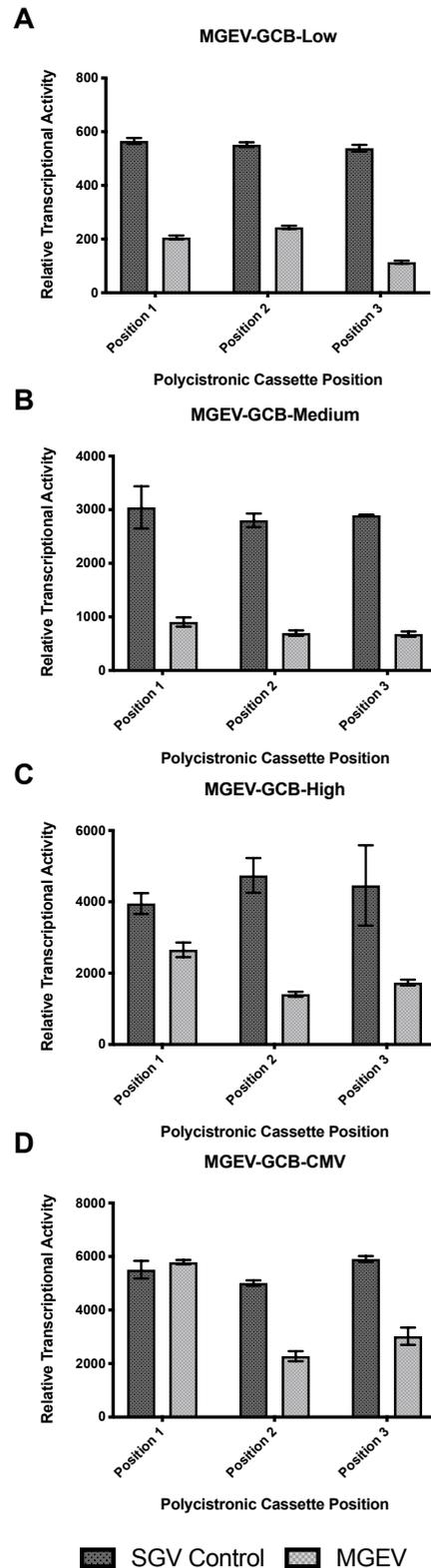


Figure 4-17: Quantifying transcriptional activity dependent positional effects within a multi-gene expression vector by qPCR.

The positional effect within the MGEV was measured by using qPCR to calculate the RTA in position 1, 2 and 3 expressing *eGFP*, *mCherry* and *tagBFP* driven by either a low (A), medium (B) or high (C) strength synthetic promoter, or a hCMV-MIE (D) promoter. The SGV control is the equimolar co-expression of the three fluorescent proteins as separate plasmids and the MGEV RTA is the expression within a tandem fix series polycistronic cassette. The error bars represent the SEM of biological triplicates.

Figure 4-17 shows the RTA was lower in all three positions within a MGEV regulated by each promoter variant. A varying degree of repression was observed apart from position 1 in the MGEV-GCB-CMV. The data showed the trend in RTA between position 1, 2 and 3 for MGEV-GCB- Low was position 2 > position 1 > position 3, for MGEV-GCB-Medium was position 1 > position 2 > position 3, and for MGEV-GCB-High and CMV was position 1 > position 3 > position 2. The fold change between the SGV control and MGEV RTA was calculated for each position and shown in Table 4-9.

Table 4-9: Fold Change of relative transcriptional activity to quantify positional effects of the multi-gene expression vector variants.

The average fold change of RTA in position 1, 2 and 3 for each MGEV variant was quantified against the expression of the counterpart SGVs by qPCR. The MGEV variants co-expressed *eGFP*, *mCherry* and *tagBFP* driven by a low, medium and high strength synthetic promoter, and a hCMV-MIE promoter in all three positions. The average and standard deviation of position fold change indicates whether the positional effect is constant or variable depending on different strength promoters.

MGEV Variants	Average Fold Change against SGV Expression			Average of Position Fold Change	Standard Deviation of Position Fold Change
	Position 1	Position 2	Position 3		
MGEV-GCB-Low	0.36	0.44	0.21	0.34	0.12
MGEV-GCB-Medium	0.30	0.25	0.23	0.26	0.03
MGEV-GCB-High	0.67	0.30	0.39	0.45	0.20
MGEV-GCB-CMV	1.05	0.45	0.51	0.67	0.33

The data indicated the fold change trend is dependent on transcriptional activity. For instance, the low strength promoter demonstrated a substantial decrease in expression, by an average of 66% with a marginal difference in expression between position 1 and 2 of 8%. The medium strength promoter showed the largest average decrease in expression by 74% with minimal difference in expression between position 2 and 3. The high strength promoter exhibited an average decrease in expression of 55% with variable expression in all three positions. The lowest effect was observed by the hCMV-MIE promoter-mediated expression, whereby an average reduction of 33% was observed. Furthermore, the standard deviation of the position fold change reiterates the positional effect is not constant.

The overall reduction and variation in in gene expression between positions within the polycistronic cassette infers a consequence of transcriptional interference by RNA polymerase II (RNA pol II) occlusion (Proudfoot, 2016; Shearwin et al., 2005). Furthermore, the transcriptional interference is dynamic and dependent on

transcriptional activity in position 1 (Palmer et al., 2011). Interestingly, the marginal increase in expression of hCMV-MIE in position 1 could be a result of promoter squelching (interaction between TFs or competition of TFs between promoter variants leading suppressed promoter activity) between hCMV-MIE and the upstream SV40 promoter (regulating *GS* expression). Previous research has indicated that a hCMV-IE promoter has exhibited squelching of SV40, repressing its transcriptional power (West, 2014), in turn minimising interference in position 1.

Following on from the transcriptional analysis, the fluorescent protein expression by flow cytometry was investigated to verify the initial observation. This was performed by transiently transfecting the MGEV variants and the same gene copies of the respective SGV controls in MedI-CHO using the HT transient transfection protocol. After 24h, the iMFI was measured by flow cytometry. The samples were gated for viable cells, followed by single cells and excluding doublets, the cellular autofluorescence was accounted for within the VL-1, BL-1 and YL-2 channels. The iMFI of the MGEV variants and the respective SGV controls were plotted in a bar chart as shown in Appendix B. The expression fold change of *eGFP*, *mCherry* and *tagBFP* within the MGEV was calculated against the respective SGV controls and plotted in the Figure 4-18 bar chart.

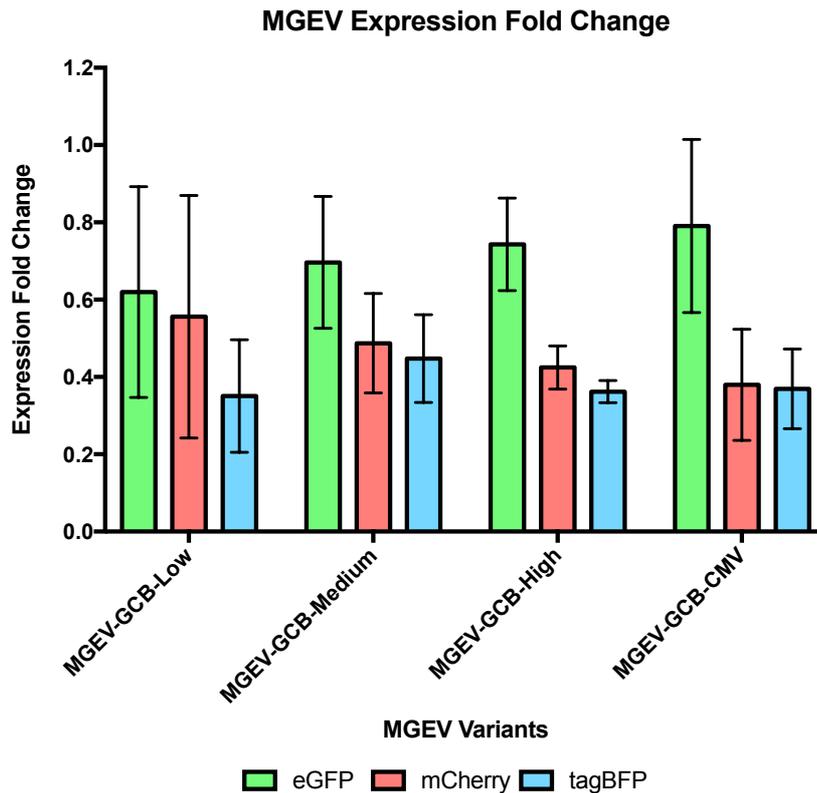


Figure 4-18: Expression fold change quantifying positional effect from fluorescent protein expression.

The relative fold change was derived by comparing the iMFI of the MGEV against the SGV control as a means to quantify the positional effect. The positional effect for each MGEV variant regulated by either a low, medium, or high strength synthetic promoter, or a hCMV-MIE promoter was depicted in the bar chart. The error bars represent SEM of biological triplicates.

The expression fold change in Figure 4-18 shows lower expression of all three fluorescent protein when compared against the respective SGV controls. The expression in position 1 was least repressed within all the MGEV variants with varying degree of repression in position 2 and 3. The data reiterates that positional effects within a MGEV is transcriptional strength dependent. The expression levels within the positions of the polycistronic cassette for all the MGEVs variants was position 1 > position 2 > position 3. The trend was inconsistent with the qPCR data and the data analysed in Chapter 3 for hCMV-MIE mediated expression, suggesting variation in expression data by flow cytometry. The variation is shown by the larger error bars and the poorer accuracy of quantification of the low and medium strength promoter, as shown in section 4.2.3.2. The average expression fold change was calculated for each MGEV variant within the polycistronic cassette as shown in Table 4-10.

Table 4-10: Average expression fold change of integrated median fluorescent intensity to quantify positional effect of multi-gene expression variants.

The average iMFI expression fold change of the MGEVs against the SGV controls for *eGFP*, *mCherry* and *tagBFP* were calculated. The data indicated the observed positional effect when expression of the fluorescent proteins is regulated by a low, medium and high strength synthetic promoter or a hCMV-MIE promoter.

MGEV Variants	Average Expression Fold Change against SGV Expression		
	Position 1	Position 2	Position 3
MGEV-GCB-Low	0.62	0.56	0.35
MGEV-GCB-Medium	0.70	0.49	0.45
MGEV-GCB-High	0.74	0.42	0.36
MGEV-GCB-CMV	0.79	0.38	0.37

The average expression fold changes varied compared to the qPCR derived data. The data indicates position 3 is most repressed in all MGEV variants ranging from 55% to 65%. Whereas, this was only observed in the MGEV-GCB-Low and Medium variants by qPCR quantification. Furthermore, the largest difference in reduced gene expression between the qPCR and flow cytometry data was for MGEV-GCB-Low and CMV. The variation maybe due to differences in post-transcriptional processes such as translation, protein folding and intracellular accumulation of the fluorescent proteins. Additionally, the lower accuracy of quantification by flow cytometry also contributed to the variation. The flow cytometry data concludes that positional effect is variable in positions 1 and 2 but less variable in position 3.

The overall conclusion from the qPCR and flow cytometry data indicated that gene repression within a MGEV is variable since the expression trend and fold changes were inconsistent. Due to the higher accuracy of qPCR quantification, it was used to determine the final conclusion that the positional effect within the polycistronic cassette is transcriptional strength dependent. Therefore, the combination of transcriptional interference is not constant due to the various biological mechanisms involved. As a result, these effects cannot be compensated for by using a library of synthetic promoters with varying transcriptional power.

4.2.7. Demonstrating titration of gene expression and determining stoichiometric ratio within a multi-gene expression vector

The characterisation of the MGEV and the synthetic promoters have shown that variation in expression profile was observed. Unfortunately, transcriptional interference and promoter squelching was dynamic and variable in nature. A current method to quantify transcriptional interference is using a nuclear run on assay, where hybridisation of radiolabelled nascent RNA on a northern blot shows the targeted region directly downstream of the 3'UTR and within the neighbouring TU's proximal promoter region. This would quantify RNA pol II occlusion-mediated repression as an indicator of inefficient transcription termination (West and Proudfoot, 2009). Whereas, a means to detect promoter squelching is by measuring repressed transcriptional activity of promoters by qPCR, on separate single gene plasmids using different unlinked reporters when paired with multiple promoter variants and expressed concurrently. The comparison of the respective reporter mRNA copies would identify potential promoter interference-mediated repression.

However, both of these methods are cumbersome and both effects are context-specific to the promoter variants used and MGEV design. Therefore, even though quantification of transcriptional interference and promoter squelching is possible, it would still be difficult to account for both of these effects within the MGEV design space. As a result, the predictability of titrated gene expression within a MGEV is more challenging than initially hypothesised. Due to these limitations, a larger library of MGEV variants were tested to demonstrate titration of gene expression and define the stoichiometric ratio by using synthetic promoters with varying transcriptional activity. Figure 4-19 summarises the configuration of the MGEV including the order of TU and genetic components employed within each TU.

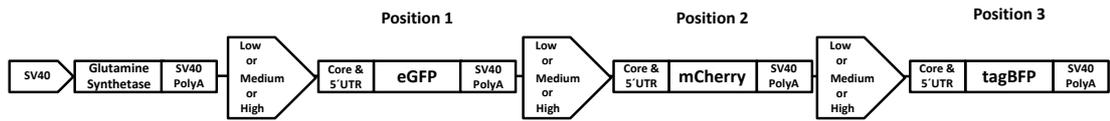


Figure 4-19: Schematic describing the MGEV configuration.

The polycistronic cassette design depicts the order of three TUs encoding for the fluorescent reporter proteins (*eGFP*, *mCherry* and *tagBFP*) in a fixed tandem series within the MGEV. Upstream of the polycistronic cassette is an industry standard glutamine synthetase selection cassette controlled by a SV40 promoter to simulate standard biopharmaceutical production plasmid configuration. The synthetic promoters were interchanged within each position to achieve every promoter combination leading to a library of 27 MGEV variants described in Table 4-11. Each TU comprises of an identical hCMV-MIE core element, 5'UTR and SV40 pA element.

4.2.7.1. Construction of 27 different multi-gene expression vector variants

A library of 27 different MGEVs were designed and constructed that co-expressed *eGFP*, *mCherry* and *tagBFP* in position 1, 2 and 3 of the polycistronic cassette, as shown in Figure 4-19. The gene expression was titrated using the 5RPU.1, 40RPU.2 and 80RPU.1 promoters representing low, medium and high transcriptional strength respectively. As a consequence of the positional effect, MGEV variants were constructed by varying the loci and combination of promoters within the polycistronic cassette. Table 4-11 describes the entire library of 27 MGEVs design and constructed.

Table 4-11: Summary of synthetic promoter combination within each multi-gene expression vector variant constructed.

The identified low (5RPU.1), medium (40RPU.2) and high (80RPU.1) strength synthetic promoters were varied in every combination possible and designated a name as shown in the table. Each MGEV variant co-expressed *eGFP* in position 1, *mCherry* in position 2 and *tagBFP* in position 3.

Library Sub-category	MGEV Variant	Synthetic Promoter Combination		
		Position 1 (<i>eGFP</i>)	Position 2 (<i>mCherry</i>)	Position 3 (<i>tagBFP</i>)
Low Strength Combinations	MGEV-GCB-1	5RPU.1	5RPU.1	5RPU.1
	MGEV-GCB-2	5RPU.1	5RPU.1	40RPU.2
	MGEV-GCB-3	40RPU.2	5RPU.1	5RPU.1
	MGEV-GCB-4	5RPU.1	40RPU.2	5RPU.1
	MGEV-GCB-5	5RPU.1	5RPU.1	80RPU.1
	MGEV-GCB-6	80RPU.1	5RPU.1	5RPU.1
	MGEV-GCB-7	5RPU.1	80RPU.1	5RPU.1
Medium Strength Combinations	MGEV-GCB-8	40RPU.2	40RPU.2	40RPU.2
	MGEV-GCB-9	40RPU.2	40RPU.2	5RPU.1
	MGEV-GCB-10	5RPU.1	40RPU.2	40RPU.2
	MGEV-GCB-11	40RPU.2	5RPU.1	40RPU.2
	MGEV-GCB-12	40RPU.2	40RPU.2	80RPU.1
	MGEV-GCB-13	80RPU.1	40RPU.2	40RPU.2
	MGEV-GCB-14	40RPU.2	80RPU.1	40RPU.2
High Strength Combinations	MGEV-GCB-15	80RPU.1	80RPU.1	80RPU.1
	MGEV-GCB-16	80RPU.1	80RPU.1	5RPU.1
	MGEV-GCB-17	5RPU.1	80RPU.1	80RPU.1
	MGEV-GCB-18	80RPU.1	5RPU.1	80RPU.1
	MGEV-GCB-19	80RPU.1	80RPU.1	40RPU.2
	MGEV-GCB-20	40RPU.2	80RPU.1	80RPU.1
	MGEV-GCB-21	80RPU.1	40RPU.2	80RPU.1
Mixed Strength Combinations	MGEV-GCB-22	5RPU.1	40RPU.2	80RPU.1
	MGEV-GCB-23	5RPU.1	80RPU.1	40RPU.2
	MGEV-GCB-24	40RPU.2	5RPU.1	80RPU.1
	MGEV-GCB-25	80RPU.1	5RPU.1	40RPU.2
	MGEV-GCB-26	80RPU.1	40RPU.2	5RPU.1
	MGEV-GCB-27	40RPU.2	80RPU.1	5RPU.1

The MGEV-GCB-1, 8 and 15 had previously been constructed for the characterisation of positional effects within the MGEV as discussed in section 4.2.6. The remaining 24 MGEV variants were constructed using the TU-1, 2 and 3 encoding *eGFP*, *mCherry* and *tagBFP* respectively, and the pExp-Vec-GG recipient vector at a 2:1 molar ratio. Transformed colonies were screened for successful construction of the MGEVs by restriction digest colony screen. The purified plasmid DNA was digested by NotI restriction endonuclease and the DNA fragment sizes were verified by gel electrophoresis. Figure 4-20 shows the results of the restriction digest colony screen to identify the successful construction of the 24 MGEV variants.



Figure 4-20: Verifying successful construction of the remaining 24 multi-gene expression vector variants. 1% agarose gel showing the outcome of a restriction digest colony screen using NotI endonuclease. The DNA bands highlighted within the red boxes were selected and progressed towards DNA amplification. This was decided based on their band size, as per predicted, and intensity, indicating the transformed colony contained a high copy number.

The screening of 72 colonies (3 per MGEV variant) showed the successful construction MGEV variants based on the number of DNA fragments and their respective sizes to represent the correct genetic elements within the MGEV. The successfully assembled MGEV exhibited 4 fragments – fragment 1 was 6214bp, fragment 2 was between 1483 to 1582 bp, fragment 3 was between 1217 to 1316 bp and fragment 4 was between 1208 and 1307 bp. The gel image in Figure 4-20 shows only 3 bands, and this was expected due to minimal separation observed between fragment 3 and 4. However, the thickness and intensity of the smallest band when compared to the other fragments within the lane indicated co-migration of two fragments. The second criteria for selection was high intensity bands indicating high copy number of plasmids within the transformed *E.coli*. The lanes highlighted in red and labelled in Figure 4-20 were the respective MGEV variants selected for amplification and progressed towards expression screens.

4.2.7.2. Quantifying gene expression stoichiometric ratios of the multi-gene expression vector library by qPCR

The gene expression stoichiometric ratios for the MGEV library were quantified using qPCR. The aim was to define the stoichiometric ratio using a set of validated promoters for each variant within the library. As a result, a list was generated of combined ratios for 3 genes to use for various engineering applications. This was achieved by transiently expressing the MGEV library and a mock (no DNA) control in MedI-CHO using the HT transient transfection protocol. After 24 h of expression, the RNA was extracted from the transfected cells and reverse transcribed into cDNA. The mRNA copies for *GS*, *eGFP*, *mCherry* and *tagBFP* were quantified by using a respective gene copy standard curve ranging from 1×10^8 to 6.4×10^3 copies of linearised template DNA. The *GS* expression was quantified as it was positioned upstream of the polycistronic cassette and utilising a SV40 promoter, which is a standard design of an industrially relevant biopharmaceutical production plasmid. Therefore, characterising the expression of *GS* would elucidate the transcriptional strength of the SV40 promoter and how it would impact expression of the downstream TUs within a MGEV context particularly for CHO cell engineering and biopharmaceutical production applications. The mRNA copies were quantified and plotted in a bar chart as shown in Figure 4-21.

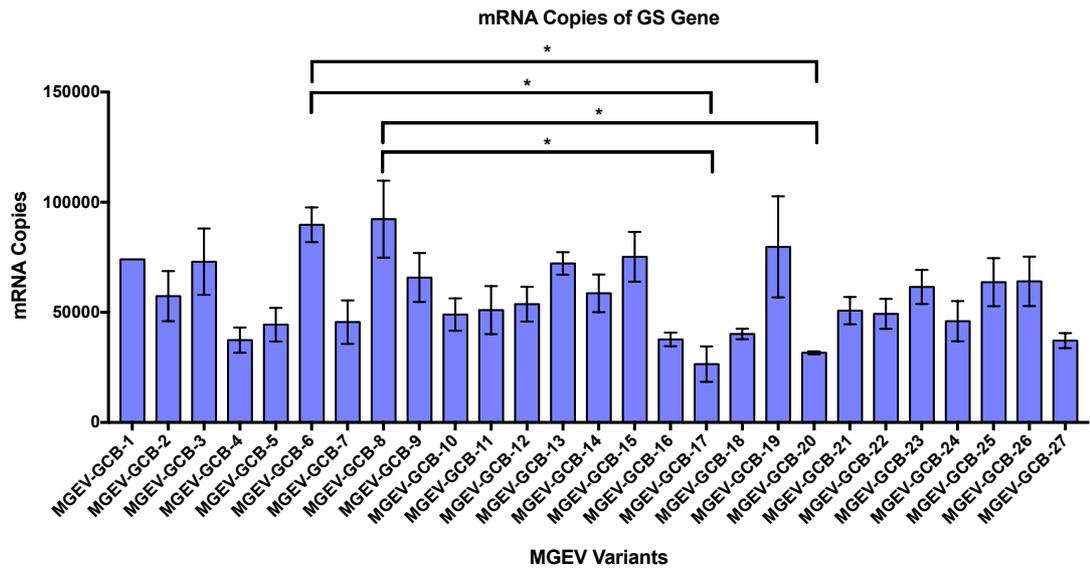


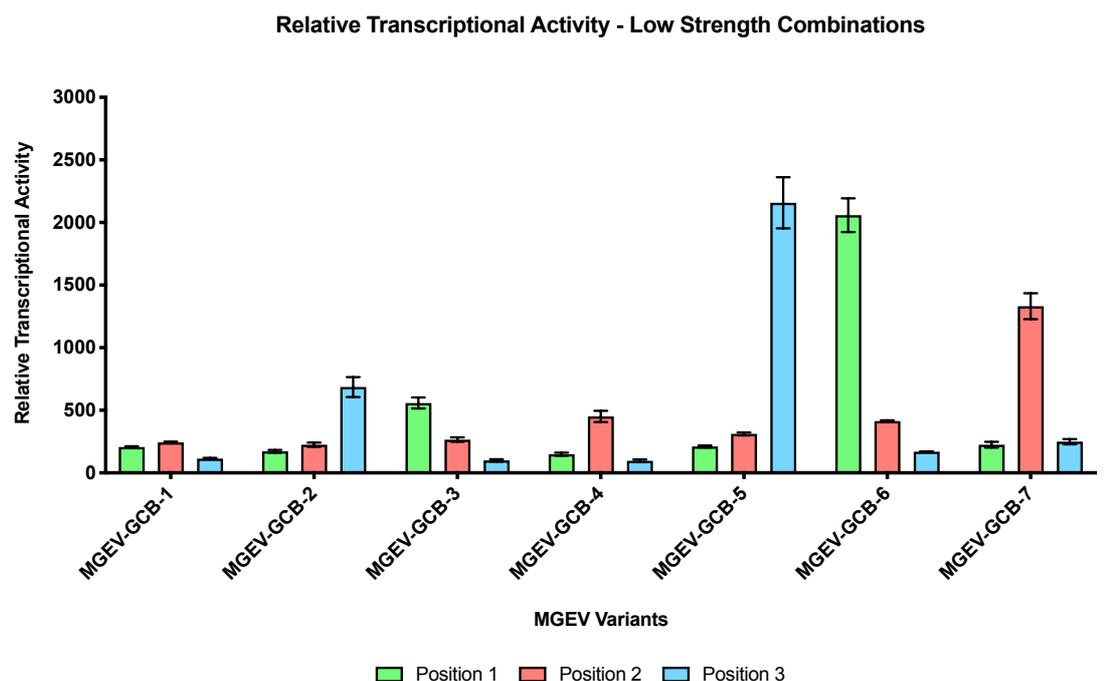
Figure 4-21: Absolute mRNA copies of *GS* gene within multi-gene expression vector variants.

The absolute quantification of *GS* mRNA copies regulated by a SV40 promoter upstream of the polycistronic cassette within the MGEV library. The mRNA was quantified by asymmetric sigmoidal regression of a *GS* template standard curve using qPCR. The error bars represent the SEM of biological triplicates. A one-way ANOVA test was performed and the ‘*’ represents significant difference ($p < 0.0332$) in *GS* mRNA copies.

Prior to quantification, the Ct values were adjusted using *mmadhc* and *fkbp1a* housekeeping genes to normalise for cDNA levels. The absolute quantification of recombinant *GS* expression was calculated by quantifying the endogenous *GS* mRNA copies from mock transfected cells and subtracting it from transfected cells containing the MGEV. The data in Figure 4-21 showed that recombinant *GS* mRNA copies quantified ranged from 26501.7 to 92362.1 mRNA copies. The average mRNA copies and percentage of standard deviation across the entire library is 56594.2 and 30.5%. Therefore, the average transcriptional activity of the SV40 was similar to the low strength synthetic promoter when comparing mRNA copies detected (data not shown). The one-way ANOVA statistical analysis showed no significant difference in *GS* mRNA copies, apart from four combinations – MGEV-GCB-6 vs 17, MGEV-GCB-6 vs 20, MGEV-GCB-8 vs 17 and MGEV-GCB-8 vs 20. Furthermore, there was no apparent correlation in increased or decreased *GS* transcription with the promoter directly downstream of the *GS* gene. However, the percentage of standard deviation observed indicates some variation in *GS* expression, which could be contributed by negative supercoiling leading to conformational change of a circular MGEV (Curtin et al., 2008; Ma and Wang, 2016) caused by RNA pol II transcription elongation complex resulting in inhibition of upstream (*GS* gene) transcription initiation. Alternatively, SV40 transcriptional activity could have been affected by the

interference with neighbouring synthetic promoters, whereby enhancer TFs between promoter variants could be interacting resulting in variable transcriptional activity (Schmidt et al., 2016). As a result, the varying recombinant *GS* mRNA copies indicates potential variability in selection stringency during stable cell line generation due to differing transcriptional activity of SV40.

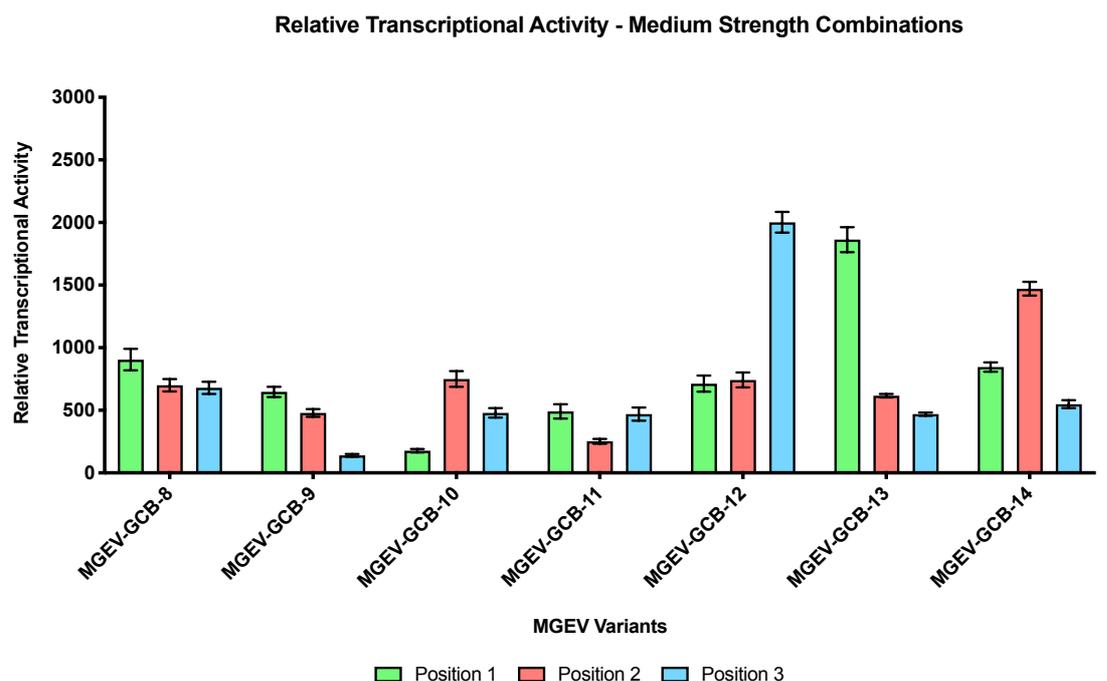
The quantified mRNA copies for *eGFP*, *mCherry* and *tagBFP* within the MGEV library were normalised for variable mRNA dynamics observed during SGV expression. The mRNA copies were normalised using the external calibration curve constructed from SGV co-expression of the fluorescent reporters (Figure 4-16). Therefore, the normalised mRNA copies was referred to as RTA. The library was split into 4 subcategories – low strength combinations, medium strength combinations, high strength combinations and mixed strength combinations as indicated on Table 4-11. The low strength combination represented the low strength synthetic promoter being constant in 2 out 3 positions with the third position comprising of a different promoter variant. This was the same for the medium and high strength combinations. The mixed strength combination category showed MGEVs where three different strength promoters were in every position within the polycistronic cassette. The RTA were divided within the sub-categories and the bar charts were plotted according to positional order as shown in in Figure 4-22, 23, 24 and 25.



MGEV Variants	Predicted Expression Level			Observed Expression Level		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
MGEV-GCB-1	Low	Low	Low	Low	Low	Low
MGEV-GCB-2	Low	Low	Medium	Low	Low	Medium
MGEV-GCB-3	Medium	Low	Low	Medium	Low	Low
MGEV-GCB-4	Low	Medium	Low	Low	Medium	Low
MGEV-GCB-5	Low	Low	High	Low	Medium	High
MGEV-GCB-6	High	Low	Low	High	Medium	Low
MGEV-GCB-7	Low	High	Low	Low	High	Low

Figure 4-22: Quantified gene expression levels within multi-gene expression vector variants by qPCR – Low strength combination subcategory.

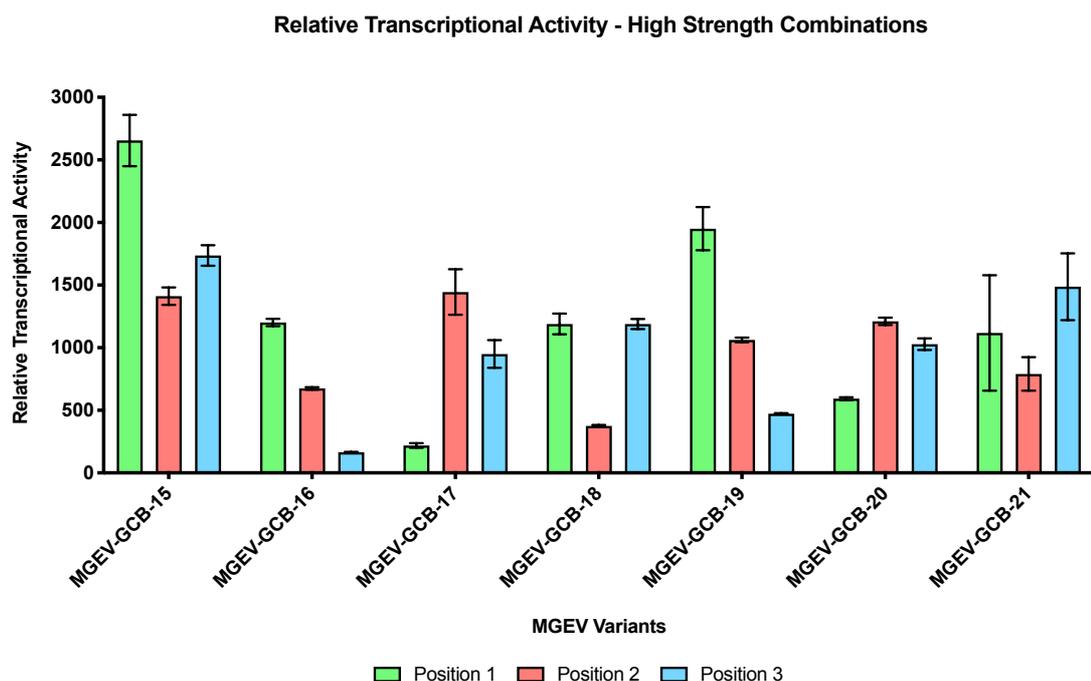
The average RTA in position 1, 2 and 3 of the MGEV variants 1 to 7 are represented in this bar chart. This depicts the gene expression level in each position within the polycistronic cassette where the low strength synthetic promoter was predominant and a medium and high strength synthetic promoter was introduced. The RTA was derived by normalising the mRNA copies detected of *eGFP*, *mCherry* and *tagBFP* using an external calibration curve (Figure 4-16) to facilitate direct comparison of the different fluorescent reporters. The error bars represent SEM of biological triplicates. The table represents the predicted vs observed strength of expression by applying a series of RTA threshold rules which were – low expression: 1 to 275; medium expression: 276 to 905; high expression: 906 to 2656.



MGEV Variants	Predicted Expression Level			Observed Expression Level		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
MGEV-GCB-8	Medium	Medium	Medium	Medium	Medium	Medium
MGEV-GCB-9	Medium	Medium	Low	Medium	Medium	Low
MGEV-GCB-10	Low	Medium	Medium	Low	Medium	Medium
MGEV-GCB-11	Medium	Low	Medium	Medium	Low	Medium
MGEV-GCB-12	Medium	Medium	High	Medium	Medium	High
MGEV-GCB-13	High	Medium	Medium	High	Medium	Medium
MGEV-GCB-14	Medium	High	Medium	Medium	High	Medium

Figure 4-23: Quantified gene expression levels within multi-gene expression vector variants by qPCR – Medium strength combination subcategory.

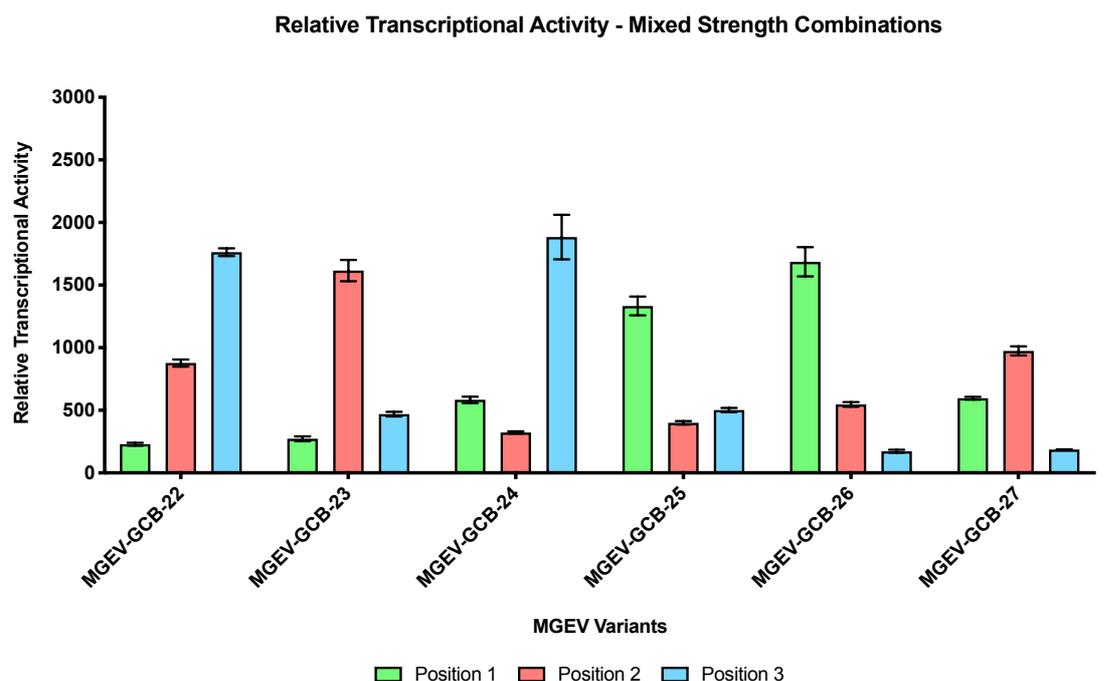
The average RTA in position 1, 2 and 3 of the MGEV variants 8 to 14 are represented in this bar chart. This demonstrates the gene expression level observed in each position within the polycistronic cassette where the medium strength synthetic promoter was predominant and a low and high strength synthetic promoter was introduced. The RTA was derived by normalising the mRNA copies detected of *eGFP*, *mCherry* and *tagBFP* using an external calibration curve (Figure 4-16) to facilitate direct comparison of the different fluorescent reporters. The error bars represent SEM of biological triplicates. The table represents the predicted vs observed strength of expression by applying a series of RTA threshold rules which were – low expression: 1 to 275; medium expression: 276 to 905; high expression: 906 to 2656.



MGEV Variants	Predicted Expression Level			Observed Expression Level		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
MGEV-GCB-15	High	High	High	High	High	High
MGEV-GCB-16	High	High	Low	High	Medium	Low
MGEV-GCB-17	Low	High	High	Low	High	High
MGEV-GCB-18	High	Low	High	High	Medium	High
MGEV-GCB-19	High	High	Medium	High	High	Medium
MGEV-GCB-20	Medium	High	High	Medium	High	High
MGEV-GCB-21	High	Medium	High	High	Medium	High

Figure 4-24: Quantified gene expression levels within multi-gene expression vector variants by qPCR – High strength combination subcategory.

The average RTA in position 1, 2 and 3 of the MGEV variants 15 to 21 are represented in this bar chart. This portrays the gene expression level observed in each position within the polycistronic cassette where the high strength synthetic promoter was predominant and a low and medium strength synthetic promoter was introduced. The RTA was derived by normalising the mRNA copies detected of *eGFP*, *mCherry* and *tagBFP* using an external calibration curve (Figure 4-16) to facilitate direct comparison of the different fluorescent reporters. The error bars represent SEM of biological triplicates. The table represents the predicted vs observed strength of expression by applying a series of RTA threshold rules which were – low expression: 1 to 275; medium expression: 276 to 905; high expression: 906 to 2656.



MGEV Variants	Predicted Expression Level			Observed Expression Level		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
MGEV-GCB-22	Low	Medium	High	Low	Medium	High
MGEV-GCB-23	Low	High	Medium	Low	High	Medium
MGEV-GCB-24	Medium	Low	High	Medium	Medium	High
MGEV-GCB-25	High	Low	Medium	High	Medium	Medium
MGEV-GCB-26	High	Medium	Low	High	Medium	Low
MGEV-GCB-27	Medium	High	Low	Medium	High	Low

Figure 4-25: Quantified gene expression levels within multi-gene expression vector variants by qPCR – Mixed strength combination subcategory.

The average RTA in position 1, 2 and 3 of the MGEV variants 22 to 27 are represented in this bar chart. This depicts the gene expression level in each position within MGEV when a low, medium and high strength synthetic promoter variant are interchanged between positions within the polycistronic cassette. The RTA was derived by normalising the mRNA copies detected of *eGFP*, *mCherry* and *tagBFP* using an external calibration curve (Figure 4-16) to facilitate direct comparison of the different fluorescent reporters. The error bars represent SEM of biological triplicates. The table represents the predicted vs observed strength of expression by applying a series of RTA threshold rules which were – low expression: 1 to 275; medium expression: 276 to 905; high expression: 906 to 2656.

The results displayed in Figure 4-22, 23, 24 and 25 shows that the overall trend of expression in different positions of the MGEV across the library generally correlated to the expected strength of the synthetic promoter. However, the specific RTAs varied between MGEV variants and within the polycistronic cassette of each MGEV that had multiple units of the same strength synthetic promoter. This was expected due to the observed promoter squelching and positional effects discussed and quantified in section 4.2.4 and 4.2.6 respectively. In Figure 4-22, 23 and 24, the array of RTAs observed adhered to the trend showing low, medium and high level of expression,

exhibiting a titrated response based on transcriptional activity. In Figure 4-25, where the promoter strengths are varied, the overall trend of low, medium or high level of expression was observed. Therefore, titrated expression has been demonstrated by using promoters with different levels of defined transcriptional activity.

The stoichiometric ratio of gene expression between position 1, 2 and 3 of the polycistronic cassette for each MGEV variant was calculated. This was achieved by identifying the lowest average RTA value in the entire library (demonstrating the lowest transcriptional activity observed within a MGEV context) which was position 3 of MGEV-GCB-4 dictated by the low strength synthetic promoter. Therefore, gene expression stoichiometry was calculated by deriving a fold change in each position of each MGEV variant relative to the lowest observed average RTA in the library (96.2). These stoichiometric ratios were tabulated alongside the promoter strength allocation in each position of the polycistronic cassette and displayed in Table 4-12.

The range of fold change for low transcriptional activity was between 1.0 and 4.3, for medium transcriptional activity was between 4.88 to 9.41 and for high transcriptional activity 7.0 to 27.6. Moreover, the percentage deviation of the low, medium and high strength synthetic promoters across the entire library irrespective of position was 37.6, 22.2 and 30% respectively. The wide range in fold changes and deviation was a consequence of differing levels of gene repression potentially caused by various mechanisms influencing transcription such as positional-mediated transcriptional interference and promoter squelching within the MGEV. However, this has also yielded a set of empirically-derived gene stoichiometric ratios to indicate an approximate level of relative gene expression using a particular configuration of TU positions and synthetic promoters where transcriptional variation has been accounted for.

Table 4-12: Gene expression stoichiometric ratios of 27 multi-gene expression vector variants.

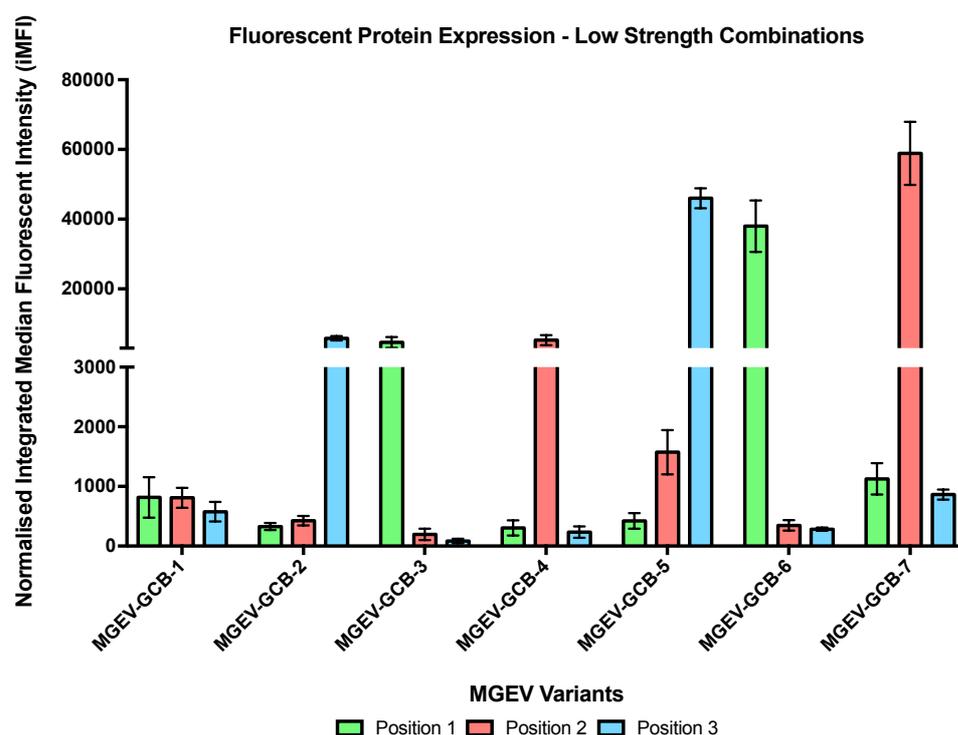
The average fold change derived relative to the lowest observed average RTA within the MGEV library (position 3 of MGEV-GCB-4 utilising a low strength synthetic promoter) was calculated to define the titration of gene expression in position 1, 2 and 3 using varying combinations of synthetic promoter strengths (low, medium and high). The data defines the stoichiometric ratio of gene expression within the MGEV transcriptionally.

Synthetic Promoter Combination			Stoichiometric Ratios		
Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
5RPU.1	5RPU.1	5RPU.1	2.14	2.54	1.19
5RPU.1	5RPU.1	40RPU.2	1.80	2.34	7.13
5RPU.1	5RPU.1	80RPU.1	2.20	3.24	22.44
5RPU.1	40RPU.2	5RPU.1	1.55	4.68	1.00
5RPU.1	40RPU.2	40RPU.2	1.84	7.79	4.99
5RPU.1	40RPU.2	80RPU.1	2.38	9.13	18.33
5RPU.1	80RPU.1	5RPU.1	2.34	13.84	2.59
5RPU.1	80RPU.1	40RPU.2	2.83	16.81	4.88
5RPU.1	80RPU.1	80RPU.1	2.28	15.02	9.88
40RPU.2	5RPU.1	5RPU.1	5.80	2.76	1.04
40RPU.2	5RPU.1	40RPU.2	5.11	2.61	4.88
40RPU.2	5RPU.1	80RPU.1	6.07	3.36	19.59
40RPU.2	40RPU.2	5RPU.1	6.73	4.96	1.45
40RPU.2	40RPU.2	40RPU.2	9.41	7.27	7.07
40RPU.2	40RPU.2	80RPU.1	7.42	7.71	20.82
40RPU.2	80RPU.1	5RPU.1	6.20	10.13	1.93
40RPU.2	80RPU.1	40RPU.2	8.79	15.28	5.71
40RPU.2	80RPU.1	80RPU.1	6.17	12.59	10.69
80RPU.1	5RPU.1	5RPU.1	21.41	4.31	1.76
80RPU.1	5RPU.1	40RPU.2	13.87	4.16	5.23
80RPU.1	5RPU.1	80RPU.1	12.37	3.91	12.37
80RPU.1	40RPU.2	5RPU.1	17.54	5.69	1.80
80RPU.1	40RPU.2	40RPU.2	19.37	6.39	4.88
80RPU.1	40RPU.2	80RPU.1	11.63	8.22	15.46
80RPU.1	80RPU.1	5RPU.1	12.49	7.02	1.71
80RPU.1	80RPU.1	40RPU.2	20.29	11.05	4.92
80RPU.1	80RPU.1	80RPU.1	27.61	14.68	18.06

4.2.7.3. Quantifying gene expression trends of the multi-gene expression vector library by flow cytometry

An orthogonal means to quantify titration of gene expression within the MGEV library was by flow cytometry. The aim of this was to observe if titration of gene expression can be observed by fluorescent intensity. This was measured by transiently expressing the MGEV library in MedI-CHO using the HT transient transfection protocol. After 24h, the fluorescent protein expression was detected by using the Attune NxT flow cytometer. The iMFI was measured by gating for viable cells, followed by identifying single cells and excluding doublets, and accounting for the cellular autofluorescence

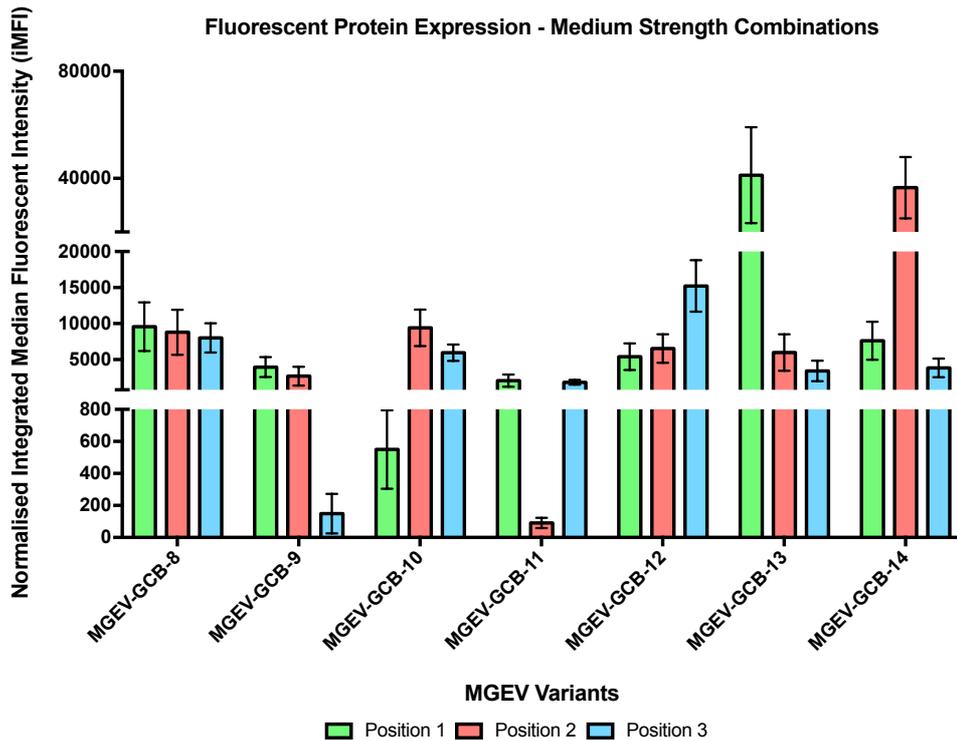
within the VL-1, BL-1 and YL-2 channels. The iMFI for *eGFP* and *mCherry* was normalised to *tagBFP* using an adjustment factor derived from SGV co-expression of the three fluorescent proteins. The normalised iMFI for position 1, 2 and 3 were plotted as bar charts and divided into the same subcategories described for the qPCR data – low strength combinations, medium strength combinations, high strength combinations and mix strength combinations. The bar charts are shown in Figures 4-26, 27, 28 and 29.



MGEV Variants	Predicted Expression Level			Observed Expression Level		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
MGEV-GCB-1	Low	Low	Low	Low	Low	Low
MGEV-GCB-2	Low	Low	Medium	Low	Low	Medium
MGEV-GCB-3	Medium	Low	Low	Medium	Low	Low
MGEV-GCB-4	Low	Medium	Low	Low	Medium	Low
MGEV-GCB-5	Low	Low	High	Low	Low	High
MGEV-GCB-6	High	Low	Low	High	Low	Low
MGEV-GCB-7	Low	High	Low	Low	High	Low

Figure 4-26: Quantified fluorescent protein expression of multi-gene expression vectors by flow cytometry – Low strength combination subcategory.

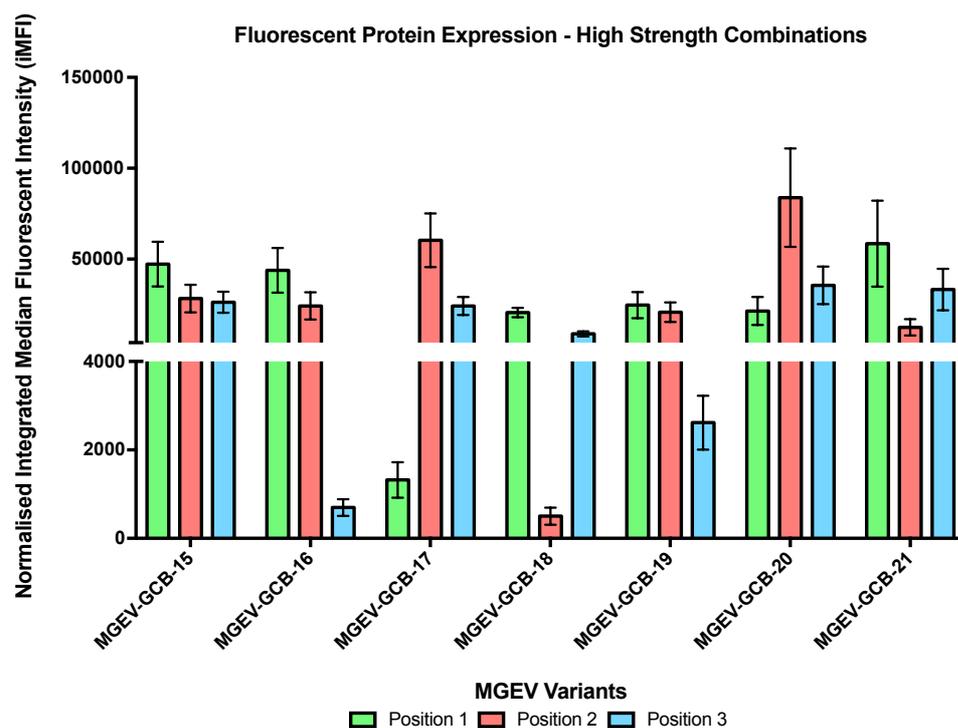
The normalised iMFI for *eGFP*, *mCherry* and *tagBFP* in position 1, 2 and 3 were plotted. The data represented the expression profile of the MGEV variants where 2 out of 3 TUs are driven by a low strength synthetic promoter while the remaining TU was regulated by either a low, medium or high strength synthetic promoter. The error bars represent SEM of biological triplicate. The table represents the predicted vs observed strength of expression by applying a series of iMFI threshold rules which were – low expression: 1 to 3000; medium expression: 3001 to 22500; high expression: 22501 to 100000.



MGEV Variants	Predicted Expression Level			Observed Expression Level		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
MGEV-GCB-8	Medium	Medium	Medium	Medium	Medium	Medium
MGEV-GCB-9	Medium	Medium	Low	Medium	Low	Low
MGEV-GCB-10	Low	Medium	Medium	Low	Medium	Medium
MGEV-GCB-11	Medium	Low	Medium	Low	Low	Low
MGEV-GCB-12	Medium	Medium	High	Medium	Medium	High
MGEV-GCB-13	High	Medium	Medium	High	Medium	Medium
MGEV-GCB-14	Medium	High	Medium	Medium	High	Medium

Figure 4-27: Quantified fluorescent protein expression of multi-gene expression vectors by flow cytometry – Medium strength combination subcategory.

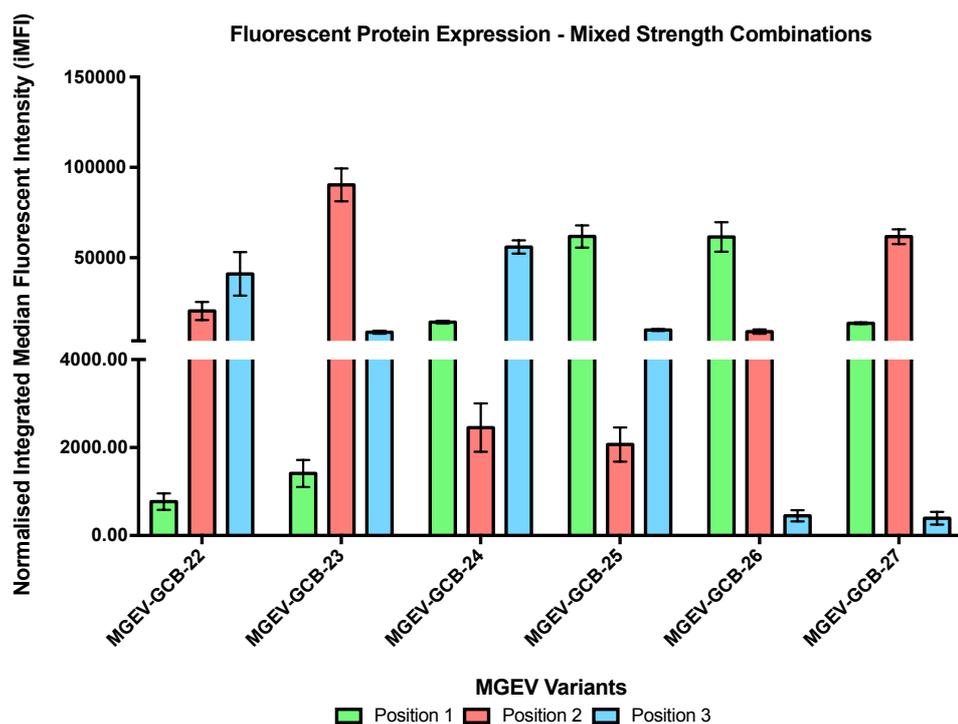
The normalised iMFI for *eGFP*, *mCherry* and *tagBFP* in position 1, 2 and 3 were plotted. The data represented the expression profile of the MGEV variants where 2 out of 3 TUs are driven by a medium strength synthetic promoter while the remaining TU was regulated by either a low, medium or high strength synthetic promoter. The error bars represent SEM of biological triplicate. The table represents the predicted vs observed strength of expression by applying a series of iMFI threshold rules which were – low expression: 1 to 3000; medium expression: 3001 to 22500; high expression: 22501 to 100000.



MGEV Variants	Predicted Expression Level			Observed Expression Level		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
MGEV-GCB-15	High	High	High	High	High	High
MGEV-GCB-16	High	High	Low	High	High	Low
MGEV-GCB-17	Low	High	High	Low	High	High
MGEV-GCB-18	High	Low	High	Medium	Low	Medium
MGEV-GCB-19	High	High	Medium	High	Medium	Low
MGEV-GCB-20	Medium	High	High	Medium	High	High
MGEV-GCB-21	High	Medium	High	High	Medium	High

Figure 4-28: Quantified fluorescent protein expression of multi-gene expression vectors by flow cytometry – High strength combination subcategory.

The normalised iMFI for *eGFP*, *mCherry* and *tagBFP* in position 1, 2 and 3 were plotted. The data represented the expression profile of the MGEV variants where 2 out of 3 TUs are driven by a high strength synthetic promoter while the remaining TU was regulated by either a low, medium or high strength synthetic promoter. The error bars represent SEM of biological triplicate. The table represents the predicted vs observed strength of expression by applying a series of iMFI threshold rules which were – low expression: 1 to 3000; medium expression: 3001 to 22500; high expression: 22501 to 100000.



MGEV Variants	Predicted Expression Level			Observed Expression Level		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
MGEV-GCB-22	Low	Medium	High	Low	Medium	High
MGEV-GCB-23	Low	High	Medium	Low	High	Medium
MGEV-GCB-24	Medium	Low	High	Medium	Low	High
MGEV-GCB-25	High	Low	Medium	High	Low	Medium
MGEV-GCB-26	High	Medium	Low	High	Medium	Low
MGEV-GCB-27	Medium	High	Low	Medium	High	Low

Figure 4-29: Quantified fluorescent protein expression of multi-gene expression vectors by flow cytometry – Mixed strength combination subcategory.

The normalised iMFI for *eGFP*, *mCherry* and *tagBFP* in position 1, 2 and 3 were plotted. The data represented the expression profile of the MGEV variants where the synthetic promoter variants positions are varied within the polycistronic cassette. The error bars represent SEM of biological triplicate. The table represents the predicted vs observed strength of expression by applying a series of iMFI threshold rules which were – low expression: 1 to 3000; medium expression: 3001 to 22500; high expression: 22501 to 100000.

The normalised iMFI demonstrated that gene expression within a MGEV had been titrated by using synthetic promoters with different transcriptional strength as shown in Figure 4-26, 27, 28 and 29. Furthermore, the impact of transcriptional variability were observed by the difference in normalised iMFI when expression was regulated by the same strength promoter. The stoichiometric ratios of intracellular protein levels were calculated using the identical approach to the RTA data set (data not shown). The ratio values did not match the transcriptionally derived stoichiometric ratios. After further investigation, it was identified that large levels of variation were observed

during quantification of low intracellular protein expression levels inferring it was out of the dynamic range of accuracy.

As a result, a semi-quantitative analysis was performed by categorising a range of normalised iMFI as either low, medium or high expression levels. Normalised iMFI ranging from 1 to 3000 was categorised as low expression, whereas 3001 to 22500 was categorised as medium expression and 22501 to 100000 was categorised as high expression. These thresholds were designed using SGV co-expression as a reference. As shown in Figures 4-26, 27, 28 and 29, the predicted expression levels were based on the verified transcriptional power of the promoters and the observed was derived from the threshold rules stipulated. The predicted and observed expression levels were matched for majority of positions within the library of 27 MGEV variants. Therefore, generally gene expression was successfully titrated and semi-predictable in a MGEV by the application of defined synthetic promoters.

4.2.8. Identifying transcriptional repression trends within the context of a MGEV

As the results have shown, there was substantial variation in fluorescent reporter gene expression demonstrating variable transcriptional activity of the synthetic promoters. Therefore, the next goal was to discern any specific trends in transcriptional repression which was observed across the library of 27 MGEV variants. This was achieved by comparing the RTA observed within a MGEV against the average RTA observed from separate SGVs co-expressed at equimolar ratios (as described Figure 4-5) at similar gene copies to the MGEV transfected DNA load, hereby referred to as expected RTA. The main observation from this comparison showed that overall gene expression within a MGEV was 70% lower. This indicates a substantial repression in transcriptional activity within the context of a MGEV.

As discussed in 4.2.6, a positional effect was apparent within the polycistronic cassette. Through interrogating the RTAs derived from expression of the 27 MGEV variants (shown in Figure 4-21 to 4-24), an overall positional effect ratio was identified as 1 : 0.85 : 0.86 for position 1 : position 2 : position 3 respectively. This ratio was derived by calculating the mean RTA in each position of the polycistronic cassette across all 27 MGEVs and comparing the average RTA of position 2 and 3 relative position 1. In order to identify specific promoter-related repression trends, the

observed RTA within the MGEV library was compensated for overall positional effects by accounting for the 15% and 14% repression in position 2 and 3 respectively. After adjusting for the overall positional effect, the MGEV and expected RTA were compared generating a percentage of transcriptional repression for each position and promoter combination within the library of MGEV variants. A frequency distribution of the percentage repression was constructed and shown in Figure 4-30.

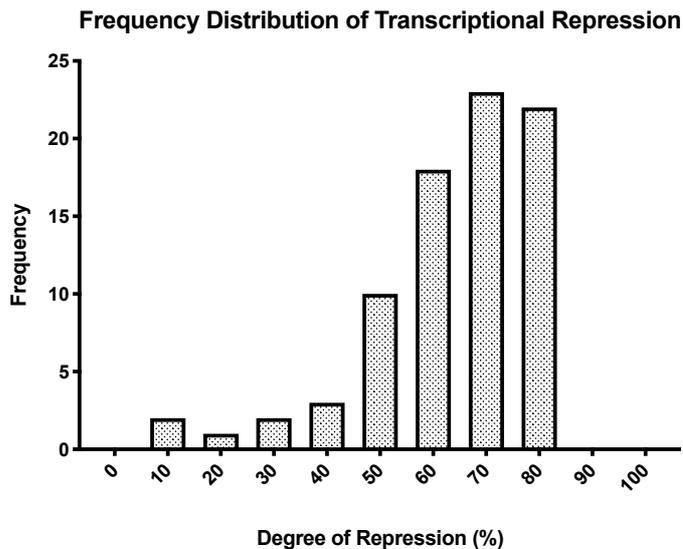


Figure 4-30: Frequency distribution of transcriptional repression within the MGEV library.

The percentage of transcriptional repression was determined by comparing the difference between the MGEV RTA normalised for average positional effect against the expected RTA. The degree of repression was then categorised into a fixed interval ranging from 0 to 100% as individual bin centers. The frequency of percentage repression within the MGEV library was quantified and presented as a distribution.

The frequency distribution shows that majority of the transcriptional activity within the MGEV library is substantially repressed where the median transcriptional repression is 68.6%. This indicates that the close proximity of the TUs within a MGEV substantially inhibits the transcriptional activity of the synthetic promoters. To further identify position specific and synthetic promoter pairing repression trends a colour gradient heat map was constructed to depict the percentage of transcriptional repression (derived from comparing MGEV and expected RTA) in each position of the polycistronic cassette across the MGEV library (Figure 4-31).

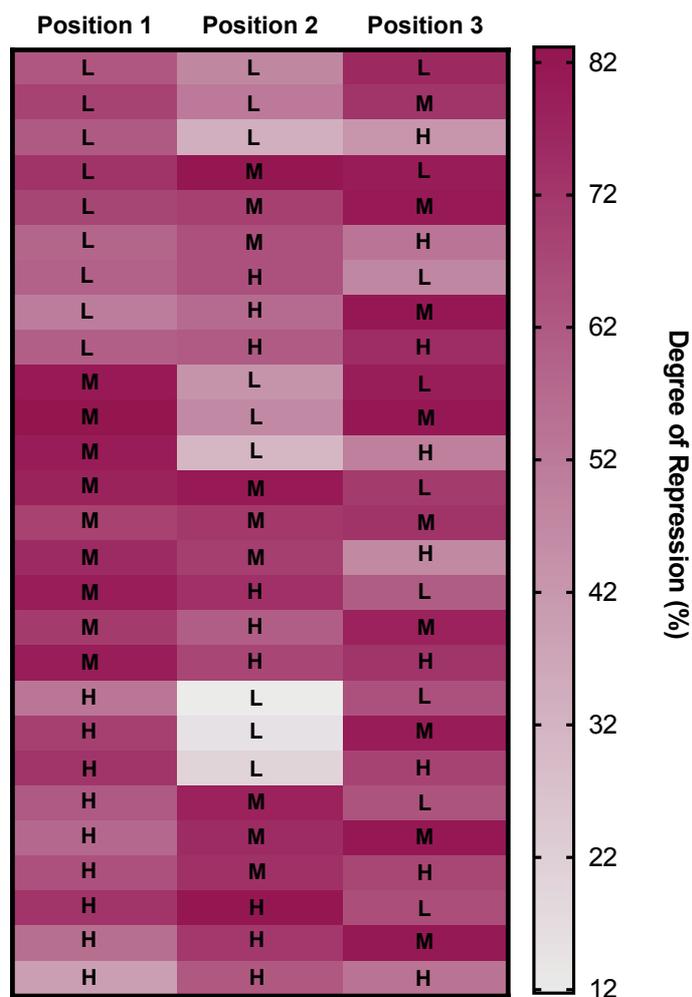


Figure 4-31: Position specific trends in transcriptional repression within the MGEV library.

The percentage of transcriptional repression was derived for every position of each MGEV variant tested. A color gradient heat map was constructed to depict the degree of repression for each MGEV variant. Shades of purple represent high repression, conversely shades of grey represent lower repression. The synthetic promoter utilised in the specific position is overlaid and abbreviated as “L”, “M” and “H” representing low, medium and high strength synthetic promoters.

The heatmap depicts that the medium strength synthetic promoter is consistently more repressed across all positions (shown by the darker shade of purple). The median percentage of repression for the medium strength synthetic promoter across all positions was 79.7%. Conversely, the low strength synthetic promoter exhibited enhanced transcriptional activity when neighbouring a higher strength synthetic promoter particularly in position 2, as shown by the lighter shade of purple/grey in heat map (Figure 4-31). Moreover, the mean transcriptional repression of the low strength synthetic promoter when neighbouring a higher strength synthetic promoter was 50.6% (which is 18% lower than the median transcriptional repression across the entire library). The high strength synthetic promoter-mediated expression did not show specific trends but boarder context-specific variation. Therefore, these observations

suggest synthetic promoter activity was influenced by interactions between variants. This could be caused by promoter squelching where competition of TFs and associated cofactors between synthetic promoter variants could inhibit transcriptional activity (Huliák et al., 2012; Schmidt et al., 2016).

4.3. Discussion

A MGEV that can demonstrate predictable titrated gene expression is advantageous in many aspects including as a tool to quantify gene expression dynamics, to enhance biopharmaceutical production and engineer mammalian cells. Furthermore, this benefits wider applications towards gene therapy and other biomedical-based treatments. Therefore, this chapter explores the application of synthetic promoters in a MGEV to titrate gene expression and quantify stoichiometric ratios in an aim to design a tool for more precise gene expression.

The first step in the study was to adapt the synthetic promoter library developed by Dr. Adam J. Brown to be compatible with the TU and MGEV system. This involved successfully PCR modifying and amplifying the proximal TFREs by introducing restriction sites. The BamHI and NheI restriction sites facilitate the interchangeability of the synthetic proximal element within the standardised TU. From the library, 10 out of the 14 synthetic promoters were successfully modified and cloned into the TU-1, 2, and 3 cassettes for MGEV construction and verified by restriction digestion colony screen. This further reiterates that the ‘plug-and-play’ feature within the TU (described in chapter 3) is functional for effective modification of the expression cassette.

As shown in Figure 4-1, the TU was designed to interchange the proximal region of a promoter (by standard restriction digestion-ligation cloning) and not the core element as it was adjoined with Medimmune’s proprietary 5’ UTR. Therefore, only the synthetic proximal element could be cloned into the TU, not the minimal core promoter as it would require to swap the 5’ UTR. This was suboptimal as the source plasmid’s 5’ UTR was uncharacterised compared to MedImmune’s proprietary 5’ UTR. As a result, the synthetic proximal element was combined with a hCMV-MIE core element. Finally, the library was curated for the synthetic promoters which were compatible with the Golden Gate assembly method (do not comprise of a BsaI site within the proximal element). Due to previous synthetic promoter transcriptional

activity being defined using a different reporter during *in vitro* testing (Brown et al., 2017) and the proximal element was cloned with a different core element than previously tested, the final set of promoters required further validation using the fluorescent protein reporters (*eGFP*, *mCherry* and *tagBFP*) to characterise promoter performance.

The synthetic promoters verified were the 5RPU, 20RPU, 40RPU, 60RPU and 80RPU. A gradient of expression was observed; however, the expression fold changes differed between the three fluorescent proteins. Additionally, the marginal difference in iMFI between 60RPU and 80RPU may infer accumulation of fluorescent protein within the cell's cytoplasm leading protein turnover and degradation. The verification data did confirm a promoter exhibiting low, medium and high level of expression. Therefore, the promoters selected were 5RPU, 40RPU and 80RPU. The 80RPU was selected over the 60RPU based on historical data, even though some unexpected variability in *eGFP* expression was observed. The reduced *eGFP* expression by the 80RPU promoter could be the result of aggregation and accumulation of *eGFP* intracellularly. Literature has shown that *eGFP* stability is pH and protein concentration dependent (Krasowska et al., 2010), therefore, unfavourable conditions within the cell cytoplasm may contribute towards the lower than expected expression.

Following from adapting the genetic components to control gene expression, two methodologies were developed to quantify the titration of gene expression. The methods selected as techniques to measure differential gene expression at a transcriptional and protein level were qPCR and flow cytometry respectively. The qPCR assay was developed by designing efficient primer pairs to quantify cDNA of *eGFP*, *mCherry*, *tagBFP* and *GS* as a function of mRNA transcript. Therefore, the assay quantified transcriptional activity of the four genes within a MGEV. Furthermore, the qPCR assay was concluded to have a sufficient dynamic range of detection and could accurately quantify differential gene expression at a transcriptional level. The data indicated that expression using the low strength synthetic promoter was less accurate compared to the medium and high strength promoters. This may arise from very low mRNA transcripts present within the transfected cell, hence lying on the lower limit of the dynamic range of quantification.

Similarly, the same analysis was performed for the flow cytometer, whereby expression was measured by iMFI. The quantification analysis demonstrated a sufficient dynamic range of gene expression when regulated by a medium and high strength promoter with poorer accuracy for low promoter strength-mediated expression. Furthermore, the expression of fluorescent protein as a function of promoter strength was less accurate than qPCR when measuring gene expression titration.

Therefore, the qPCR methodology provides more accurate absolute quantitative data on gene expression and was ideal for measuring gene expression dynamics and titration using synthetic promoter components. The flow cytometry provided a semi-quantitative analysis of gene expression trends and could be used as an orthogonal preliminary HT screen due to its lower labour intensive requirements. Additionally, the flow cytometer can be used as a means to measure protein levels when influenced by other genetic elements, such as 5' and 3' UTRs.

To achieve predictability, the understanding and characterisation of simultaneous gene expression driven by identical or different promoter variants was required. To investigate whether cross reactivity was observed within the synthetic promoter library, a low, medium and high strength synthetic promoter was selected. The promoters were co-expressing equimolar ratios of *eGFP*, *mCherry* and *tagBFP* as separate SGV plasmids. The self-interaction was measured by increasing the promoter units (in the form SGVs encoding a different fluorescent reporter by intervals of 200 ng of transfected DNA) by 2- and 3-fold. The cross-interaction was measured by combining different promoter variants. The results showed repression in fluorescent reporter gene expression when multiple units of promoters were expressing different reporters simultaneously (measured by flow cytometry). The promoter variants exhibiting the highest and lowest average repression by both self and cross-interaction was the low and high strength synthetic promoters respectively. The medium strength synthetic promoter average repression was in between this range. The decreased gene expression was hypothesised to be a consequence of promoter squelching and increased biosynthetic burden from expressing multiple proteins (Brown et al., 2017; Glick, 1995; Ley et al., 2015; Yusufi et al., 2017).

Promoter squelching is defined as the repression of transcription caused by sequestering of enhancer or basal TFs leading to competition for TF-TFRE interaction. This is an alternative form of transcriptional interference (Cahill et al., 1994). One proposed mechanism of promoter squelching and the repression of gene expression could be due to the competition of essential TFs or any accessory cofactors to initiate transcription (Curtin et al., 2008; Huliák et al., 2012). Alternatively, interference could also be caused by post-transcriptional limitations such as competition for ribosomes during translation (Curtin et al., 2008). However, the exact mechanism has not been fully defined and understood, hence it remains a challenge to account for promoter interference during the design of synthetic libraries. Additionally, a potential means to more accurately quantify promoter squelching and decipher from biosynthetic burden is by using qPCR-mediated quantification.

The positional effect within a MGEV was first quantified in chapter 3, whereby relatively consistent repression of gene expression was observed between MGEV variants regulated by a hCMV-MIE promoter. The following step was to determine if the positional effect was constant or transcriptional activity dependent. The positional effect was quantified by qPCR and flow cytometry by measuring expression of *eGFP*, *mCherry* and *tagBFP* regulated by a low, medium and high strength synthetic promoter. The results indicated that the positional effect was dynamic and transcriptional strength dependent. As shown by qPCR, the RTA fold change relative to the SGV controls in position order of 1, 2 and 3 for the low strength promoter was 0.42: 0.49: 0.19, medium strength promoter was 0.57: 0.44: 0.41, and high strength promoter was 0.67: 0.42: 0.53 respectively. Therefore, showing the level of repression was non-constant. The flow cytometry data showed decreased expression, however the fold change and the position most affected differed to the qPCR data for MGEV-GCB-Low and High. The difference for the MGEV-GCB-Low could be due to limitations in accuracy of quantification by both qPCR and flow cytometry. Whereas, for the MGEV-GCB-High, the difference could be due to post-transcriptional limitations (as mentioned in section 4.2.5).

The positional effect was hypothesised to be a consequence of transcriptional interference. The qPCR quantification infers this was true due to the variation in mRNA copies during transcription. The dynamic nature of transcriptional interference is a product of various gene repression mechanisms. For example, transcriptional interference comprises of RNA pol II occlusion-mediated interference, TF or pre-initiation complex (PIC) dislodgement (Palmer et al., 2011; Shearwin et al., 2005) and promoter squelching alongside negative supercoiling (Curtin et al., 2008).

Moreover, the range of transcriptional activity from low to high in position 1 showed an increase in repression in position 2. This was theorised to be a consequence of increased recruitment of TFs and RNA pol II leading to increased promoter squelching in position 2 and 3 coupled with occlusion in the downstream promoter region of the position 2 TU, hindering initiation of transcription. Conversely, in position 3, under low transcriptional activity the repression was highest due to a cumulative effect of interference caused by position 1 and 2. Whereas, under high transcriptional activity, the impact of interference in position 2 may have led to lower levels of RNA pol II occlusion due to reduced recruitment of the upstream gene. However, decreased gene expression was still observed which could be caused by promoter squelching.

As a result of the complexity and multiple mechanisms contributing towards transcriptional interference, the positional effect is non-predictable. Therefore, to achieve predictability of gene expression titration using a synthetic promoter library, the transcriptional interference has to be sequestered. Publications have theorised the use of transcription terminators to avoid RNA pol II occlusion (Eaton et al., 2018; Palmer et al., 2011; Proudfoot, 1991). Alternatively, publications have explored and applied insulator elements to avoid promoter squelching, RNA pol II occlusion and negative supercoiling by maintaining the TU in a euchromatin state (Hasegawa and Nakatsuji, 2002; Liao et al., 2018; Torella et al., 2014b; Yahata et al., 2007). Both these elements would assist by introducing modulation within the MGEV and provide the ability to predict gene expression stoichiometric ratios.

The question remained, can gene expression be titrated, and stoichiometric ratios derived using synthetic promoters with varying transcriptional strength? This was addressed by using a low, medium and high strength synthetic promoter and varying the combinations to generate a MGEV library of 27 variants co-expressing *eGFP*, *mCherry* and *tagBFP*. The 27 variants were quantified by qPCR and semi-quantitatively by flow cytometry. Both assays demonstrated that gene expression was titrated dependent on the transcriptional activity of the synthetic promoter. Additionally, the *GS* mRNA copies were quantified upstream of the polycistronic cassette of the 27 variants. The stoichiometric ratios were generated using the RTA values from qPCR forming a database of observed gene expression performance within each MGEV variant when using a discrete combination of synthetic promoters in specific positions within the polycistronic cassette.

However, the data also indicated variable transcriptional activity of the synthetic promoters within a MGEV context. As a result, the MGEV library's RTAs were compared against a set of expected RTAs (derived from co-expression of separate SGVs encoding the fluorescent reporters at equimolar ratios and similar gene copies). This comparison identified that overall transcriptional activity was substantially repressed by 70%. This overall repression in transcription within the MGEV context could be a consequence of negative supercoiling. Negative supercoiling is a change in the plasmid structure left in the wake of RNA pol II-mediated transcription elongation resulting in transcription inhibition of the upstream genes (Corless and Gilbert, 2017; Curtin et al., 2008; Ma and Wang, 2016). Another cause for transcriptional repression could be the potential inherent bidirectional behaviour of promoters, where transcription can occur both on the sense and antisense strand (Seila et al., 2008; Wei et al., 2011). This leads to promoter interference by inhibiting transcription of neighbouring TUs (Curtin et al., 2008).

Additionally, specific trends in synthetic promoter activity repression was also observed. The medium strength synthetic promoter within the MGEV context was the most transcriptionally repressed out of the three synthetic promoters utilised. When referring to the TFRE composition of the synthetic promoters (refer to Figure 4-32), the medium strength promoter had common TFREs between both the low and high strength promoters. Therefore, the medium strength synthetic promoter could be

squelched due to competing for the cognate TFs with the other synthetic promoter variants (Cahill et al., 1994; Huliák et al., 2012; Schmidt et al., 2016).

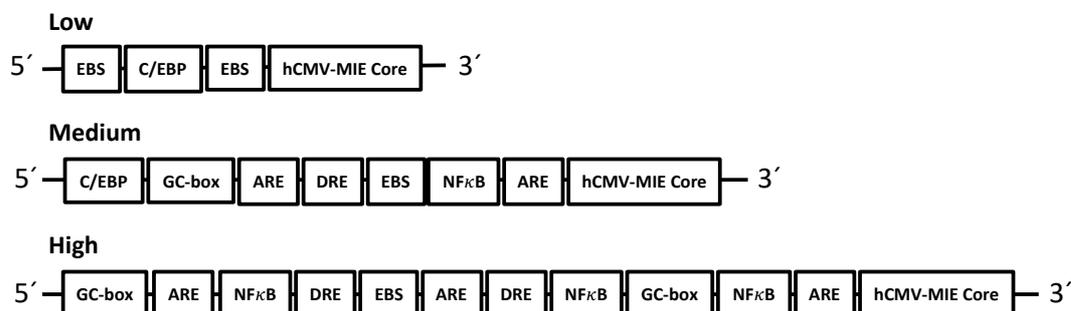


Figure 4-32: The transcription factor regulatory element (TFRE) composition of the synthetic promoters tested in the multi-gene expression vector library.

The schematic describing the TFRE composition of the low (5RPU.1), medium (40RPU.2) and high (80RPU.1) strength synthetic promoters. The TFREs within each promoter were specifically selected to vary the promoter's transcriptional activity and separated by a 2 bp spacer. The promoters were positioned upstream of the hCMV-MIE core. The list of TFRE abbreviations are as follows – Antioxidant regulatory element (ARE), CCAAT-enhancer binding protein (C/EBP), Dioxin regulatory element (DRE), ETS binding site (EBS), GC-box, Nuclear factor kappa B (NFκB).

Interestingly, the low strength synthetic promoter demonstrated enhanced transcriptional activity when neighbouring a more transcriptionally active promoter particularly in position 2. This could be caused by synergistic functionality of TFs between the synthetic promoters. For example, literature has shown derivatives of the C/EBP and NFκB TFs functioning cooperatively to initiate transcription in immune cells (Maehara et al., 1999). Therefore, a potential hypothesis is the abundance and potency of NFκB driven transcription (Brown et al., 2015, 2017) within the medium and high strength synthetic promoters and interaction with the C/EBP TF could inadvertently enhance transcriptional activity of the low strength synthetic promoter within the MGEV context.

Therefore, these findings indicate that gene expression within a MGEV context is substantially impacted by a range of mechanisms. Therefore, it would be difficult to engineer or account for these mechanisms in order to achieve predictable gene expression dictated by the transcriptional strength of synthetic promoters within a MGEV. An alternative approach to achieving precise and predictable stoichiometric gene expression ratios is by characterising and quantifying every variant in the MGEV library. The data is then compiled in a database of empirically derived approximate ratios which can be employed to suggest a combination of synthetic promoters and

their respective positions in the polycistronic cassette. These ratios can be applied towards precise recombinant gene expression of an optimal *LC* and *HC* ratio alongside an optimal dose of an accessory gene, such as *XBP-1s*, to enhance recombinant mAb expression in CHO cells. Therefore, this study has shown that expression of multiple recombinant genes can be controlled simultaneously using synthetic promoters of defined transcriptional activity within a MGEV.

Chapter 5

5. Proof of Concept Applications of Multi-gene Expression Vectors as a Research Tool

A multi-gene expression vector (MGEV) is a synthetic biology tool with potential applications in various areas of the biological sciences. This chapter demonstrates the application of a MGEV in synthetic biology and the biopharmaceutical industry. One proof of concept study was using a bicistronic MGEV co-expressing secreted alkaline phosphatase (SEAP) and Cypridina luciferase (CL) to measure the capacity of gene regulation and modularisation of transcription units (TUs) using synthetic transcription terminators. This was achieved by designing a library of terminator elements based on knowledge derived from literature and screening these elements (36) using high throughput (HT) transient expression. The data demonstrated one synthetic polyadenylation variant to be functional based on preliminary screens and identified favourable design features for a 2nd generation library. A second proof of concept study was stably engineering the CHO cell by co-expressing a difficult-to-express (DTE) monoclonal antibody (mAb) and x-box binding protein-1 spliced (XBP-1s) to enhance mAb expression. The data demonstrated a 37.3% increase in mAb expression and increased growth capacity of the stable pool. Furthermore, qPCR analysis showed increased XBP-1s expression indicating functionality of the MGEV. However, some variability in XBP-1s expression was also observed particularly when utilising a higher strength promoter. A potential hypothesis is that the stable pool may be self-regulating gene expression during selection by silencing recombinant XBP-1s expression to avoid toxicity and maintain survivability.

Acknowledgements

I would like to acknowledge Danielle Meehan who assisted and collaborated in the *in silico* design of the transcription termination library. Furthermore, I would like to acknowledge Dr. Susie Sou and the rest of the Cell Line Development team at MedImmune for the collaboration study of stably engineering the CHO cell to recombinantly express a DTE mAb and XBP-1s, particularly performing the fed-batch overgrow study, and measuring the growth of the stable pools, mAb titres expressed and mRNA copies of XBP-1s.

5.1. Introduction

A MGEV is a synthetic biology tool with a broad range of potential applications but not limited to development of orthogonal synthetic biology components, cell engineering and biopharmaceutical production (Duportet et al., 2014; Guye et al., 2013). Data in Chapter 4 has shown that gene expression can be titrated in a MGEV using validated synthetic promoters with a defined level of transcriptional activity. However, it was also observed that positional-mediated repression of recombinant genes was occurring within a polycistronic cassette. Therefore, after characterising the gene expression dynamics within a MGEV, the following step was to explore the functionality. A number of theorised applications of a MGEV have been suggested throughout the thesis. In this chapter, two proof of concept approaches demonstrating the capacity of a MGEV were explored. One aspect is in the area of synthetic biology and how a MGEV assists in the design and screening of synthetic genetic elements. Another aspect explores the functionality of the MGEV as a tool for engineering CHO cells.

5.1.1. Design and screening of synthetic transcription terminator elements

Previous publications have inferred that positional-based gene repression is observed within a MGEV by utilisation of insulator elements (Torella et al., 2014b). Furthermore, more extensive characterisation in Chapter 4 has shown that the degree of repression is in a state of flux and dependent on relative transcriptional activity. Through literature this occurrence is rationalised as a consequence of transcriptional interference (Eszterhas et al., 2002; Shearwin et al., 2005). Transcriptional interference is a cumulation of multiple mechanisms that contribute to reduction of downstream gene expression. The mechanisms include RNA polymerase II (RNA pol II) occlusion,

transcription factor (TF) or pre-initiation complex (PIC) dislodgement (Palmer et al., 2011; Shearwin et al., 2005) and promoter squelching by competition of TFs and associated cofactors (Curtin et al., 2008; Huliák et al., 2012; Schmidt et al., 2016).

West et al have previously investigated the process of transcription termination, specifically the development or application of elements which improve the efficiency of termination in human and mouse cells (West and Proudfoot, 2009; West et al., 2006). An efficient transcription termination element can assist in modularisation of TUs within a MGEV by mitigating RNA pol II occlusion and TF or PIC dislodgement mediated transcriptional interference. This would be achieved by the efficient displacement of the RNA pol II from the sense strand, as well as contributing towards gene expression stability and cellular localisation of mRNA post-transcription (Gasnov et al., 2015; Porrua and Libri, 2015; Proudfoot, 2016).

The specific mechanism of transcription termination is still debated and not fully understood. Currently, there are two proposed models of termination, the allosteric and the torpedo model, as described in section 1.7.2. The allosteric model is a conformational change of the RNA pol II after interaction with the polyadenylation signal (PAS) leading to displacement of the enzyme. The torpedo model is the interaction between the 5'-3' exonuclease, Xrn2, and the RNA pol II by digesting the nascent RNA, leading to displacement of the RNA pol II from the sense strand (Eaton et al., 2018; Porrua and Libri, 2015). For instance, research has shown transcription terminator elements, such as the human β -globin or mouse serum albumin (MSA) terminators, improving transcription termination efficiency in mammalian cells using the torpedo model (Gasnov et al., 2015; West and Proudfoot, 2009; West et al., 2006). Alternatively, polyadenylation (pA) elements have also been shown to terminate transcription using a SV40 pA element (Gasnov et al., 2015).

However, the major limitation of existing elements from mammalian or viral genomes is the length of the sequence. For example, the human β -globin terminator element is 800 bp (West and Proudfoot, 2009), the MSA terminator element is 2.4 kb (West et al., 2006) and the SV40 late pA is 240 bp. Additionally, all of these sequences are poorly defined apart from the known conserved functional motifs. Therefore, the full functionality of these elements is undetermined and could cause the transcription

termination process to be unpredictable. Conversely, synthetic transcription terminator elements are shorter and consist of only functional domains contributing towards transcription termination. These components have been developed for *E.coli* (Chen et al., 2013) and yeast (Curran et al., 2015). The advantage of synthetic elements is the reduced reliance on native sequences, in turn reducing homology-based undesired recombination with the host genome (Curran et al., 2015).

Currently, the only published data on the design and development of functional synthetic terminator elements for mammalian cells was demonstrated in the context of a single gene plasmid (Cheng et al., 2019). There is a demand for such elements to assist in achieving increased predictability of gene expression within a MGEV by attaining modularity. Moreover, there are a number of advantages of using MGEVs as a tool to screen synthetic transcription terminator elements – i) the rapid and efficient construction of a MGEV compared to traditional cloning techniques allows for a larger number of constructs to be assembled, in turn facilitating screening of larger libraries; ii) the ‘plug-and-play’ feature allows ease of interchangeability of variants to minimise cost of synthesis; iii) the standardised TU cassette composition controls other expression variables, making it a better tool to measure individual component functionality.

As a result of the demand and capabilities of a MGEV, it was selected as a HT tool to screen a library of transcription terminators including both native and synthetic derivatives, to identify functional synthetic variants to control gene expression. This was achieved by *in silico* design and synthesis of a library of native and synthetic transcription terminator elements using literature. The library of MGEVs were constructed for screening using a 2-gene expression cassette co-expressing *SEAP* and *CL*. The functionality of the variants was tested by measuring the transient expression of *SEAP* and *CL* within the bicistronic cassette.

5.1.2. Engineering mammalian cells to enhance cellular phenotype and monoclonal antibody production

There are limitations in mammalian cell engineering specifically to engineer CHO cells for biopharmaceutical production. For example, various publications have shown the co-expression of transactivators such as *XBP-1* and *ATF6*, chaperones or foldases

like *BiP*, *CypB* and *PDI*, when combined with a DTE recombinant therapeutic to synergistically increase production titres (Johari et al., 2015; Pybus et al., 2014).

Currently, approaches to increase biopharmaceutical production by stably engineering the CHO cell factory has been achieved by performing multiple cell line constructions. For example, a two-step sequential stable cell line construction process was performed to over-express *XBP-1s* followed by *Ero1 α* . This engineered host was then used to increase transient expression of a mAb (Cain et al., 2013). An alternative approach is performing genome editing of the CHO cell factory by CRISPR/Cas9 including stably expressing the mAb (Lee et al., 2018). However, there are still limitation in the sgRNA to achieve multiplexing and multi-gene expression (Tian et al., 2017). Therefore, designing and developing a MGEV that would allow for concurrent stable multi-gene engineering of the CHO cell factory by co-expressing a recombinant protein and an enhancing accessory gene would be beneficial. Furthermore, the application of synthetic promoters to titrate gene expression (as shown in Chapter 4) would facilitate precise gene expression of multiple recombinant genes.

XBP-1s is an enhanced TF with increased efficacy from its unspliced state. This is achieved by the removal of a 26 bp intron in the *XBP-1* mRNA by the IRE1 element. *XBP-1s* is a global TF which is known to contribute towards activation of the unfolded protein response (UPR) and endoplasmic reticulum stress (ERS) elements. *XBP-1s* has also contributed towards secretion and controlling cell survival during induced stress (Tigges and Fussenegger, 2006). Publications have shown that transient co-expression of *XBP-1s* and a DTE mAb have increased specific productivity (qP) of the mAb (Pybus et al., 2014). Other studies have shown the stable co-expression of *XBP-1s* and *Ero1 α* or only *XBP-1s* have improved mAb expression within CHO cells by modifying the ER (Becker et al., 2008; Cain et al., 2013). More recently, a study performed by Cartwright et al showed that a titration of *XBP-1s* gene copies had a correlation with increased transient expression of a DTE mAb such that the optimal gene dose increased expression by 237% (Cartwright et al., 2020).

Therefore, a MGEV was designed to co-express a DTE mAb and *XBP-1s*, whereby the heavy chain (*HC*) and light chain (*LC*) genes are driven by a hCMV-MIE promoter and the *XBP-1s* gene was titrated using a range of synthetic promoters. This MGEV

was then used to stably engineer the CHO cell. To accurately observe the dose response of *XBP-1s* upon mAb expression, the variation of integrated copies was controlled by using MedImmune's proprietary targeted integration (TI) cell line. This was achieved by first adapting the pExp-Vec-GG recipient vector to be compatible with the TI system. Four prototype MGEVs co-expressing a DTE mAb and titrated amounts of *XBP-1s* were constructed. These prototypes were tested by TI to stably engineer MedImmune's proprietary CHO cell line. The successfully established cell lines were then characterised for growth, mAb titres and qP. Finally, the mRNA quantification of *XBP-1s* was performed as a preliminary screen to see relative expression levels within the cell line.

5.2. Results

5.2.1. Design and screening of synthetic transcription terminator elements using a multi-gene expression vector

The results of the *in silico* design process, synthesis, MGEV construction and screening of the native and synthetic transcription terminators by co-expression of *SEAP* and *CL* is described in this section.

5.2.1.1. Design of the synthetic transcription terminator element library

The aim was to investigate the transcription termination process in mammalian cells and design synthetic elements with a goal to:

- i) Design elements which are shorter than native sequences but maintain functionality to terminate transcription.
- ii) Be able to use the synthetic element to regulate recombinant gene expression.
- iii) Facilitate modularity of TUs within a MGEV, in an aim to predict expression of multiple recombinant genes.

The design process began by dissecting the transcription termination process and identifying the known mechanisms involved. This highlighted that transcription termination involves the synergistic functionality of polyadenylation and termination. A range of publications have explored different approaches to termination including polyadenylation (pA), co-transcriptional cleavage (CoTC), pause and ribozyme-mediated transcription termination, as described in section 1.7.2.1, 2, 3 and 4

respectively. A combination of literature and known native transcription termination elements were used to explore each of these approaches to identify the conserved functional motifs within the DNA sequence. Therefore, generating a design space for a pA, CoTC, pause and ribozyme element. The summary of the design space is shown in Table 5-1.

Table 5-1: Identification of conserved functional groups of transcription terminators and the design space for synthetic elements.

The combination of literature research and molecular dissection of native transcription terminators had highlighted the conserved functional motifs and the design space for synthetic pA, CoTC, pause and ribozyme elements.

Mechanism of Transcription Termination	Conserved Functional Elements	Design Space
Polyadenylation (pA)	<ul style="list-style-type: none"> - Upstream element motif – ‘TGTA’. However, the remainder of the sequence is less conserved. - PAS motif – ‘AATAAA’ with an alternative of ‘ATTAAA’. - Downstream element is a motif of a G/T rich sequence. (Colgan and Manley, 1997; Elkon et al., 2013) 	<p>Using the conserved functional domains, the following features could be investigated:</p> <ul style="list-style-type: none"> - Number of PAS motifs within the element - Length of spacer nucleotides between PAS motifs - Length of G/T rich motif - Number of G/T rich motif - G/T rich motif sequence variants - Length of spacer nucleotides between PAS and G/T rich motifs - Omission of upstream ‘TGTA’ motif
Co-transcriptional cleavage (CoTC)	<ul style="list-style-type: none"> - Long A/T rich motifs within the sequence of β-globin terminator element. - A/T tracts were found in 70% of putative terminator regions within the genome. - A conserved motif was identified as [A/T(5)+G/C(1)+N(0-35)]x8 – A/T(5) (West et al., 2008; White et al., 2013) 	<p>Using the conserved motif described, two features can be varied:</p> <ul style="list-style-type: none"> - A/T rich motif length - Length of spacer nucleotides between repeat motifs
Pause	<ul style="list-style-type: none"> - MAZ element – is the binding site for the TF MAZ and has shown to pause RNA pol II. - The MAZ element motif is G(5)AG(5). - Four MAZ motifs positioned downstream of the PAS showed termination. 	<p>Using the conserved motifs, the following features can be varied:</p> <ul style="list-style-type: none"> - Number of MAZ motifs - Distance of MAZ motifs from the PAS

Mechanism of Transcription Termination	Conserved Functional Elements	Design Space
Pause	<ul style="list-style-type: none"> - Proximity of the MAZ motif from the PAS influenced termination. (Gromak et al., 2006) 	
Ribozyme	<ul style="list-style-type: none"> - Hammerhead ribozyme is a well characterised ribozyme - Shown to cleave mRNA transcript with no PAS - Use ribozyme to efficiently cleave pre-mRNA transcript to allow for Xrn2 mediated termination (Dower et al., 2004) 	Use a hammerhead ribozyme element downstream of a PAS

In the aim to design efficient transcription terminators, other commercially available pA elements were identified to compare against the currently employed SV40 late pA element. Additionally, CHO-K1 homologous pA and CoTC elements were identified downstream of the top 3 transcriptionally active genes in the CHO-K1 genome. The hypothesis is that an endogenous sequence downstream of a highly transcriptionally active gene may be more favourable and lead to efficient transcription termination. Therefore, the transcription termination library comprises of 5 categories – 1) Commercially available pA elements; 2) Synthetic pA elements; 3) Synthetic CoTC-mediated terminators; 4) CHO-K1 pA and CoTC homologues; 5) Synthetic pause-mediated & ribozyme-mediated terminators.

Group 1 – Commercially available pA elements

The industry standard pA element employed in biopharmaceutical production is the SV40 late pA. However, through literature there are a range of commercially available pA which were identified. The list of alternative pA elements that were tested are as follows:

- Viral-derived pA elements
 - SV40 late
 - Herpes simplex virus (HSV) thymidine kinase (tk)
 - Human immunodeficiency virus (HIV)-1
 - Adenovirus L1
- Mammalian-derived pA elements
 - Lambda LC
 - Human growth hormone (hGH)
 - Bovine growth hormone (bGH)
 - Chinese hamster elongation factor-1 α (chef1 α)

Figure 5-1 summarises the composition of each of the commercially available pA elements.

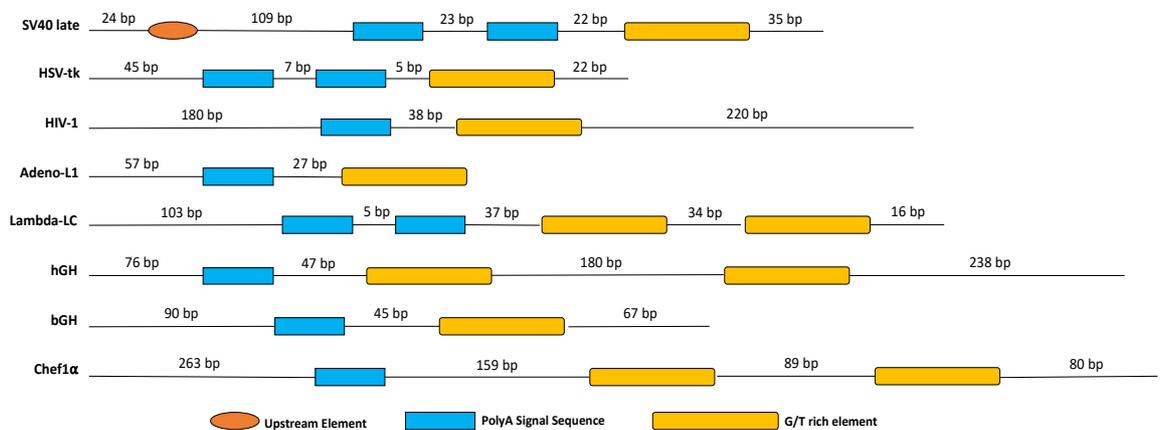


Figure 5-1: Design schematic summarising the commercially available polyadenylation variants synthesised and tested.

Graphic showing a schematic summarising the commercially available pA variants that were *de novo* synthesised and tested. The composition of each pA is summarised by identifying the upstream element, polyA signal and G/T rich motif.

Group 2 – Synthetic pA elements

Using the design space highlighted in Table 5-1, a library of pA elements were designed comprising of the essential functional motifs. Ten synthetic pA elements were designed as shown in Figure 5-2. The library was designed using an SV40 late pA as a template. The design features tested included the number of G/T rich motifs, different G/T rich motif variants, nucleotide spaces between multiple motifs and the presence of an upstream element motif, while maintaining two PAS motifs in each variant.

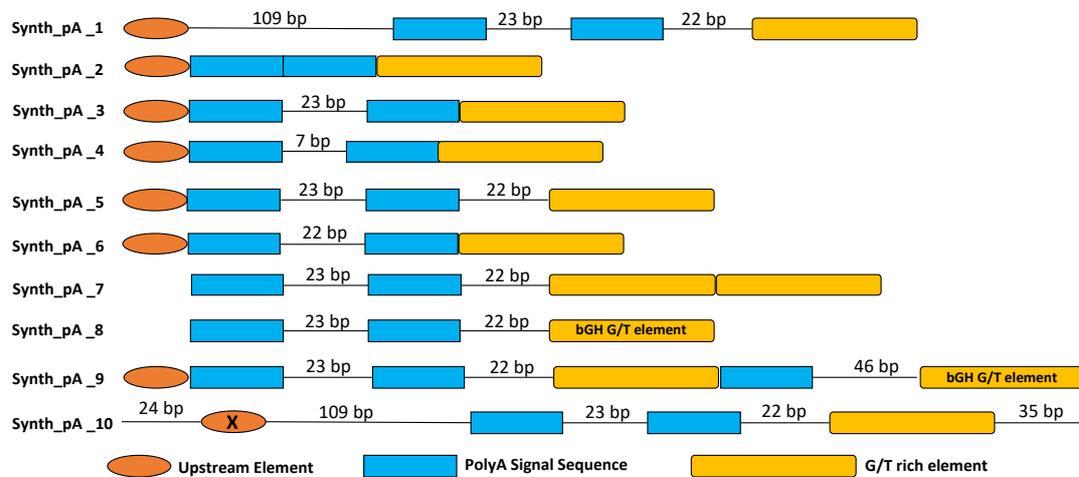


Figure 5-2: Design schematic summarising the synthetic polyadenylation library synthesised and tested.

Graphic showing a schematic summarising the pA library designed and *de novo* synthesised based on the template and conserved functional motifs within an SV40 late pA element. The library tests a number of features including number and different variants of G/T rich motifs (SV40 and bGH variants), nucleotide spacers between motifs and the presence of an upstream motif.

Group 3 – Synthetic CoTC elements

A library of 12 synthetic CoTC elements were designed and synthesised using the conserved functional motifs described in Table 5-1. The elements comprise of A/T tracts, a G/C and nucleotide (N) spacers based on the following design consensus sequence– [A/T(5) + G/C(1) + N(0-35)] x 8 A/T(5) derived from genome analysis (White et al., 2013). The variables changed within the library were length of A/T tracts and nucleotide spacers as shown in Figure 5-3. However, the *de novo* synthesis of the Synth_CoTC_10 element was unsuccessful due to large repeat units.

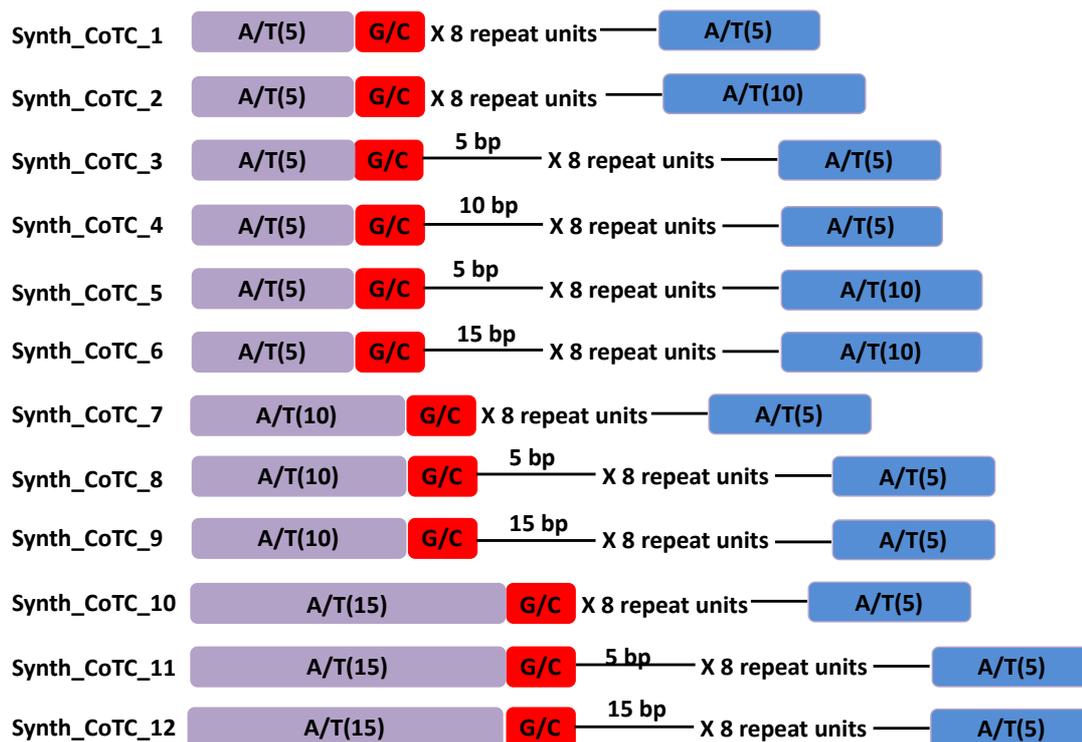


Figure 5-3: Design schematic summarising the synthetic co-transcriptional cleavage library synthesised and tested.

Figure showing a schematic summarising the CoTC library designed and *de novo* synthesised based on conserved motifs described in literature. The library varies the A/T tract repeat unit length, the final A/T tract length, and the nucleotide spacer within the repeat units.

Group 4 – CHO-K1 pA & CoTC homologues

It was hypothesised that transcription termination would be efficient in highly transcriptionally active genes as these genes are important within the host cell. This was rationalised by literature identifying a functional terminator downstream of essential genes such as the β -globin gene in human cells (West et al., 2008). Therefore, the top 3 transcriptionally active genes were *ActB*, *S100a6* and *Gadph* as identified by transcriptomic data. These genes were located in the CHO-K1 genome and the sequence directly 3' of the stop codon was analysed ranging from between 810 to 2000 bp downstream. The downstream sequence was screened for pA and CoTC functional motifs to indicate a presence of endogenous derivatives of these elements. A pA element was identified downstream of all three genes, however the *ActB* gene pA element was scattered with BsaI restriction sites rendering it non-compatible to the MGEV system. A CoTC-like element was found downstream of *ActB* and *S100a6*; once again, *ActB* was omitted due to the presence of multiple BsaI sites. Therefore, the following homologues were *de novo* synthesised and tested– S100a6_pA, Gadph_pA and S100a6_CoTC.

Group 5 – Synthetic Pause & Ribozyme elements

The pause and ribozyme-mediate transcription termination either slows down the RNA pol II during elongation or self-cleave the mRNA transcript to potentially initiate Xrn2-mediated transcription termination (Gromak et al., 2006). The synthetic pause elements were designed using a MAZ TF binding element downstream of a PAS. The variables tested were the number of MAZ motifs and the nucleotide spacer between the PAS and MAZ element. The synthetic ribozymes involved the addition of a hammerhead ribozyme element downstream of the PAS to trigger transcription termination. The library of synthetic pause elements is shown in Figure 5-4.

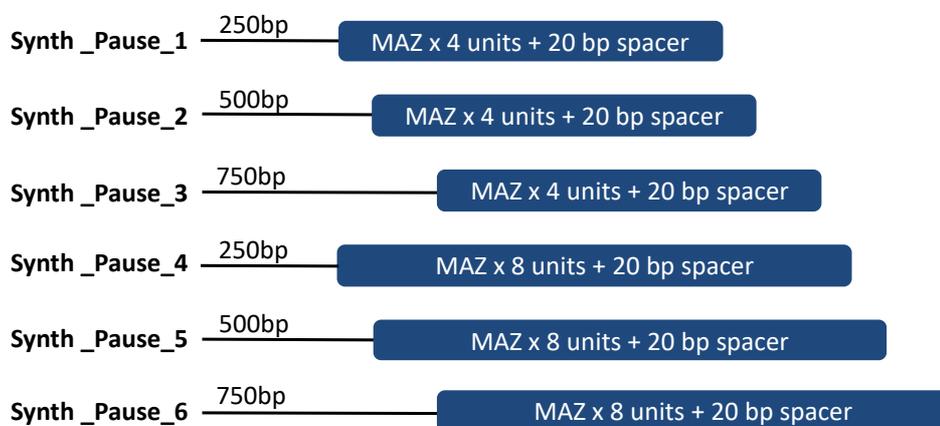


Figure 5-4: Design schematic summarising the synthetic pause-mediated transcription termination library synthesised and tested.

Figure showing a schematic summarising the pause-mediated transcription terminator library designed and *de novo* synthesised based on conserved motifs described in literature. The library varies the number of MAZ elements and the proximity from the pA element.

5.2.1.2. Construction of multi-gene expression vectors to screen the transcription terminator library

The approach utilised to screen the functionality of the terminators was using a 2-gene system within a MGEV to co-express *SEAP* and *CL*. Both of these recombinant proteins are commonly used gene expression reporters and are secreted out of the CHO cell. Figure 5-5 shows a schematic of the bicistronic cassette assembled and the loci for the pA or terminator element screened.

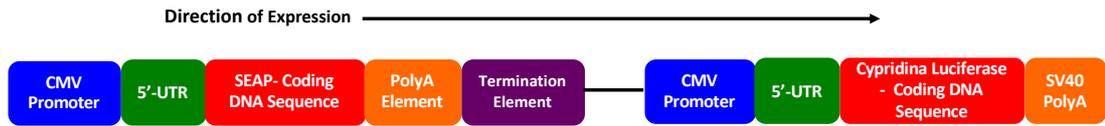


Figure 5-5: Schematic summarising the multi-gene expression vector used for terminator library screening.

The schematic shows the bicistronic cassette designed to screen the 5 subcategories within the terminator library for functionality and impact on gene expression. The expression of *SEAP* would indicate the impact on efficient 3' end processing and mRNA transcript generation. The expression of *CL* would indicate the impact on efficient transcription termination and the reversal of transcriptional interference. The MGEV also contains a *GS* expression cassette, mammalian episomal origin of replication, a β -lactamase gene and a bacterial origin of replication.

The functionality of the pA and terminator elements were tested by cloning them downstream of the *SEAP* gene and upstream of the *CL* TU. Each TU comprises of a hCMV-MIE promoter, 5' untranslated region (UTR) and an SV40 late pA element downstream of the *CL* gene. The *SEAP* expression within the bicistronic cassette would infer the effect of the pA and/or terminator element on 3' end mRNA processing. Whereas, *CL* expression would infer the efficiency of termination and potential modulation by reversing transcriptional interference.

The pA and terminator library was cloned into the *SEAP* expression cassette post-*de novo* synthesis by GeneArt. A total of 36 MGEV variants were designed to screen the functionality of the pA and terminator library as shown in Table 5-2.

Table 5-2: Library of 36 multi-gene expression vector variants for terminator library screen.
The table summarises the 36 MGEV variants designed to screen the pA and terminator library for functionality.

Transcription Terminator Category	MGEV Variant	Transcription Terminator Element
Industry Standard	2ST-Control	SV40 pA
Synthetic Polyadenylation Elements	2ST-1	Synth pA 1
	2ST-2	Synth pA 2
	2ST-3	Synth pA 3
	2ST-4	Synth pA 4
	2ST-5	Synth pA 5
	2ST-6	Synth pA 6
	2ST-7	Synth pA 7
	2ST-8	Synth pA 8
	2ST-9	Synth pA 9
	2ST-10	Synth pA 10
Synthetic Co-transcriptional Cleavage (CoTC) Elements	2ST-11	Synth CoTC 1
	2ST-12	Synth CoTC 2
	2ST-13	Synth CoTC 3
	2ST-15	Synth CoTC 5
	2ST-16	Synth CoTC 6
	2ST-17	Synth CoTC 7
	2ST-18	Synth CoTC 8
	2ST-19	Synth CoTC 9
	2ST-20	Synth CoTC 11
	2ST-21	Synth CoTC 12
Synthetic Pause & Ribozyme Elements	2ST-22	Synth Puase 1
	2ST-23	Synth Puase 2
	2ST-24	Synth Puase 3
	2ST-25	Synth Pause 5
	2ST-26	Synth Pause 6
	2ST-27	Synth Ribo 1
CHO-K1 Homologous polyadenylation & CoTC Element	2ST-28	S100a6 pA
	2ST-29	Gadph pA
	2ST-30	S100a6 CoTC
Commercially Available Polyadenylation Elements	2ST-31	HSVtk pA
	2ST-32	HIV-1 pA
	2ST-33	Adeno L1 pA
	2ST-34	bGH pA
	2ST-35	Lambda LC pA
	2ST-36	hGH pA

The construction of the MGEV variants was performed using Golden Gate cloning of the *SEAP* +pA/terminator variants and *CL* expression cassette, and the pExp-Vec-GG recipient vector at a 2:1 ratio. Transformed *E.coli* colonies were selected and screened by restriction digest colony screen using AgeI endonuclease and verified by gel electrophoresis as shown in Figure 5-6.

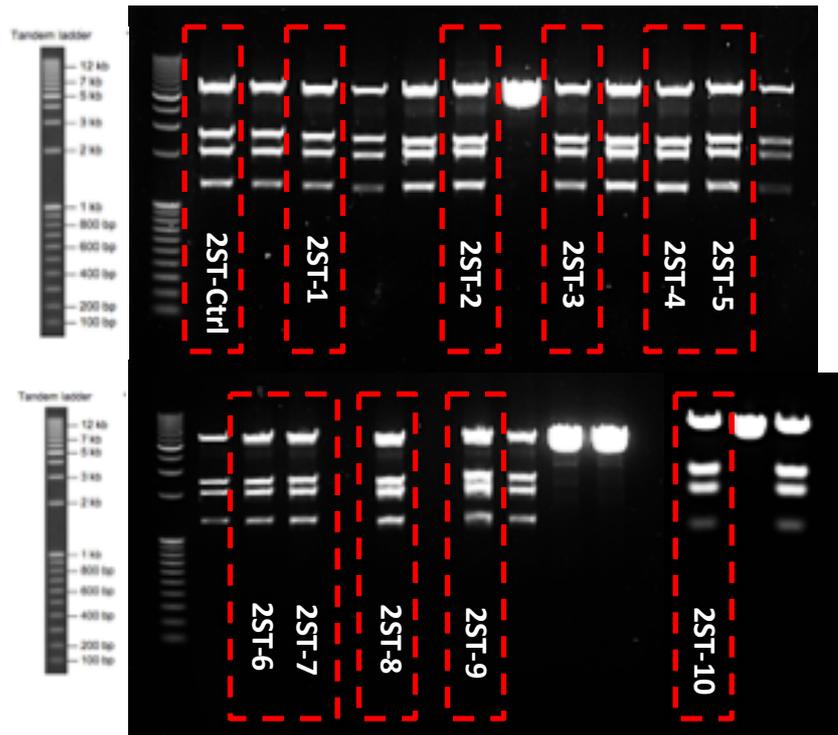


Figure 5-6: Restriction digest colony screen of multi-gene expression vector variants containing the control and synthetic polyadenylation elements.

A 1% agarose gel showing the results of a restriction digest colony screen using *AgeI* endonuclease. The image indicates the successful construction of the MGEV variants comprising of the SV40 late pA control and the synthetic pA elements. The lanes highlighted with the red box are the positive colonies based on fragment sizes and progressed to DNA amplification and testing.

The positive colonies are highlighted in the red box within Figure 5-6. These positive colonies represent the successful construction of the MGEV variants containing either the SV40 late pA control element or the synthetic pA elements. The positive colonies were identified based on the expected DNA fragments post-digestion with *AgeI* when compared against the DNA ladder. Four DNA fragment sizes were expected – fragment 1 - 5844 bp, fragment 2 - 2594 to 2635 bp, fragment 3 – 1971 bp and fragment 4 – 1304 bp. The same restriction digest colony screens were performed for the remaining MGEV variants described in Table 5-2. The results of the colony screen are presented in Appendix C.

5.2.1.3. Screening of the transcription terminator library

The MGEV system was used to screen the transcription terminator library by co-expressing *SEAP* and *CL* for 72 h using the HT transient transfection protocol. This approach would elucidate the functionality of the terminator by the impact on *SEAP* expression. Furthermore, it would identify preliminary information of transcription termination efficiency by measuring downstream *CL* expression (refer to Figure 5-5).

The expression of *SEAP* and *CL* was quantified using a plate-based absorbance and luminescence assay and compared against the control. In this study, the control was the industrial standard system where an SV40 late pA, referred to as SV40_pA, was used as a transcription terminator downstream of each recombinant gene. As shown in Chapter 3 and 4, transcriptional interference was observed when an SV40 late pA was used. However, for the purpose of a HT screen of the terminator library, the interference was normalised to a ratio of 1:1 for *SEAP* and *CL* expression within the control. Therefore, any differences in expression compared to the control during screening was an effect of the transcription terminator variant. A total of 35 different terminator variants were screened and the data was divided into 4 subcategories– 1) pA element, 2) synthetic CoTC elements, 3) CHO-K1 pA and CoTC homologous elements and 4) synthetic pause and ribozyme elements. The results of the screens for each subcategory are expanded below.

5.2.1.3.1. Screening of polyadenylation elements

A subcategory of the pA elements screened was a set of commercially available pA elements. The aim was to compare SV40_pA against other viral and mammalian derived pA elements. This was achieved by substituting the pA element downstream of the *SEAP* gene while maintaining the SV40 late pA downstream of the *CL* gene within the bicistronic cassette. The expression of both reporters was measured after 72h and fold change was compared against SV40_pA as shown in Figure 5-7.

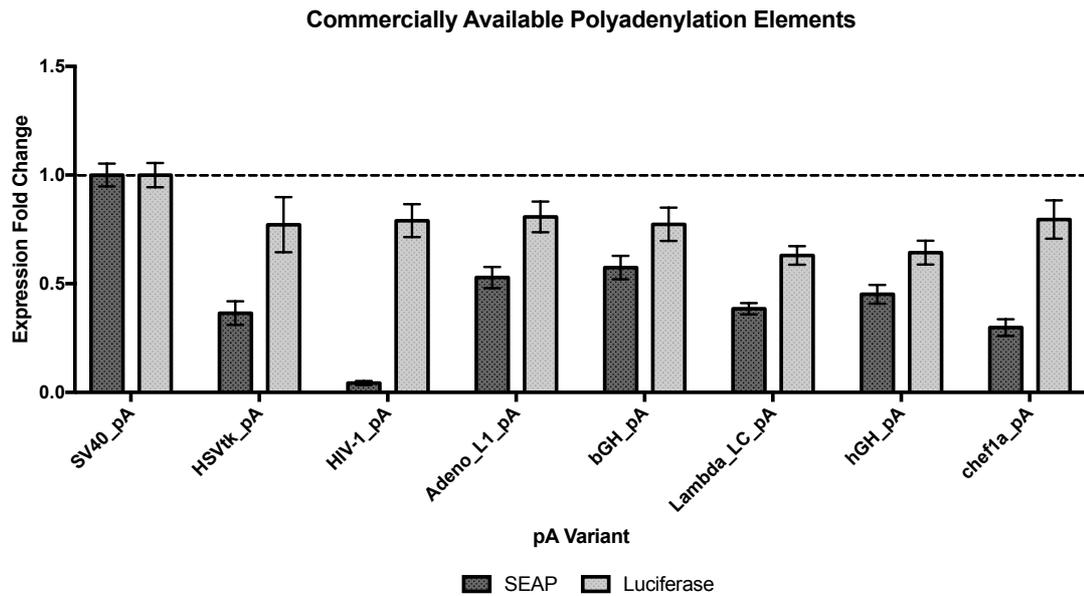


Figure 5-7: Screening of commercially available polyadenylation element within a multi-gene expression vector.

The bar chart compares the expression fold change of *SEAP* and *CL* after 72h of transient expression within a MGEV. A range of viral and mammalian derived commercially available pA element were screened and compared against the industry standard SV40_pA element. The error bars represent SEM of biological triplicates.

Figure 5-7 shows that the alternative viral and mammalian derived pA elements reduced expression of *SEAP* and *CL* by varying degrees. *SEAP* expression was more effected, showing a reduction of $\geq 42.5\%$, whereas *CL* was reduced by $\geq 19.2\%$. The data infers that the alternative commercially available pA elements had lower efficiency of termination by the substantial decrease in *SEAP* expression compared to *CL* expression. Additionally, the lower *SEAP* expression suggests that the pA element may have an impact on gene regulation and 3' end processing of mRNA as well as transcription termination. For example, the pA element could affect efficiency of nuclear export, mRNA stability and translation of *SEAP* (Li et al., 2017). Therefore, indicating that the SV40_pA was the preferred and optimal commercially available pA element.

Another subcategory of pA variants screened were synthetic pA elements derived from an SV40 late pA sequence. The aim of the synthesised library was to generate a fully defined synthetic pA element which was smaller and functional. This was determined by substituting the SV40_pA sequence for a synthetic variant within the bicistronic cassette. Expression of *SEAP* and *CL* was measured after 72h post-

transfection and fold change was quantified when comparing against the SV40_pA control as shown in Figure 5-8.

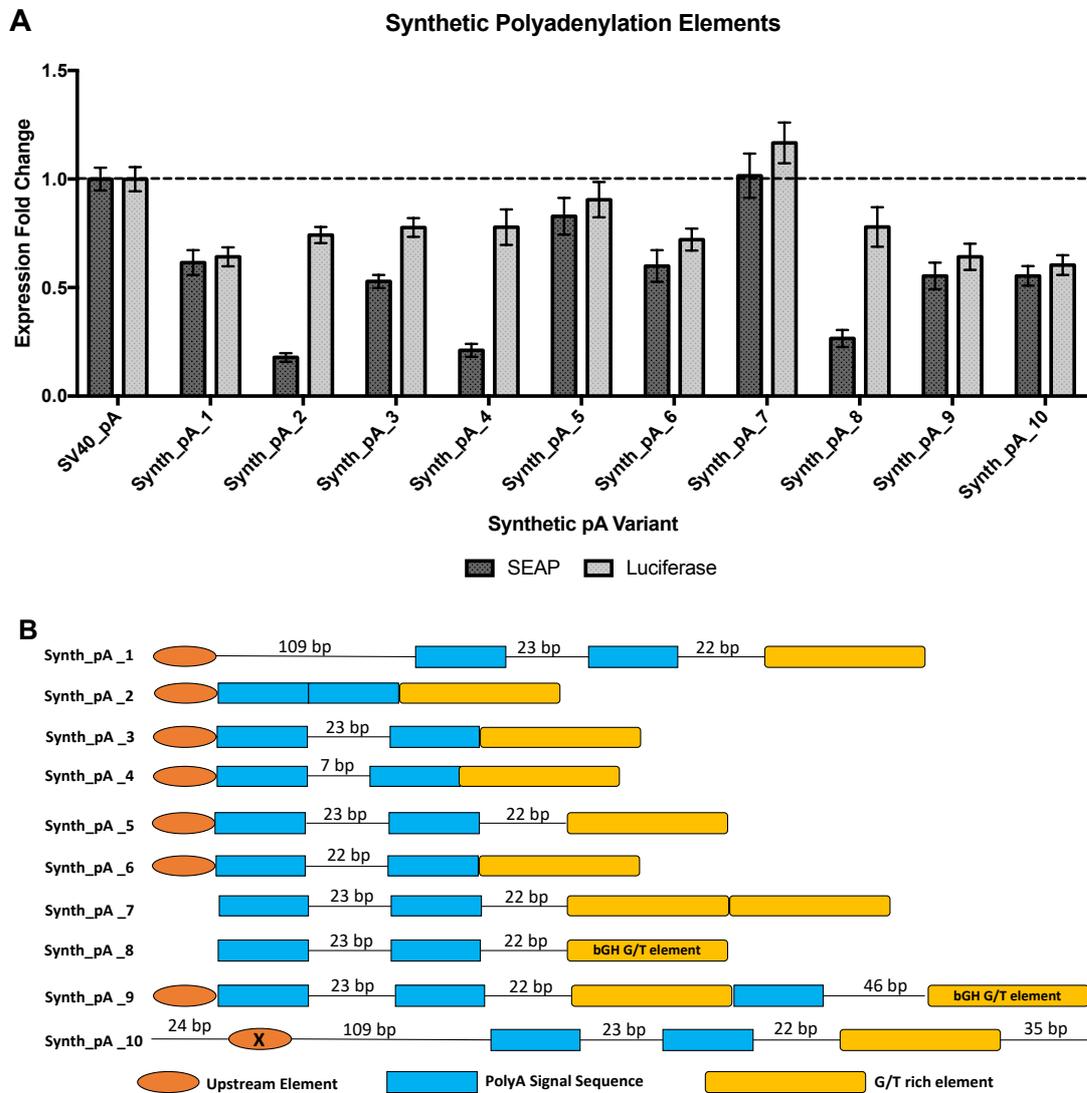


Figure 5-8: Screening of synthetic polyadenylation elements within a multi-gene expression vector.

A) Bar chart comparing the expression fold change of *SEAP* and *CL* after 72h of transient expression within a MGEV. A panel of synthetically designed pA elements derived from an SV40 late pA sequence were screened for impact on expression when compared to the SV40_pA element. The error bars represent SEM of biological triplicates. **B)** Schematic summarising the library of synthetic pA variants designed and substituted for the SV40 late pA element.

The expression fold change of *SEAP*, as shown in Figure 5-8A, was either matched or reduced by the synthetic pA variants when compared against the SV40_pA control by a varying degree. Similarly, *CL* expression was reduced by a differing levels in most variants apart from synth_pA_7. Synth_pA_7 element was the best performing variant by maintaining *SEAP* expression and increasing *CL* expression by 16.7% when

compared against the SV40_pA control. Furthermore, this variant was 82 bp compared to the SV40_pA being 241 bp, therefore minimising the length of the pA element and maintain functionality. Similar to the commercially available pA screen, *SEAP* expression was affected more than *CL* expression, indicating that the pA element plays a more crucial role in 3' processing of mRNA and influencing protein expression by nuclear export, mRNA stability and translation (Li et al., 2017) compared to minimising interference of *CL*.

Additionally, the screen of library assisted in narrowing the design space by identifying essential design criteria to maintain pA functionality when comparing the expression data of Figure 5-8A with the schematics of Figure 5-8B. The presence of an extra G/T rich element and omission of the upstream element improved *SEAP* and *CL* expression as shown by comparing expression between synth_pA_5 and 7 variants. Moreover, the sequence of the G/T rich motif can impact functionality of the pA element as shown by switching the SV40 late pA G/T rich motif for the bGH pA G/T rich motif between synth_pA_5 and 8 leading to substantial reduction in *SEAP* expression.

Another essential feature was the DNA spacer between the PAS motifs as demonstrated by a 35% reduction in *SEAP* expression between synth_pA_2 compared to synth_pA_3. Similarly, a 7 bp spacer between PAS motifs as shown in synth_pA_4 did not improve *SEAP* expression compared to synth_pA_3. The spacer between the PAS and G/T rich motifs further improves pA functionality as shown by increasing *SEAP* expression by 30.1% between synth_pA_3 and 5. Finally, the upstream element is undetermined on functionality of a pA element as minimal change in *SEAP* expression was observed between synth_pA_1 and 10.

5.2.1.3.2. Screening of co-transcriptional cleavage elements

The synthetic CoTC elements were designed using conserved motifs identified through literature and a library was generated by varying the A/T tract length and DNA spacers length between A/T tracts. A library of 12 variants was designed of which 10 were screened (Synth_CoTC_4 and 10 were omitted due to cloning and synthesis issues respectively). Literature had indicated that the CoTC element functioned

downstream of a pA sequence (West and Proudfoot, 2009), therefore the synthetic elements were cloned downstream of an SV40 late pA element in the *SEAP* TU within the bicistronic cassette (as shown in Figure 5-5). The synthetic CoTC elements were screened within the MGEV by transient expression for 72h. The *SEAP* and *CL* expression was quantified, and fold change was compared against the SV40_pA control as shown in Figure 5-9A.

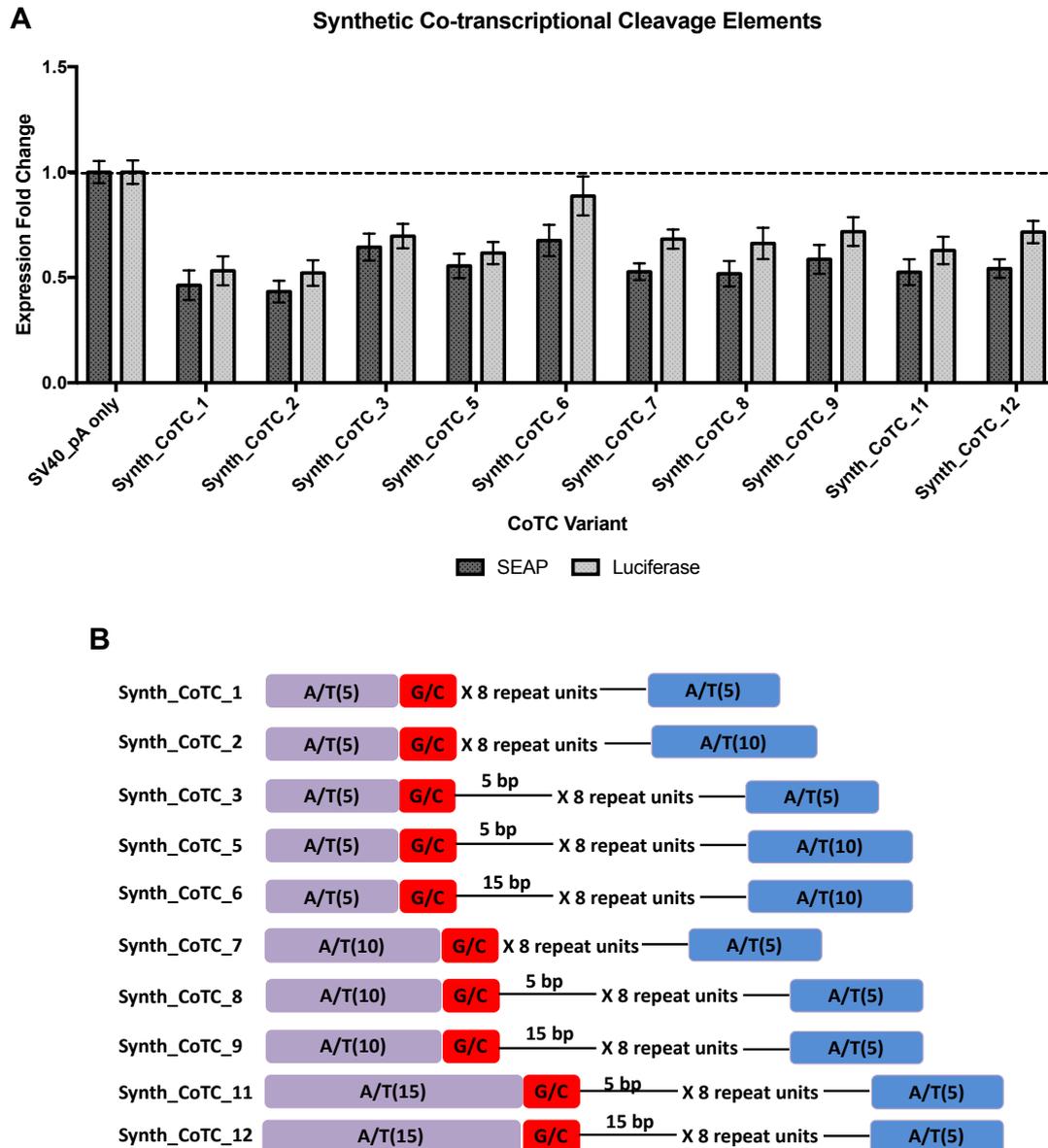


Figure 5-9: Screening of synthetic co-transcriptional cleavage elements within a multi-gene expression vector.

A) Bar chart comparing the expression fold change of *SEAP* and *CL* after 72h of transient expression within a MGEV. A panel of synthetically designed CoTC elements derived from conserved motifs found in literature were screened for impact on expression when compared against the industry standard SV40_pA only. The error bars represent SEM of biological triplicates. **B)** Schematic summarising the library of synthetic CoTC elements designed when positioned downstream of an SV40 late pA and refers to the expression data in the bar chart.

Figure 5-9A shows that the expression of *SEAP* and *CL* were both lower than the SV40_pA control when adding the synthetic CoTC element downstream of the *SEAP* gene with a SV40 late pA element, by a varying degree. The synthetic CoTC element which had the least negative impact on expression was Synth_CoTC_6 with an average of 32.4% and 11.3% reduction in *SEAP* and *CL* respectively. Furthermore, intra-cassette comparison of relative gene repression of *SEAP* and *CL* against the control was performed. As mentioned previously, the *SEAP* and *CL* expression within the control MGEV was normalised to a 1:1 ratio, therefore variation would indicate change in the expression dynamics within the bicistronic cassette. The difference in relative expression fold change between *SEAP* and *CL* was the largest at an average increase of 21.1% for Synth_CoTC_6. Therefore, even though expression was lower than the control, the relative *CL* expression within the bicistronic cassette was higher inferring potential reduction in downstream gene repression. Dissection of the element highlighted the 15 bp DNA spacer between A/T tracts had the lowest negative impact on expression of *CL* as shown by Synth_CoTC_6 and 12, with a difference of 21.1% and 17.4% respectively. Although, the A/T tract lengths may vary, the comparison of Synth_CoTC_11 and 12 variants acted as a process of elimination.

5.2.1.3.3. Screening of CHO-K1 homologues

The CHO-K1 pA and CoTC homologues were derived from 3' end genome analysis of highly transcriptionally active genes. The aim was to investigate if native elements would have a positive effect on transcription termination. This was determined by either substituting the pA element or cloning the CoTC element downstream of a SV40 late pA sequence within the *SEAP* expression cassette. The elements were screened within an MGEV co-expressing *SEAP* and *CL* after 72h of culturing. The expression of *SEAP* and *CL* was quantified, and the fold change was calculated against the SV40_pA control as shown in Figure 5-10.

CHO-K1 Polyadenylation & Co-transcriptional Cleavage Homologue Elements

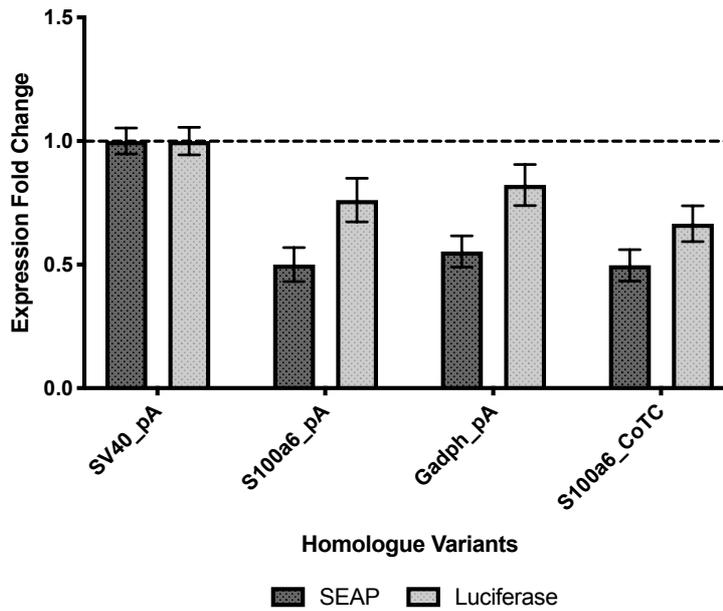


Figure 5-10: Screening of CHO-K1 polyadenylation and co-transcriptional cleavage homologous elements within a multi-gene expression vector.

The bar chart compares the expression fold change of *SEAP* and *CL* after 72h of transient expression within a MGEV. Two pA homologues from highly transcriptionally active genes were screened in substitution of a SV40 late pA. Whereas, the one CoTC homologue was positioned downstream of the SV40 late pA element. Both were screened to observe impact on gene expression. The error bars represent SEM of biological triplicates.

Once again, the homologous pA and CoTC elements derived from the CHO-K1 genome showed reduced expression of *SEAP* and *CL* when compared against the SV40_pA control. The average decrease in *SEAP* expression ranged from 44.7% to 50.3%, while *CL* expression was decreased between 17.8% and 33.5%. The S100a6_CoTC element had the most detrimental effect on both *SEAP* and *CL* expression as shown in Figure 5-10. The data indicate the pA and CoTC homologues had a more negative effect on *SEAP* expression compared to *CL*, suggesting lower efficiency of transcription termination of *SEAP* gene. Interestingly, the difference in relative fold change between *SEAP* and *CL* within each MGEV showed a bigger difference within CHO-K1 pA homologues of 26.1% and 26.9% for S100a6_pA and Gadph_pA respectively. Whereas, it was lower for S100a6_CoTC at 16.8%. This indicated the expression of *SEAP* and *CL* was affected differently within the MGEV and could either suggest a negative effect on *SEAP* expression or reduction in repression of *CL*. However, the number of endogenous elements screened were limited, therefore a more extensive genome analysis could be performed to find

conserved motifs in CHO cells and then design CHO derived synthetic terminator elements.

5.2.1.3.4. Screening of pause and ribozyme elements

Literature had indicated the position of a pause or ribozyme element downstream of a pA element could assist in improving transcription termination by slowing the RNA pol II (Dower et al., 2004; Gromak et al., 2006). The conserved motifs of a MAZ element were tested by varying the proximity from the pA and the number of MAZ motifs. A hammerhead ribozyme was positioned downstream of the pA element to promote self-cleavage of the pre-mRNA transcript and expose nascent mRNA to Xrn2-mediated degradation, in turn terminating transcription. The synthetically designed elements were screened by positioning them downstream of a SV40 late pA behind the *SEAP* gene within the bicistronic cassette. The expression of *SEAP* and *CL* was measured after 72 h of incubation. The fold change was calculated by comparing against the SV40_pA control as shown in Figure 5-11A.

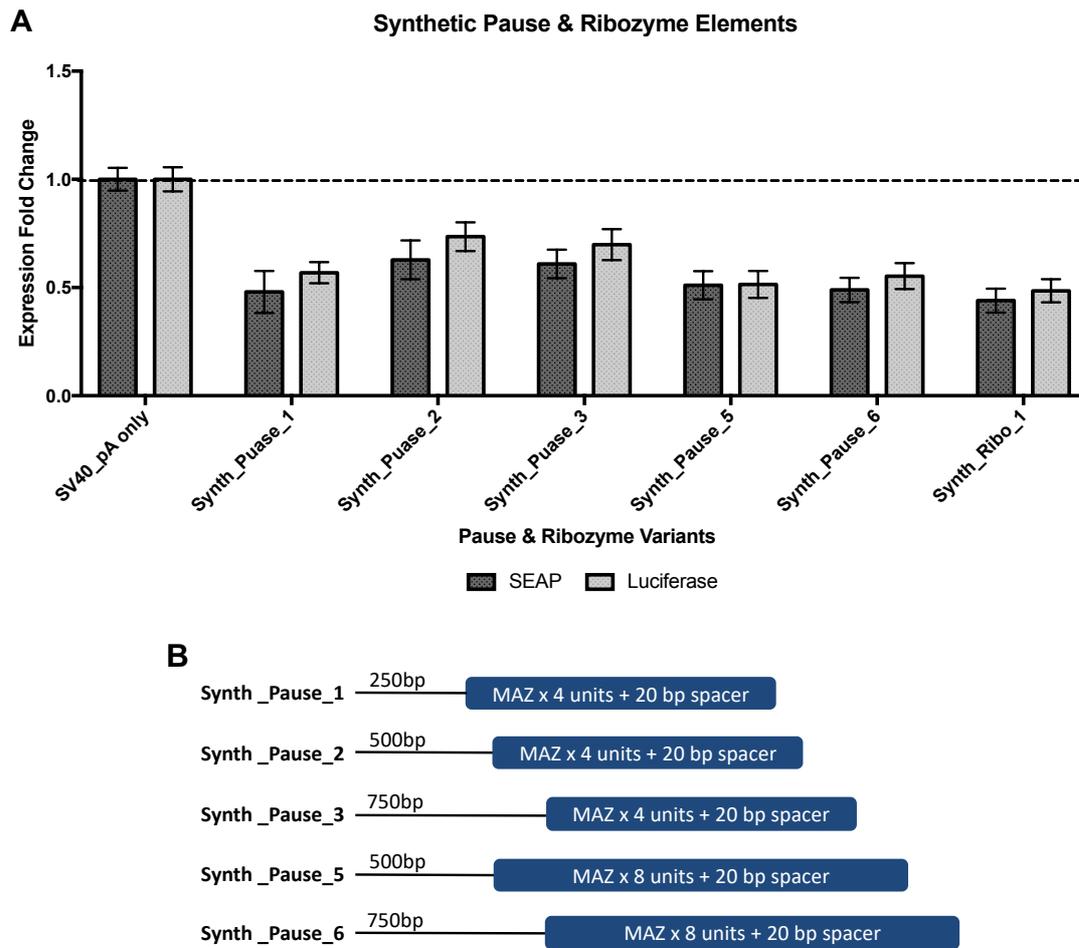


Figure 5-11: Screening of synthetic pause and ribozyme elements within a multi-gene expression vector.

A) Bar chart comparing the expression fold change of *SEAP* and *CL* after 72h of transient expression within a MGEV. A library of synthetically designed pause elements derived from conserved motifs found in literature and a hammerhead ribozyme element were screened for impact on expression when compared against the industry standard SV40_pA element only. The error bars represent SEM of biological triplicates. **B)** Schematic summarising the library of synthetic pause elements designed when positioned downstream of a SV40 late pA sequence and refers to the expression data in the bar chart.

The expression levels of *SEAP* and *CL* were both lower than the SV40_pA control with a reduction of $\geq 37.2\%$ and $\geq 26.5\%$ respectively, as shown in Figure 5-11A. This indicated the pause and ribozyme element had a negative impact on expression of both genes. Furthermore, the relative difference in *SEAP* and *CL* expression within each MGEV was minimal ranging from 0.4% to 10.7%. Therefore, the addition of the synthetic pause or ribozyme element between two genes within a bicistronic cassette had a similar level of *SEAP* and *CL* repression. Due to lower *SEAP* expression, it could be inferred that the addition of pause or ribozyme element may have negatively impacting on mRNA processing and transcription termination, as improved

termination efficiency has shown to enhance recombinant protein expression (West and Proudfoot, 2009).

5.2.2. Enhancing monoclonal antibody production in a stable CHO cell line by co-expressing *XBP-1s* using a multi-gene expression vector

Previous data has indicated the transient co-expression of *XBP-1s* and a DTE mAb (provided by MedImmune) increased mAb titres. Furthermore, a correlation of increased *XBP-1s* gene dosage led to an increase in mAb expression. Therefore, a MGEV combined with a range of synthetic promoters to titrate *XBP-1s* was performed using the same DTE mAb. This section will cover the adaptation of the MGEV system for TI-mediated engineering, construction of MGEVs co-expressing a DTE mAb and *XBP-1s*, and generation and characterisation of engineered TI stable pool variants.

5.2.2.1. Adapting the pExp-Vec-GG recipient vector for compatibility with MedImmune's targeted integration platform

The MedImmune TI platform comprises of a CHO host cell line which has a landing pad engineered into the genome at a highly active gene expression locus. The recombination system used was adapted from P1 bacteriophage utilising a cause recombination (Cre) recombinase and non-palindromic recognition sites called locus of crossover in P1 bacteriophage (LoxP). The Cre/LoxP system leads to site-specific integration with reciprocating LoxP sites (Kühn and Torres, 2002). Therefore, the landing pad within the host cell and the counterpart plasmid require LoxP sites to facilitate the Cre-mediated recombination. As a result, the pExp-Vec-GG recipient vector for MGEV assembly required modification by introducing a LoxP site and a promoter-less hygromycin resistance gene. This was performed by restriction digestion-ligation cloning using *SspI* and *NotI*, to excise the episomal origin of replication element and the *GS* cassette but retaining the SV40 late pA element. The fragment re-introduced was from a MedImmune provided plasmid containing the LoxP site and the promoter-less hygromycin resistance gene. After transformation, *E.coli* colonies were screened by restriction digest colony screen and verified by gel electrophoresis as shown in Figure 5-12.

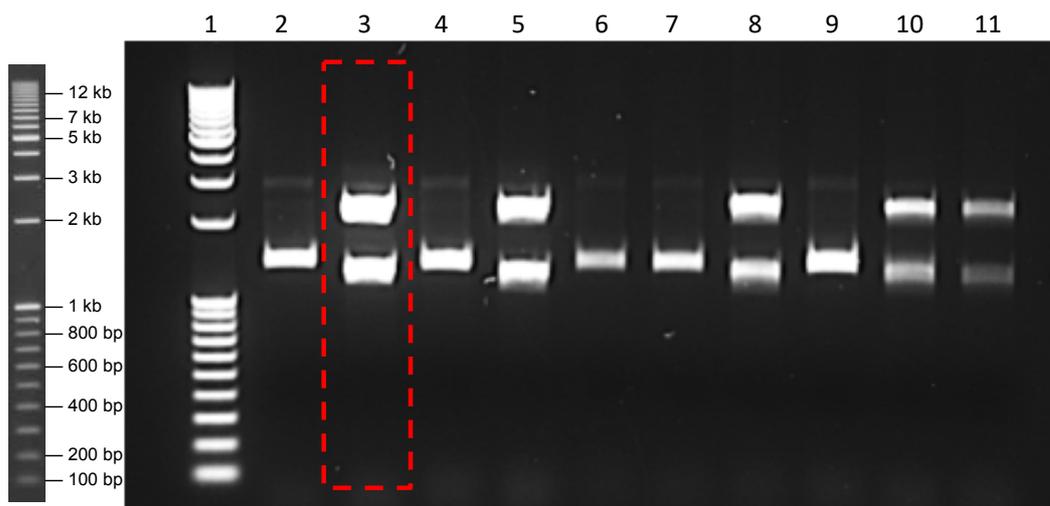


Figure 5-12: Restriction digest colony screen verifying successful cloning of hygromycin resistance gene and LoxP site into the pExp-Vec-GG recipient vector.

A 1% agarose gel showing the screening and identifying of an *E.coli* colony containing a pExp-Vec-GG recipient vector successfully modified by cloning the promoter-less hygromycin resistance gene and LoxP site. This was achieved by restriction digest of the DNA plasmid by SspI and NotI. The DNA in lane 3 and highlighted within the red box indicates a successfully modified vector. This colony was progressed to DNA amplification and MGEV construction in this study.

Figure 5-12 shows the results of 10 colonies screened by SspI and NotI restriction digest to verify successful cloning of the promoter-less hygromycin resistance gene and the LoxP site. The colony containing the digested DNA in lane 3 and highlighted in the red box was selected based on the size of the two fragments (2287 bp and 1281 bp) and the high intensity of the bands indicating high plasmid copy number within the colony. The TI compatible MGEV recipient vector backbone was referred to as pExp-Vec-GG (TI).

5.2.2.2. Design and construction of multi-gene expression vectors co-expressing a DTE monoclonal antibody and XBP-1s

Data generated by Dr. Joe Cartwright and Dr. Claire Arnall (within the lab group) demonstrated a gene dose-dependent response when expressing *XBP-1s* to increase DTE mAb expression transiently. Additionally, it was highlighted that 10% of total mAb DNA load was the best *XBP-1s* gene dose to increase mAb titre by 240% (Cartwright et al., 2020). Therefore, it was investigated whether a MGEV with titrated expression of *XBP-1s* could demonstrate a similar response when the CHO cell is stably engineered. The titration of *XBP-1s* was achieved by using synthetic promoters developed by Dr. Adam J. Brown, where a series of promoter variants were selected based on their validated transcriptional strength (Brown et al., 2017). Whereas, the *LC*

and *HC* genes were regulated using a hCMV-MIE promoter. The final design and construction of the prototype MGEVs co-expressing a DTE mAb and *XBP-1s* were described in Table 5-3.

Table 5-3: Design of prototype multi-gene expression vectors co-expressing a DTE monoclonal antibody and *XBP-1s*.

The table summarises the prototype MGEV variants designed and constructed. This includes the position of each gene within the polycistronic cassette, the pA elements used and the promoter allocation to regulate mAb expression and titrate *XBP-1s* expression.

Promoter Variants	Genes for Assembly			Promoter Allocation		
	Position 1	Position 2	Position 3	Position 1	Position 2	Position 3
	TU1-Lambda_pA	TU2-SV40_pA	TU3-SV40_pA			
Prototype-Control	<i>DTE-LC</i>	<i>DTE-HC-E</i>	-	hCMV-MIE	hCMV-MIE	-
Prototype-1	<i>DTE-LC</i>	<i>DTE-HC</i>	<i>XBP-1s</i>	hCMV-MIE	hCMV-IE	5RPU
Prototype-2	<i>DTE-LC</i>	<i>DTE-HC</i>	<i>XBP-1s</i>	hCMV-MIE	hCMV-MIE	10RPU
Prototype-3	<i>DTE-LC</i>	<i>DTE-HC</i>	<i>XBP-1s</i>	hCMV-MIE	hCMV-MIE	20RPU
Prototype-4	<i>DTE-LC</i>	<i>DTE-HC</i>	<i>XBP-1s</i>	hCMV-MIE	hCMV-MIE	hCMV-MIE

The synthetic promoters allocated are predicted to express relative expression strengths of 5%, 10% and 20% for 5RPU, 10RPU and 20RPU respectively. Although, the observed expression strength will be lower due to positional-mediated gene repression (as shown in Chapter 4). No promoter stronger than the hCMV-MIE was selected as it would be counter-intuitive to over-express the accessory gene over the mAb of interest for biopharmaceutical production applications. The polycistronic cassette was located downstream of the hygromycin selection marker to maintain alignment with industry performed TI stable pool generation. An example vector map showing assembly of Prototype-3 is shown in Figure 5-13.

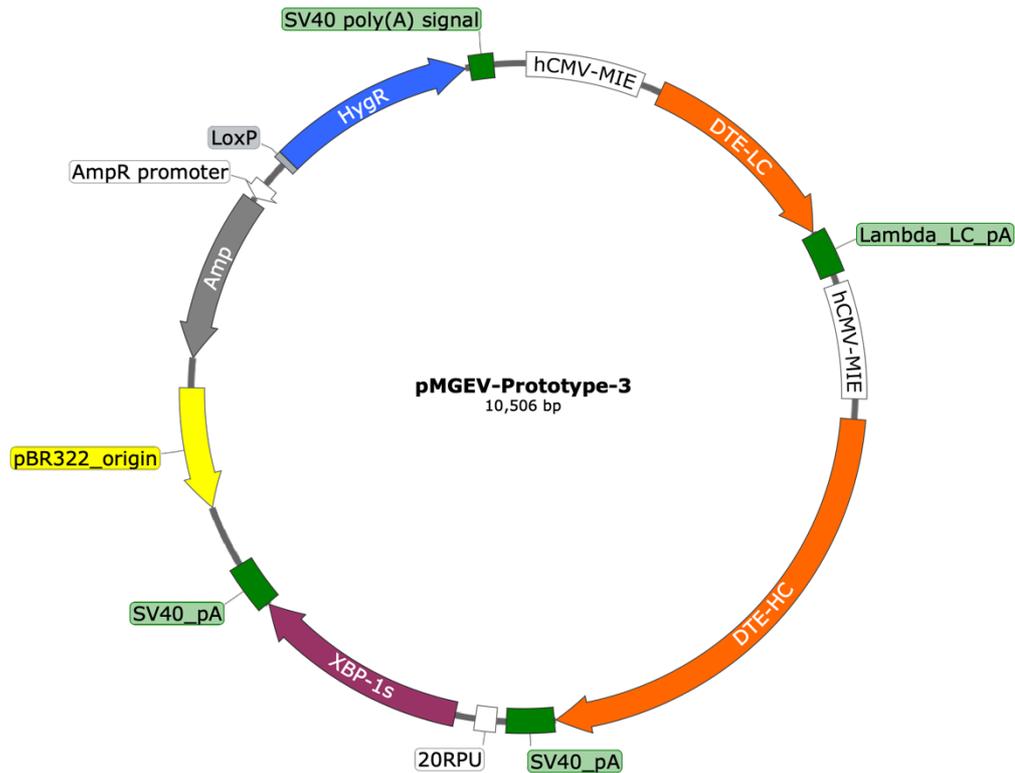


Figure 5-13: Prototype-3 multi-gene expression vector for engineering CHO cells with *XBP-1s*. Vector map showing a MGEV that co-expresses the *LC* and *HC* CDSes of a DTE mAb and *XBP-1s* CDS at a titrated level using hCMV-MIE and a synthetic promoter. In this vector map a 20RPU promoter was used to control the gene expression level of *XBP-1s*. Furthermore, the MGEV is compatible with MedImmune's proprietary TI platform.

The prototype MGEV variants were constructed by Golden Gate assembly combining the TU1, 2 and 3 inserts and pExp-Vec-GG(TI) recipient vector at a 2:1 ratio. The transformed *E.coli* colonies were screened by restriction digest colony screen using AgeI and verified by gel electrophoresis as shown in Figure 5-14.

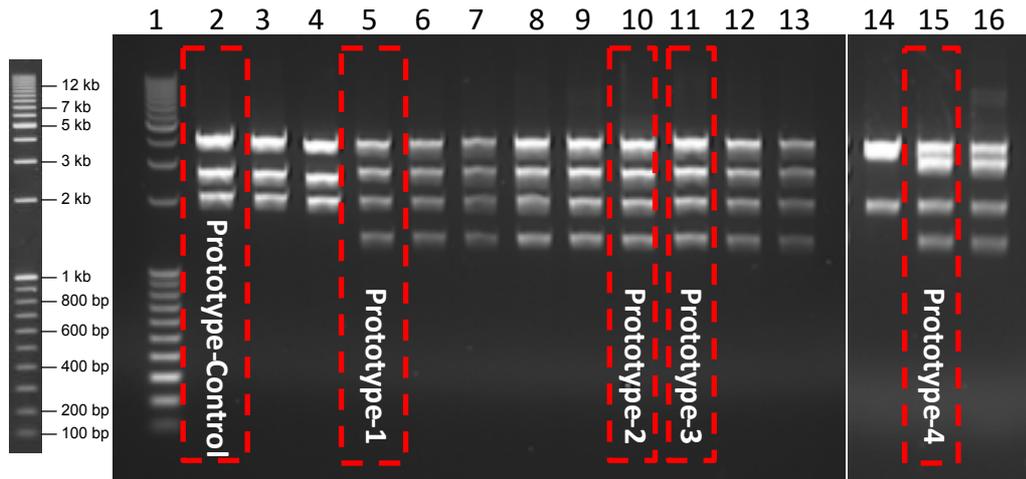


Figure 5-14: Restriction digest colony screen verifying successful cloning of prototype multi-gene expression vectors.

A 1% agarose gel showing the restriction digest colony screen of 15 different colonies containing plasmid DNA (3 colonies per prototype MGEV variant) by *AgeI* restriction digestion. The DNA in lanes 2, 5, 10, 11 and 15 (highlighted within red boxes) were positive colonies for Prototype-control, 1, 2, 3 and 4 respectively.

Figure 5-14 shows the successful construction of the prototype MGEV variants by restriction digest colony screen using *AgeI* endonuclease. The positive colonies were deduced based on the number and size of the fragments, and the band intensity. The gel shows 3 fragments for the Prototype-Control at the following sizes – fragment 1: 4198 bp, fragment 2: 2854 bp and fragment 3: 2024 bp, whereas, 4 fragments were observed for the Prototype 1,2, 3 and 4. The prototypes with the synthetic promoters (Prototype-1,2 and 3) showed DNA fragment sizes as follows – fragment 1: 4198 bp, fragment 2: 2854 bp, fragment 3: 2024 bp and fragment 4: 1422 bp. Prototype-4 had different size fragments and were as follows – fragment 1: 4198bp, fragment 2: 3321 bp, fragment 3: 2024 bp and fragment 4: 1422 bp. Therefore, the colonies highlighted in the red boxes in lanes 2, 5, 10, 11 and 15 were selected and progressed to DNA amplification and stable engineering. Furthermore, the prototype MGEV variants were verified by performing DNA sequencing and showing 100% sequence homology to the reference sequence (data not shown).

5.2.2.3. Establishing targeted integrated stable pools expressing the prototype multi-gene expression vector variants

The TI stable pools were established by co-transfecting the prototype MGEV and a plasmid expressing Cre recombinase at a 2:1 ratio into MedImmune's proprietary TI CHO host cell line. A triplicate of stable pools were set up per prototype MGEV variant leading to a total of 15 pools. The cultures were left to recover for 48h before supplementation of hygromycin at a final concentration of 500 µg/ml. The hygromycin was used to select for cells with successful recombination of the prototype MGEV into the landing pad, as this would functionalise the hygromycin resistance gene by facilitating a SV40 promoter upstream. The TI stable pools were monitored periodically and adapted to suspension when the viable cell concentration (VCC) and viability of the cultures were $> 3 \times 10^5$ cells/ml and $>75\%$ respectively. The recovery and adaptation to suspension of the TI stable pools was deduced by VCC and viability as shown in Table 5-4.

Table 5-4: Viable cell concentration and viability of TI stable pools post-suspension adaptation.

The table summarises the TI stable pools (with the prototype MGEV variants integrated into the genome) VCC and viability post-adaptation to suspension. The quantification indicates the recovery of the cells. The pools labelled with an asterisk (*) required an additional 72 h to recover and adapt to suspension compared to the remaining pools.

Prototype Variant	TI Stable Pool Name	Viable Cell Concentration ($\times 10^6$ cells/ml)	Viability (%)
Prototype-Control	A1	2.24	79.7
	A2	2.45	81.0
	A3*	0.72*	77.3*
Prototype-1	B1*	2.07*	89.6*
	B2	1.02	79.7
	B3	1.01	79.6
Prototype-2	C1	1.34	66.3
	C2	3.91	88.1
	C3	1.79	81.7
Prototype-3	D1	3.83	90.7
	D2	4.04	92.8
	D3	3.94	92.3
Prototype-4	E1	0.99	66.5
	E2*	0.87*	67.8*
	E3*	0.71*	69.6*

Majority of the TI stable pools were recovered and subcultured 72 h post-adaptation to suspension apart from A3, B1, E2 and E3. These pools required an additional 72 h to recover prior to subculturing. Clarified supernatant samples from day 7 overgrows were collected and tested for the detection of mAb. All pools showed a positive expression of the DTE mAb, indicating the prototype MGEVs were successfully functioning in the CHO cell genome (data not shown). The established TI stable pools were cryopreserved for long term storage and adapted to MedImmune's proprietary media for fed-batch overgrow (FBOG) analysis.

5.2.2.4. Growth profile of *XBP-1s* engineered stable pools during a fed-batch overgrow

The growth profile of the TI stable pools was investigated to observe the impact of stably co-expressing a DTE mAb and *XBP-1s* at titrated levels. This was achieved by performing a 13 day FBOG using MedImmune's proprietary media and feeding regime. The VCC and cell viability of the triplicate pools was measured by Dr. Susie Sou on days 3, 5, 7, 10, 11 and 13 as shown in Figure 5-15.

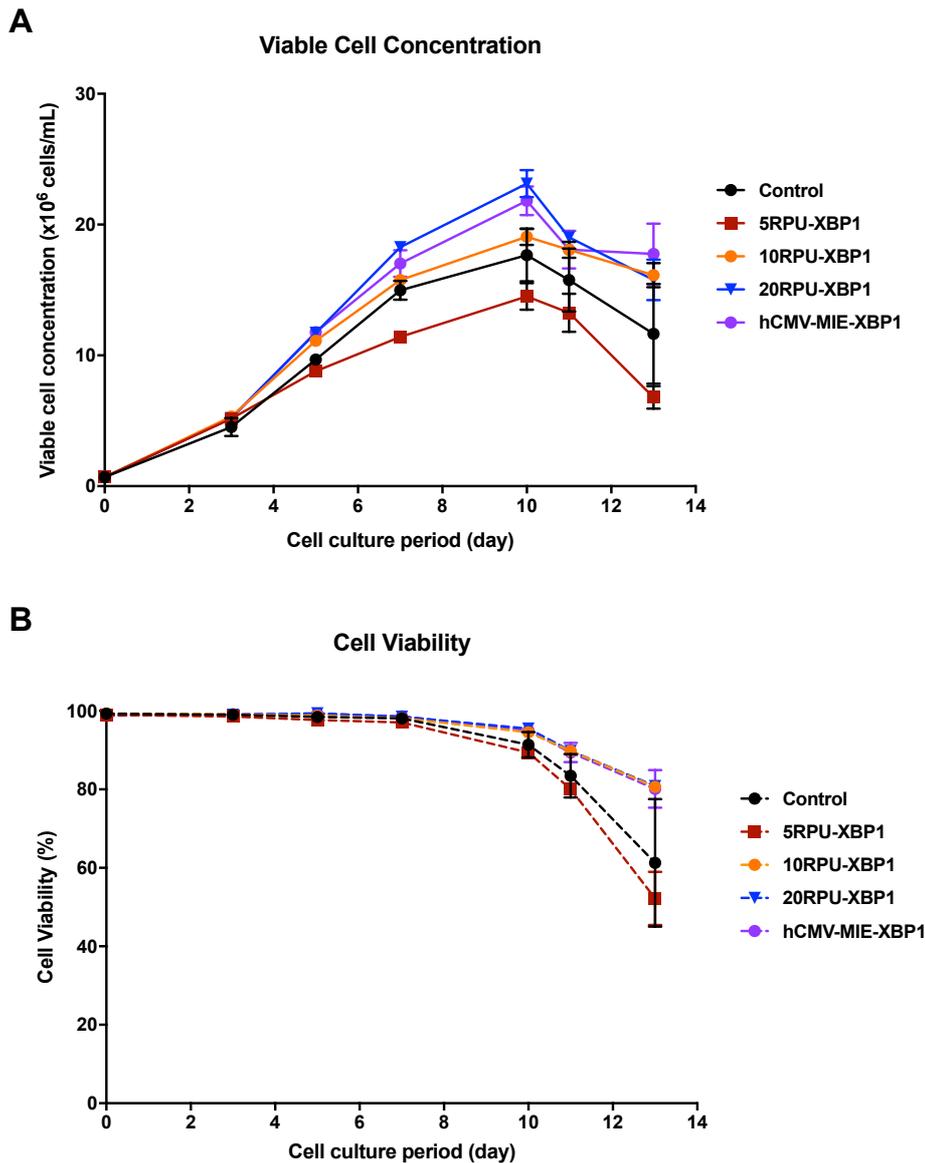


Figure 5-15: Viable cell concentration and cell viability of TI stable pools during a 13 day fed-batch overgrow.

Viable cell concentration (**A**) and cell viability (**B**) were measured during a 13 day FBOG using MedImmune's proprietary protocol. This was performed to characterise the growth profile of the *XBP-1s* engineered stable pools using the prototype MGEV variants and MedImmune's TI host. In both plots, the error bars represent the SEM of biological triplicates at each time point.

Figure 5-15A shows the growth trend through the 13 day FBOG where a higher average VCC for the triplicate pools expressing the 10RPU-XBP1, 20RPU-XBP1 and hCMV_MIE_XBP1 compared to the control (Prototype Control) pools. Conversely, the 5RPU-XBP1 triplicate pool showed lower VCC than the control through the FBOG. All cultures achieved their highest VCC on day 10 and the 20RPU-XBP1 pools attained the highest average concentration of 23.14×10^6 cells/ml. On day 11 onwards, the cultures were within the decline phase of growth as shown by reduction in VCC

and cell viability. The cell viability trend, as shown in Figure 5-15B, indicated the cultures were healthy and consistent through the duration of the FBOG and only deviated from the day 10 onwards, inferring variation in culture health during decline phase of the FBOG.

The integral viable cell density (IVCD) was measured over the 13 day FBOG for the triplicate of TI pools. The data indicated the accumulation of the viable cell biomass over the duration of the FBOG. The IVCD, as shown in Figure 5-16, demonstrated the identical trend to the VCC. Whereby, the average IVCD was higher for 10RPU-XBP1, 20RPU-XBP1 and hCMV-MIE-XBP1 pools when compared to the control and lower for the 5RPU-XBP1 pools. Furthermore, the graph showed the rate of biomass accumulation deviated between the pools after day 7 and the highest average IVCD was observed by 20RPU-XBP1 pools at 173.8×10^6 cells day/ml. Therefore, a correlation is observed between the VCC and IVCD over time that the co-expression of the *XBP-Is* has enhanced growth rate and biomass accumulation apart from for the 5RPU-XBP1 pools. This inferred that the *XBP-Is* may be alleviating protein expression-dependent stress and avoiding the onset of apoptosis.

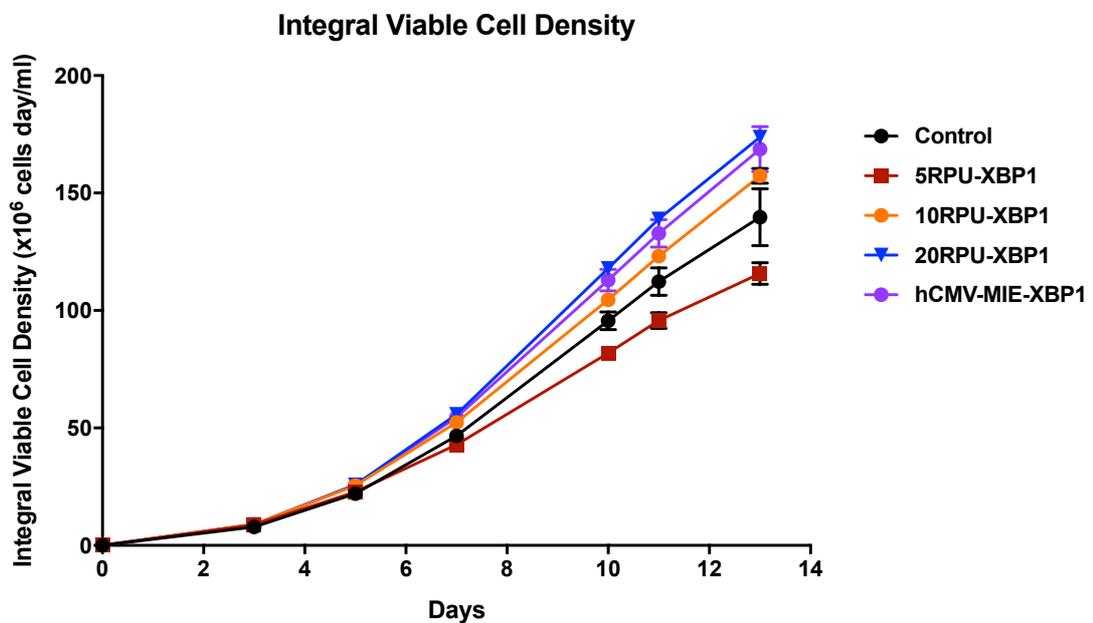


Figure 5-16: Integral viable cell density of TI stable pool during a 13 day fed-batch overgrow. Graph showing the IVCD of the *XBP-Is* engineered stable pools variants. The data demonstrates the trend and rate of biomass accumulated during a 13 day FBOG using MedImmune's proprietary protocol. In the plot, the error bars represent the SEM of biological triplicates at each time point.

5.2.2.5. Monoclonal antibody expression trend of *XBP-1s* engineered stable pools during a fed-batch overgrow

The trend of DTE mAb expression was investigated to measure the effect of *XBP-1s* co-expression on mAb titres. This was deduced by performing a 13 day FBOG using MedImmune's proprietary feeding regime. The mAb titre was measured by Dr. Susie Sou on day 3, 5, 7, 10, 11 and 13 by taking clarified supernatant samples and quantifying by Protein A HPLC. The quantified titre was plotted against the period of cell culture as shown in Figure 5-17A. Additionally, Figure 5-17B shows the titre upon harvest on day 13 for each replicate in a bar chart to demonstrate the variation.

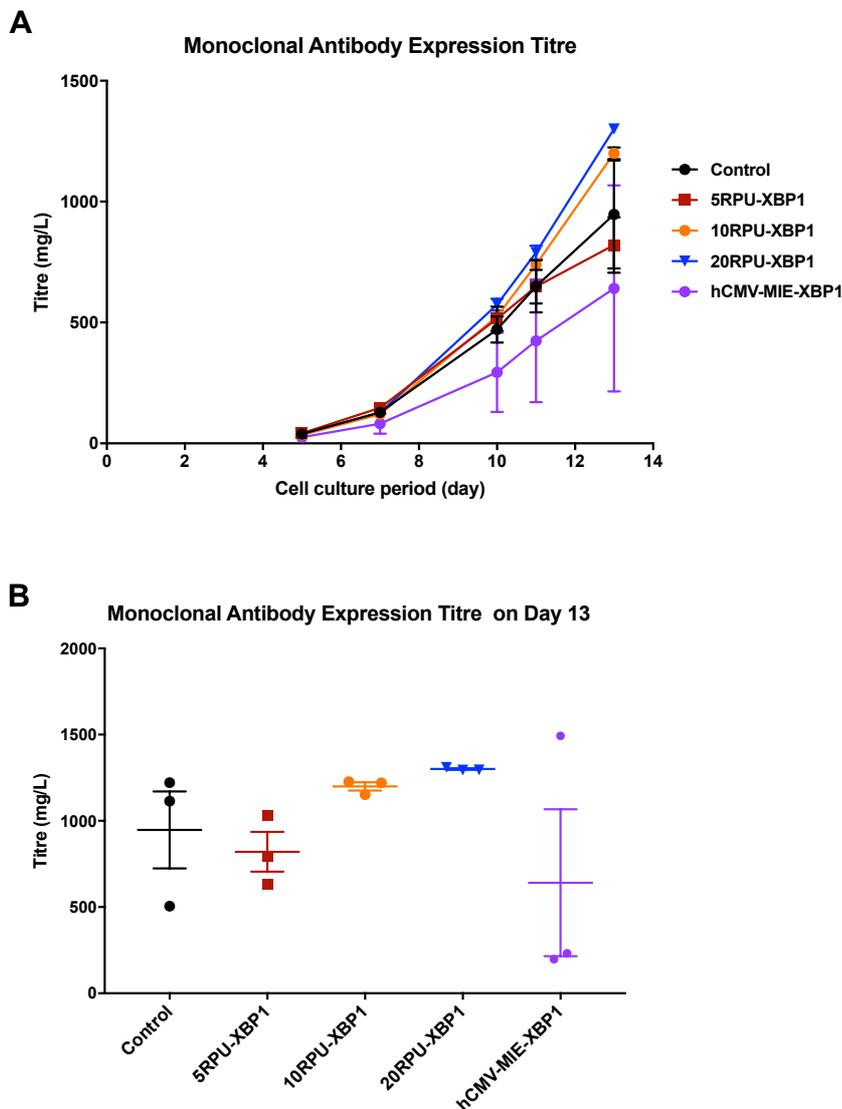


Figure 5-17: Monoclonal antibody expression titres of *XBP-1s* engineered stable pools during a 13 day fed-batch overgrow.

Graphs showing the mAb expression titre trends of the *XBP-1s* engineered stable pool variants during the FBOG process using MedImmune's proprietary feeding regime. **A)** Graph showing the expression trend of mAb over the duration of 13 days, whereas **B)** Graph showing the titre upon harvest of each pool on day 13 and the variation between biological replicates. In both plots, the error bars represent the SEM of biological triplicates.

Figure 5-17A shows the increase in mAb titre over time and demonstrated the average mAb expression of the replicates was higher in 10RPU-XBP1 and 20RPU-XBP1 pools over the control. The highest expression was observed by the 20RPU-XBP1 pools with an average titre on day 13 of 1300 mg/L, which is a 37.3% increase in titre compared to the control. The next highest expressing pools were the 10RPU-XBP1 pools with a 26.7% increase in mAb titre over the control. Additionally, the trend indicates the titration of *XBP-1s* had an additive effect on titre from the 5PRU to the 20RPU promoter. However, a lower average expression was observed for the 5RPU-XBP1 pools compared to the control. A potential hypothesis for the lower expression of the 5RPU-XBP1 pools to the control was the composition of the plasmid integrated into the genome, whereby the control is a double gene integrant and the 5RPU-XBP1 variant is a triple gene integrant. The difference in plasmid composition integrated and the additional transcriptional interference of a third gene plasmid could contribute to lower expression. Interestingly, hCMV-MIE-XBP1 stable pools showed the lowest average mAb titre. However, after investigating the variation between replicates, as shown in Figure 5-17B, 2 out of the 3 replicates demonstrated substantially lower levels of expression upon harvest and throughout the entire culture duration. A similar observation was made for one of the control replicates. A correlation was observed that the low expressing pool variants were the same variants that struggled to recover post-selection. Therefore, suggesting a potential consequential effect during stable pool generation.

The qP of the *XBP-1s* engineered stable pools was measured on day 13 to determine the average rate of mAb expression of the stable pool population. This was calculated by accounting for the IVCD of the pools on day 13. Figure 5-18 is a bar chart comparing the average qP over biological triplicates for each stable pool variant.

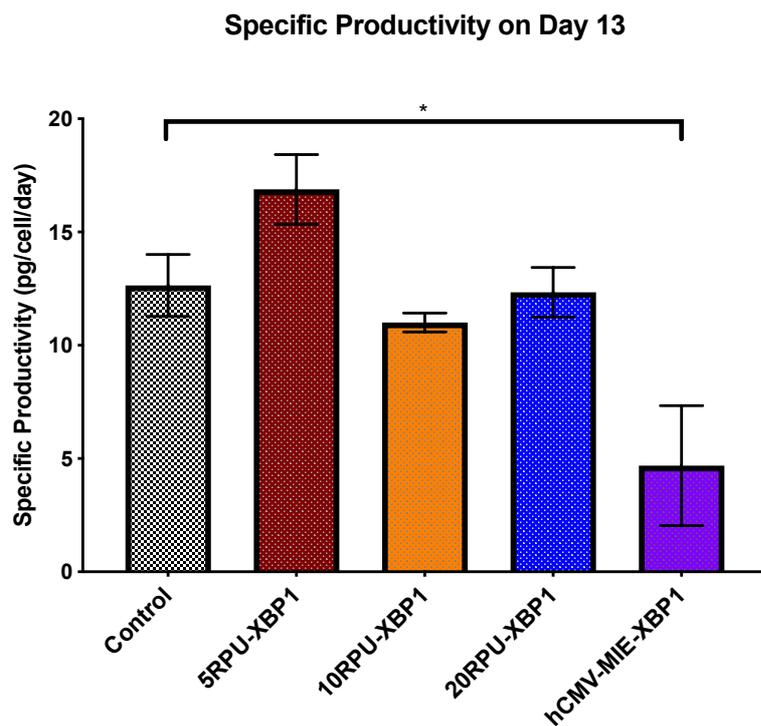


Figure 5-18: Specific productivity of monoclonal antibody expression within *XBP-1s* engineered stable pools during a 13 day fed-batch overgrow.

Bar chart showing the average qP of DTE mAb from each *XBP-1s* engineered stable pool variant. The data highlights the average rate of mAb expression by each stable pool variant, indicating the variant most effectively expressing the recombinant protein. The error bars represent the SEM of biological triplicates. The '*' represents a significant difference ($p < 0.0332$) by a one-way ANOVA test with a Dunnett correction.

Figure 5-18 shows that the average qP of 5RPU-XBP1 pools was the highest among all the stable pool variants. Expectedly, the hCMV-MIE-XBP1 pools showed the lowest average qP and correlated well with the mAb titre shown in Figure 5-17, reiterating two of the biological replicates were compromised during stable pool generation. Interestingly, the average qP of the 10RPU-XBP1 and 20RPU-XBP1 pools were marginally lower than the control (not significantly difference by a one-way ANOVA test with a Dunnett correction), indicating the average rate of mAb expression was lower. Therefore, the enhanced volumetric titres observed in Figure 5-17 were caused by higher biomass accumulation rather than qP of the stable pool population.

5.2.2.6. Measuring mRNA levels of *XBP-1s* within engineered stable pool variants

The extraction and reverse transcription of RNA to generate cDNA was taken from engineered stable pool variants sampled on day 3 of routine subculturing. The mRNA fold change of *XBP-1s* was quantified by qPCR using the double delta Ct analysis by comparing against the control stable pools that represent no recombinant expression of *XBP-1s*. The average fold changes for each *XBP-1s* engineered stable pool were plotted in a bar chart as shown in Figure 5-19.

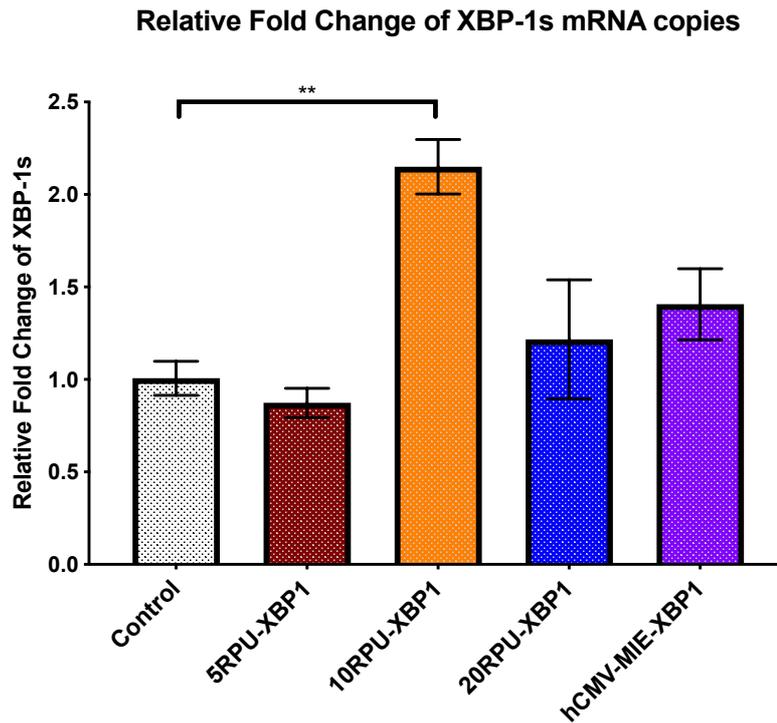


Figure 5-19: Fold change of *XBP-1s* mRNA copies of targeted integrated stable pool variants.

Bar chart showing the average fold change in *XBP-1s* mRNA copies of the engineered stable pools when comparing against the control pools. The data was generated using qPCR of each stable pool variant and fold change was calculated by double delta Ct analysis. The error bars represent the SEM of biological triplicates. The ‘**’ represents a significant difference ($p < 0.0021$) by a one-way ANOVA test with a Dunnett correction.

The highest average fold change in *XBP-1s* mRNA copies were observed by the 10RPU-XBP1 pools (2.15-fold) which was significantly different when using a one-way ANOVA test with a Dunnett correction, whereas the lowest fold change was observed by the 5RPU-XBP1 pools (0.87-fold). The difference between the 5RPU-XBP1 and control pools was marginal (not significantly different) and could be due to biological variation. The overall trend in fold change did not correlate with the relative transcriptional activity of the promoters, whereby an increasing gradient of *XBP-1s* expression was expected from the 5RPU-XBP1 to hCMV-MIE-XBP1 (1.41-fold).

Although, transcriptional interference from *LC* and *HC* expression would repress *XBP-1s* expression by a varying degree (as shown in Chapter 4). The difference in transcriptional activity of the promoter variants would still be observed.

Interestingly, the data provided a snapshot into the behavior and self-regulation of cells within the stable pool. For example, it has been documented that *XBP-1s* is a toxic protein associated with ERS and UPR stresses. Therefore, during selection of stable integration of the prototype, the cell population is also altered, whereby some cells could self-regulate genes by epigenetic effects for survivability. Therefore, the 10RPU-XBP1 pools over-expressed the *XBP-1s* gene at the tolerable level as shown by higher quantity of *XBP-1s* mRNA. Conversely, the 20RPU-XBP1 and hCMV-MIE-XBP1 were recombinantly expressing *XBP-1s* gene over the tolerable levels and hence may be repressed to survive selection. This also correlates with 2 out of 3 of the hCMV-MIE-XBP1 pools requiring an extended recovery period post-selection and demonstrating substantially lower DTE mAb expression.

However, even though *XBP-1s* mRNA copies were variable, the stable pools (10RPU-XBP1 and 20RPU-XBP1) expressing recombinant *XBP-1s* did show an average increase in mRNA copies relative to the control. When comparing with the growth profile and qP, it could be inferred that the increase in growth was a product of recombinant *XBP-1s* expression. Studies have shown that relative increases in *XBP-1s* expression can alleviate the stress within the ER and protein folding during recombinant therapeutic expression (Becker et al., 2008; Tigges and Fussenegger, 2006), and potentially benefit cell growth.

5.3. Discussion

MGEVs are a synthetic biology tool that has a vast range of applications within the biological sciences. This chapter explores two proof of concept studies that apply the MGEV as facilitating tool. The first study was the design and screening of various transcription terminator elements, where the MGEV system was used to determine the functionality of different variants. Hence, demonstrating the use of a MGEV to assist in the development of orthogonal or new genetic components for manipulating gene expression. The second study was applying the MGEV system to co-express a DTE mAb and *XBP-1s* TF in an aim to enhance stable mAb production. This is an example

of an engineering strategy that would be employed to increase biopharmaceutical production within industry using MGEVs. The conclusion of both studies are further explored in this section.

5.3.1. Design and screening of a transcription terminator library using a multi-gene expression vector

Data generated in Chapter 3 and 4 has shown variable levels of gene repression within a MGEV polycistronic cassette, as a consequence of transcriptional interference. The dynamic nature of interference is a by-product of multiple mechanisms contributing to the effect including promoter squelching (Hasegawa and Nakatsuji, 2002; Huliák et al., 2012) and RNA pol II occlusion of cis-acting genes. As shown in Section 1.7.2, literature has shown that the efficiency of transcription termination is a contributing factor towards RNA pol II occlusion-mediated interference and RNA pol II run-on could further lead to TF and PIC dislodgement of downstream genes (Callen et al., 2004; Palmer et al., 2011; Shearwin et al., 2005). Additionally, the current industry gold standard terminator is a SV40 late pA, and alternative elements have been identified within the mammalian genome. However, all of these elements are >240 bp in length and poorly characterised. Therefore, a limited number of commercially available transcription terminators which are fully characterised exist to achieve predictable gene regulation (Cheng et al., 2019). This warrants the design of multiple synthetic mammalian transcription terminators comprising of only functional motifs to minimise size, maintain termination functionality and provide some sequence diversity within a MGEV (to avoid undesired homologous recombination) (Curran et al., 2015) while facilitating modularity between TUs (Gasnov et al., 2015).

Although some functional conserved motifs have been identified from naturally occurring transcription terminator variants, limited knowledge is available for the construction and combination of motifs for efficient termination. In this study, commercially available transcription terminators were combined with synthetically designed terminators to attain a better understanding of transcription terminator design and in turn functionality. The aim was to identify optimal transcription terminator elements and design shorter elements to regulate gene expression and facilitate modulation within a MGEV. The first step was to identify and understand the mechanisms of transcription termination using previously published data and reviews

(Gromak et al., 2006; Proudfoot, 1991, 2011; White et al., 2013). This highlighted that pA and downstream elements potentially played a synergistic role in transcription termination, however, the exact mechanisms involved are not fully defined. The proposed mechanism included pA, CoTC, pause and ribozyme-mediated transcription termination (Libri, 2015; Porrua and Libri, 2015; Proudfoot, 2016). Therefore, terminator elements were designed based on the different mechanistic groups.

The conserved functional motifs and the design space were identified using literature leading to a library of synthetic pA, CoTC, pause and ribozyme elements (Colgan and Manley, 1997; Dower et al., 2004; Elkon et al., 2013; Gromak et al., 2006; West et al., 2008; White et al., 2013). Moreover, alternative commercially available pA elements were identified, and CHO-K1 genome analysis was performed to isolate native pA and CoTC elements. The entire library comprised of 42 sequences. The library was subdivided into 5 groups:

- Group 1 : Commercially available pA elements – a combination of viral and mammalian derived pA elements used in commercial vectors.
- Group 2: Synthetic pA elements – a library of 10 synthetic elements using functional motifs and element structures from a SV40 late pA.
- Group 3: Synthetic CoTC elements – a library of 12 synthetic elements derived from literature published conserved motifs (White et al., 2013).
- Group 4: CHO-K1 pA and CoTC homologous elements – identified downstream of transcriptionally active genes during transcriptomic and genome analysis.
- Group 5: Synthetic pause & ribozyme elements – using a conserved MAZ motif demonstrated in publications for pause-mediated termination (Gromak et al., 2006) and a hammerhead ribozyme motif for self-cleavage of the pre-mRNA transcript.

The library was *de novo* synthesised and cloned into the *SEAP* expression cassette, where the control terminator is a SV40_pA sequence. The various pA elements were substituted in exchange for the SV40 late pA element, while the CoTC, pause and ribozyme elements were cloned downstream of the SV40_pA element as per mechanistic functionality described in literature. The MGEV library was formed by

the construction of a bicistronic cassette co-expressing *SEAP* and *CL* (as shown in Figure 5-5). The dual gene system would screen for gene expression control and modularisation of the TUs. During synthesis, the Synth_CoTC_10 and Synth_Pause_4 could not be successfully made, furthermore during MGEV construction ActB_pA, ActB_CoTC and Synth_CoTC_4 were not assembled due to either presence of multiple BsaI sites or failed construction on multiple attempts. The remaining MGEV variants were constructed using Golden Gate assembly and positive variants were identified by restriction digest colony screen. Therefore, the final library of 36 MGEV variants were screened by HT transient transfection. The *SEAP* and *CL* expression were quantified by absorbance and luminescence assays and compared relatively against the control –SV40_pA.

The screen of commercially available pA elements demonstrated a decrease in expression of both recombinant proteins with a greater reduction of *SEAP* ($\geq 42.5\%$). Furthermore, there was no distinguishable trend of whether viral or mammalian pA variants were preferable. The data highlighted the current industrial standard (SV40_pA) is the best performing commercially available pA element. The library of synthetic pA variants derived from the SV40_pA showed lower expression of *SEAP* and *CL* by a varying degree, apart from Synth_pA_7 which showed matched *SEAP* and increased *CL* expression (16.7%). Additionally, comparing *SEAP* expression levels of different variants identified various design rules for creating a functional pA element. Firstly, DNA spacers between PAS motifs helped improve *SEAP* expression (Synth_pA_2 vs Synth_pA_3), moreover, the length of the DNA spacer was important for functionality as inferred by relative *SEAP* expression when utilising the Synth_pA_3 and 4 variants. Secondly, the DNA spacer between PAS and G/T rich motifs was also important towards gene expression as shown by the substantial increase in relative *SEAP* expression when utilising Synth_pA_2 vs Synth_pA_5 or 7. Literature has shown that two multi-protein complexes bind to the PAS and G/T rich motif to execute nascent RNA cleavage followed by polyadenylation (Elkon et al., 2013). Therefore, the importance of DNA spacers was rationalised as a potential means to avoid steric hinderance between the two essential complexes involved in 3' mRNA processing.

Finally, the number of G/T rich motifs (shown by marginal increase in relative *SEAP* expression when using the Synth_pA_5 and 7) and the sequence of G/T rich motifs (shown by decrease in *SEAP* expression employing the Synth_pA_8 variant) influences functionality of the pA. This suggests that the G/T rich motif composition such as frequency of G/Ts may affect the recruitment and binding affinity of the polyadenylation polymerase and associated poly(A) binding proteins (Nunes et al., 2010; Proudfoot, 2011). These observations have aligned well with multi-sequence alignments of eukaryotic genome analysis and other mammalian synthetic terminator designs which have identified similar conserved regions and importance of distance between the PAS and G/T rich motifs for pA functionality (Cheng et al., 2019; Proudfoot, 1991, 2011).

The synthetic CoTC library showed a decrease in both *SEAP* and *CL* expression compared to the SV40_pA by an average of 45.3% and 33.4% respectively. However, Synth_CoTC_6 demonstrated the least negative impact on expression and the comparison of relative expression fold change between *SEAP* and *CL* within the MGEV indicated the largest difference of 21.1%. This suggests the *CL* expression within the bicistronic cassette was relatively higher and could infer reduced repression of the downstream gene. Dissection of the element highlighted the 15 bp DNA spacer between A/T tracts had the lowest negative impact on expression of both *SEAP* and *CL*. When comparing naturally occurring CoTCs such as the human β -globin CoTC, the sequence comprises of A/T rich elements but are separated by longer nucleotide spacers leading to these elements being much larger (West et al., 2004, 2008; White et al., 2013). Therefore, it is unsurprising that longer DNA spacers are favourable. However, the spacer length could be extended further to potentially match wildtype CoTC functionality with the caveat to still keep the sequence length shorter than natural derivatives.

The CHO-K1 homologue screen also showed a reduction in *SEAP* and *CL* expression. Furthermore, after comparison of the difference in relative expression fold change within the MGEV variant, the pA homologues showed a larger difference between *SEAP* and *CL* than the CoTC. This may suggest either *SEAP* expression is hindered, or *CL* gene repression is marginally alleviated. The synthetic pause and ribozyme library showed a reduction in both *SEAP* and *CL* expression. Additionally, the

difference in gene expression between *SEAP* and *CL* within the bicistronic cassette was marginal indicating the presence of the pause or ribozyme element was equally affecting *SEAP* and the downstream *CL* gene expression.

In conclusion, the preliminary screening of the transcription terminator library indicated that the SV40_pA element is a proficient terminator in terms of *SEAP* and *CL* expression titres within the MGEV. The Synth_pA_7 element was the only synthetically designed variant to match and improve *SEAP* and *CL* expression respectively. Furthermore, this element was substantially shorter at 82 bp compared to the 241 bp SV40_pA. The remaining terminators demonstrated lower expression inferring an effect on 3' processing of the *SEAP* gene and transcriptional interference of *CL* due to inefficient transcription termination. However, the data narrowed the design space and identified additional important design attributes apart from the conserved motifs already mentioned in literature, such as DNA spacers between PAS and G/T motifs or between A/T rich motifs within the context of a pA and CoTC respectively. Therefore, a MGEV is functional tool for assisting in the development of synthetically designed genetic components.

In addition, to more accurately quantify and determine the degree of functionality of the terminators, a single gene vector (SGV) expressing *SEAP* or *CL* could be used. Whereby each vector would have a *GS* expression cassette upstream of the recombinant protein gene. Both of these SGVs would be co-transfected at equimolar ratio and identical gene copies to the MGEV. This control would demonstrate *SEAP* and *CL* expression with no RNA pol II occlusion-mediated interference. Therefore, expression regulated by the synthetic terminators within a MGEV could be directly compared to the SGVs to identify improvements in *CL* expression within the MGEV as a function of improved transcription termination. Furthermore, to characterise the impact of the terminator on gene regulation, *SEAP* only expression in a SGV context with the terminator variant downstream of the CDS would be informative. Apart from measuring *SEAP* and *CL* titre, using qPCR to perform absolute quantification of mRNA copies as a function of transcriptional activity would measure the rate of transcription termination to determine the efficiency. As a result, these additional experiments would comprehensively characterise the functionality of the terminator

library to provide further information on the design space of synthetic transcription terminators.

Although the preliminary screen has indicated a reduction in overall gene expression, this is still a form of gene regulation, specifically repression. Therefore, in combination with synthetic promoters, the synthetic terminators could facilitate additional control and gene regulation. For example, synthetic promoters, as shown in Chapter 4, demonstrate a fixed step-wise increase in titrated gene expression. However, combining with synthetic terminators, it could be possible to tune gene expression with precision by the marginal repressing nature of the current terminator library.

5.3.2. Application of multi-gene expression vectors to stably engineered CHO cells for enhancing biopharmaceutical production

Previous data published had shown the transient co-expression of chaperones and TFs alongside a DTE mAb can enhance mAb titres (Pybus et al., 2014; Tigges and Fussenegger, 2006). For example, Dr. Joe Cartwright and Dr. Claire Arnall (within the lab group) had demonstrated the transient co-expression of a DTE mAb and a specific gene dose of *XBP-Is* (10% of mAb DNA load during transfection) had increased titre by 240% (Cartwright et al., 2020). Therefore, the designed MGEV system was applied in combination with a range of synthetic promoters to express the DTE mAb and titrate recombinant *XBP-Is* expression in a stable cell line using MedImmune's proprietary TI system.

The first step to stable engineer MedImmune's CHO host cell line was to adapt the MGEV backbone vector by removing the *GS* expression cassette and introduce a promoter-less hygromycin resistance gene as a selection marker. As well as, a LoxP sequence element to control genome integration by Cre-mediated recombination into the landing pad within MedImmune's TI CHO host cell line. A series of prototype MGEVs were designed and constructed, where the *LC* and *HC* genes were regulated by MedImmune's hCMV-MIE promoter and the *XBP-Is* gene was regulated by a range of promoters (5RPU, 10RPU, 20RPU, hCMV-MIE) with varying transcriptional strength. The successfully constructed MGEVs were verified by restriction digest colony screen.

Triplicate TI stable pools were generated for each prototype MGEV variant by using MedImmune's proprietary protocol including selection by hygromycin over a period of 18 days. All pools were successfully established, however 4 pools (a control replicate pool, a 5RPU-XBP1 replicate pool and two hCMV-MIE-XBP1 replicate pools) required an extended recovery period. The established pools were characterised for growth and mAb expression through a 13 day FBOG using MedImmune's proprietary feeding regime.

The 20RPU-XBP1 demonstrated the highest VCC and the growth rate inferred by the gradient of the curve which was steeper compared to the other variants. Furthermore, the average mAb titre through the FBOG and upon harvest on day 13 was the highest at 1300 mg/L. The IVCD and qP analysis of the 20RPU-XBP1 pools indicated that the higher volumetric titre was the outcome of biomass accumulation rather than the rate of mAb expression per cell. Whereas, the 5RPU-XBP1 pools showed lowest VCC and IVCD with similar titres to the control apart from day 13 where it was lower by 13.3%. However, the average qP of these pools was the highest by 25.4% over the control. Interestingly, the 10RPU-XBP1 pools showed an increase in average volumetric titre over the control and was also a by-product of increased biomass accumulation due to the lower observed qP (12.7% lower than the control). Conversely, the hCMV-MIE-XBP1 pools showed the second highest average VCC and IVCD through the FBOG. However, the average volumetric titre and qP were the lowest, whereby upon harvest the average titre was 32.3% lower than the control. After assessing the replicates, it was highlighted that 2 out of the 3 replicates had substantially lower titres throughout the entire FBOG. Moreover, the lower expressing replicate pools were the same pools which struggled to recover post-selection.

The data observed did not correlate to previous transient studies where recombinant *XBP-Is* expression increased the qP of the MedI-CHO cells (Cartwright et al., 2020). Moreover, previous stably engineered *XBP-Is* CHO cell lines showed an increased qP when recombinantly expressing a mAb rather than an increase in biomass (Becker et al., 2008). However, the results from the Cartwright et al and Becker et al studies cannot be directly compared due to different experimental models. For example, Cartwright et al observations were derived from transient expression which may not be directly comparable to stable expression performance. Becker et al performed a

two-step stable cell line generation by first overexpressing *XBP-1s* followed by a mAb, both using random homologous recombination. Whereas, the study describe here used targeted integration and concurrent expression of a DTE mAb and *XBP-1s*, both of which could have generated a different phenotypic cell line than previously published data.

Additional analysis of *XBP-1s* expression was performed by quantifying mRNA copies by qPCR using the double delta Ct analysis against the control pools. The data highlighted that the 10RPU-XBP1 pools showed the highest average fold change in mRNA copies to the control by 2.15-fold. However, the 20RPU-XBP1 and hCMV-MIE-XBP1 pools were marginally higher than the control by 1.22 and 1.41-fold respectively. Therefore, the data indicate successful recombinant expression of XBP-1s within the MGEV. The mRNA copy fold change of *XBP-1s* in the 5RPU-XBP1 pools was marginally lower than the control, which is could be expected as the 5RPU promoter is the weakest strength and coupled with transcriptional interference-mediated gene repression could potentially lowering the *XBP-1s* expression to basal levels.

Interestingly, variable expression of XBP-1s was observed (by mRNA copy fold change) and did not align with the defined transcriptional activity of the respective promoter. This may be attributed to differential regulation of *XBP-1s* in both the control (no recombinant *XBP-1s* expression) and recombinantly expressing stable pools (10RPU-XBP1, 20RPU-XBP1 and hCMV-MIE-XBP1). For instance, an assumption is made when using double delta Ct analysis that endogenous *XBP-1s* expression between stable pool variants is constant, therefore, the fold change is indicative of recombinant expression. However, it is known that *XBP-1s* is part of an induced response by ER stress and UPR pathway where its expression and splicing is regulated by *ATF6* and *IRE1* respectively (Walter and Ron, 2011). Moreover, *XBP-1s* has also been known to self-regulate its own expression (Prashad and Mehra, 2015). Therefore, due to the potential variability of *XBP-1s* expression within control pools, absolute quantification of *XBP-1s* mRNA copies would more effectively discern trends in recombinant *XBP-1s* expression. Moreover, through targeted integration, an assumption is made that a single copy of the plasmid was integrated into the host cell genome. Therefore, a means to verify this integration step would confirm that variable

recombinant *XBP-Is* expression is dictated by the transcriptional activity of the promoter rather than multiple unexpected copies of the integrant within the genome.

Another potential hypothesis for the variation in *XBP-Is* expression regulated by the 10RPU, 20RPU and hCMV-MIE promoter is that *XBP-Is* is a stress inducer TF, and is known to be toxic to cells (Wang et al., 2015; Zeng et al., 2009). Therefore, unregulated overexpression of *XBP-Is* is toxic to CHO cells and can cause poor recovery of stable pools due to induced apoptosis (Becker et al., 2010). Alternatively, other publications have established *XBP-Is* to halt apoptosis and extend mammalian cell tolerance to ERS and UPR events (Gomez et al., 2007; Jäger et al., 2012). This suggests the level and regulation of recombinant *XBP-Is* expression is crucial towards survivability and avoiding an apoptotic reaction (Gulis et al., 2014). Therefore, the observations of lower recombinant expression of *XBP-Is* driven by 20RPU and hCMV-MIE promoters than the 10RPU promoter could be a by-product cellular self-regulation by gene silencing during selection or culturing to maintain survivability. This is because the 20RPU and hCMV-MIE promoters' transcriptional activity may be too high and the accumulation of *XBP-Is* would have led to apoptosis. Conversely, the 10RPU promoter strength may have been on the level of tolerance for the stable pools to survive with the overexpression of *XBP-Is*.

This hypothesis also aligns with the poor recovery and mAb expression of 2 out of 3 hCMV-MIE-XBP1 pools. The substantial suppression of mAb expression within replicate pools could be caused by subpopulations silencing recombinant gene expression either by epigenetic mechanisms or gene loss within the heterogenous population (Kim et al., 2011) during selection. Therefore, subpopulations with suppressed mAb expression would experience a growth advantage over producing cells due to reduced stress and metabolic burden invoked by overexpression of a DTE mAb and *XBP-Is* (Kim et al., 2011; Lee et al., 1991) leading to a stable pool surviving hygromycin selection but not expressing the recombinant protein of interest.

Interestingly, although lower recombinant *XBP-Is* expression driven by the 20RPU promoter was observed (as quantified by qPCR), the growth (by IVCD) and mAb expression (by titre) were the highest of all stable pool variants. This could be attributed to the selection process isolating a subpopulation where *XBP-Is* expression was regulated by epigenetic effects to a favourable level yielding improved DTE mAb expression. This does reduce the predictability of genetic interventions on host cell engineering. However, in the end, the engineering process did lead to a faster growing and higher DTE mAb producing cell line which is an ideal outcome for biopharmaceutical production applications.

Chapter 6

6. Conclusion and Future Work

6.1. Conclusion

The results generated throughout the PhD project have led to, firstly, the design and characterisation of a multi-gene expression vector (MGEV) for Chinese hamster ovary (CHO) cell engineering. Secondly, application of synthetic promoters within a MGEV has allowed titration of gene expression and defining stoichiometric ratios. Finally, demonstrating the functionality of a MGEV as a research tool for the development of synthetic transcription terminators and stably engineering CHO cells for biopharmaceutical production. The specific conclusions drawn from each of these outcomes throughout the project are further explored in this section.

6.1.1. Design and characterisation of a multi-gene expression vector for CHO cell engineering

After an extensive design process including identifying the essential design space and the desired features, a MGEV system was developed for functionality in CHO cell engineering that is applicable in both an academic and industrial environment. The system comprises of two major components, a transcription unit (TU) and a pExp-Vec-GG recipient vector. The *in silico* designed TU facilitates the *in vitro* manipulation of gene expression by utilising various genetic components such as a proximal element of a promoter, coding DNA sequence (CDS), polyadenylation (pA) elements and alternative transcription terminator or insulator elements to generate self-functioning expression cassettes. The pExp-Vec-GG vector provides stable and long

term transient expression capabilities along with performing well-established DNA amplification strategies.

The rapid assembly of MGEVs was achieved by using the unique features of Golden Gate assembly and *in silico* designed linkers to achieve controlled, directional and multi-TU cloning into the pExp-Vec-GG vector. The *in vitro* verification of the cloning technique showed rapid, reproducible and robust construction of 70 MGEVs across the entire project duration, indicating an effective design for a MGEV construction. The co-expression of three fluorescent proteins within a MGEV demonstrated successful functionality of multiple recombinant gene expression. Furthermore, the quantification displayed differential levels of downstream gene repression within a polycistronic cassette when each gene was regulated by a hCMV-MIE promoter. It was hypothesised that the gene repression was a consequence of transcriptional interference.

6.1.2. Application of synthetic promoters in a multi-gene expression vector to titrate gene expression at stoichiometric ratios in CHO cells

A library of synthetic promoters developed by Dr. Adam J. Brown (Brown et al., 2017) were first adapted for compatibility with the MGEV system and verified for expression strength using an alternative reporting system. The verified synthetic promoters with differing transcriptional strength were applied within a MGEV to display titratable gene expression. This was shown by using a combination of a low, medium and high strength synthetic promoter within a MGEV. The differential strength of gene expression within the MGEV was detected by qPCR and flow cytometry.

These promoters were further characterised by measuring combinatorial gene expression regulating different fluorescent reporters within a single gene plasmid context by flow cytometry to demonstrate repressed expression. This repression could be caused by promoter squelching where the synthetic promoters are competing for transcription factors (TFs) and other transcription regulating cofactors (Huliák et al., 2012; Schmidt et al., 2016). Moreover, the metabolic burden endured by the cells during co-expression of multiple fluorescent proteins simultaneously could also contribute towards repression. Additionally, the positional-mediated interference within a MGEV was quantified by qPCR and flow cytometry to demonstrate that the

effect was dynamic and dependent on transcriptional strength. This was due to interference being a cumulation of multiple mechanisms such as RNA polymerase II (RNA pol II) occlusion, TF and pre-initiation complex (PIC) dislodgement or promoter squelching (Curtin et al., 2008; Huliák et al., 2012; Palmer et al., 2011; Shearwin et al., 2005; West, 2014). Due to the positional effect being dependent on transcriptional activity, it would be difficult to predict expression of multiple recombinant gene based on the transcriptional strength of the promoter.

Therefore, a library of 27 MGEV variants was constructed encompassing every combination of a low, medium and high strength synthetic promoter within the polycistronic cassette. The normalised mRNA copies (RTA) was compared against a set of expected values (derived from equimolar single gene vector (SGV) co-expression at similar gene copies) to identify substantial (70%) transcriptional repression within the MGEV context. This general repression could be attributed to negative supercoiling where a change in the plasmid structure post-RNA pol II-mediated transcription elongation results in transcription inhibition of the upstream genes (Corless and Gilbert, 2017; Curtin et al., 2008; Ma and Wang, 2016). Alternatively, the potential inherent bidirectional behaviour of promoters, where transcription can occur both on the sense and antisense strand (Seila et al., 2008; Wei et al., 2011) can inhibit transcription of neighbouring TUs (Curtin et al., 2008). Unsurprisingly, gene expression variation was observed between each MGEV variant (potentially caused by transcriptional interference), however, other specific transcriptional repression trends were also identified. For example, the medium strength synthetic promoter was consistently repressed and potentially caused by competition of TFs, as the cognate transcription factor regulatory elements (TFREs) were shared between the different synthetic promoters leading to squelching (Huliák et al., 2012; Schmidt et al., 2016). Additionally, the low strength synthetic promoter exhibited enhanced activity when neighbouring a higher strength synthetic promoter which could be a consequence of interacting TFs between C/EBP and NF κ B derivatives (Maehara et al., 1999).

Due to the variability in multi-gene expression, empirically derived gene expression stoichiometric ratios (as demonstrated in this study) can be an alternative approach to achieving predictable expression with a MGEV. This possible since the gene

expression performance within a MGEV context has been characterised by testing every combination of synthetic promoter permutation within a polycistronic cassette. However, the derived ratios are specific to the recombinant genes expressed and can provide an approximate guidance towards relative gene expression levels. However, expression testing is a pre-requisite to achieve precise and predictable stoichiometry when utilising a different panel of recombinant genes.

Additionally, through this study's observations, new design parameters can be considered for future MGEV construction in order to achieve predictable and defined gene expression stoichiometry. Firstly, synthetic promoters can be designed using distinct TFRE composition, where TFRE-blocks are not shared between promoter variants to avoid competition of TFs. Additionally, naturally occurring viral promoters (hCMV-MIE) have exhibited lower transcriptional repression within a MGEV context (shown in this study to be 32.5%). This suggests viral promoters have evolved to alleviate repression by developing a more complex TFRE composition (Brown et al., 2015; Stinski and Isomura, 2008). Therefore, increasing the complexity of synthetic promoters by using non-canonical TFRE motifs (Wong et al., 2011) can expand the repertoire of TFs recruited and alleviate transcriptional repression.

Secondly, in order to eliminate positional repression, the introduction of an efficient transcription terminator and insulator element could be included between TUs within the MGEV (Tian and Andreadis, 2009). For example, the introduction of a β -globin CoTC element can reduce transcriptional interference by improving transcription termination efficiency (White et al., 2013), whereas a chicken hypersensitivity site 4 (cHS4) insulator would avoid distal promoter mediated transcription by functioning as an enhancer-blocker (Liao et al., 2018). However, both of these elements are large (ranging from 800 to 1200 bp) (Aker et al., 2007; West et al., 2008) resulting in increased plasmid size and risking cellular toxicity and poor transfectability of the MGEV (Hornstein et al., 2016; Lesueur et al., 2016). More recently, the development of shorter (250 bp) core CCCTC-enriched elements derived from the cHS4 element has maintained insulator functionality and could be a viable option to alleviate positional effects (Aker et al., 2007; Liao et al., 2018).

In conclusion, gene expression stoichiometry can be achieved for complex multi-gene engineering by either utilising empirically-derived stoichiometric ratios or applying new MGEV design criteria to counteract mechanisms contributing towards transcriptional repression.

6.1.3. Proof of concept applications of multi-gene expression vectors as a research tool

Among the vast number of applications possible using MGEVs, two proof of concept studies were performed to demonstrate the functionality of a MGEV towards synthetic component design and stable multi-gene engineering of CHO cells.

The *in silico* designed synthetic terminator elements were screened using a bicistronic cassette within the MGEV to simultaneously measure regulation of gene expression and achieving modularisation. The preliminary screen assisted in identifying one potential synthetic pA element demonstrating the desired functionality of maintaining secreted alkaline phosphatase (*SEAP*) and increasing *Cypridina* luciferase (*CL*) expression within a MGEV. Additionally, the MGEV system provided a rapid means to narrow the design space on synthetic terminators by measuring the impact of conserved motifs on gene expression and transcription termination functionality. However, additional and alternative means of screening would be required to definitively characterise the entire terminator library for gene expression regulation and modularisation of TUs. The study also demonstrated that majority of synthetic terminator variants exhibited reduced recombinant gene expression within a MGEV context. A potential application of this observed behaviour is to attain more tuneable gene expression by combining the repressive and enhancing nature of the synthetic terminators and promoters respectively.

A second proof of concept was demonstrating the capacity of a MGEV to perform CHO cell engineering. This was demonstrated by enhancing expression of a difficult-to-express (DTE) monoclonal antibody (mAb) by co-expressing the x-box binding protein 1 spliced (*XBP-1s*) TF within a targeted integrated (TI) stable pool. Previous data had indicated the specific gene dose of *XBP-1s* during transient expression significantly increased mAb titres. Therefore, a range of synthetic promoters were used to titrate recombinant *XBP-1s* expression within the MGEV. The engineered stable

pools showed enhanced growth and biomass accumulation compared to the control indicating phenotypic engineering of the CHO cell by using a MGEV. The increase in mAb titre was lower than shown by transient expression, furthermore, the qPCR data showed uncorrelated levels of *XBP-1s* mRNA copies with the respective promoter strength driving expression. A potential reason for the variation observed is that endogenous *XBP-1s* is induced by the UPR pathway and is known to regulate its own expression (Prashad and Mehra, 2015; Walter and Ron, 2011). Therefore, the relative recombinant *XBP-1s* expression fold change could be skewed relative to the control pools due to dynamic regulation of endogenous *XBP-1s* expression. Therefore, double delta Ct analysis can provide an indication of recombinant *XBP-1s* expression, but absolute mRNA copy quantification would more accurately demonstrate trends in recombinant *XBP-1s* expression. Another hypothesis for the varying *XBP-1s* expression levels could be that unregulated overexpression of *XBP-1s* is toxic to CHO cells and can cause poor recovery of stable pools due to induced apoptosis (Becker et al., 2010). Therefore, subpopulations with suppressed *XBP-1s* expression would experience a growth advantage over producing cells due to reduced stress and metabolic burden invoked by overexpression of *XBP-1s* (Kim et al., 2011; Lee et al., 1991) leading to a stable pool population surviving hygromycin selection but exhibiting lower than expected recombinant *XBP-1s* expression.

In summary, the MGEV system has demonstrated successful functionality to facilitate synthetic genetic component design and a tool for CHO cell engineering to improve biopharmaceutical production.

6.2. Future Works

This section explores the future work that would be required to further improve the design of the MGEV system for CHO cell engineering. This includes to further characterise the gene expression dynamics within the MGEV to better quantify and understand the effects. To design elements to counteract transcriptional interference within a MGEV. Finally, to combine all the knowledge, to design and test an insulated MGEV with predicted gene expression stoichiometric ratios as a tool to successfully engineer CHO cells. The specific aspects of the future studies are expanded further in this section.

6.2.1. Improved design of the multi-gene expression vector for CHO cell engineering

Through the duration of the PhD various genetic components including promoters, CDSes and transcription terminator sequences were changed within the TU using the designed ‘plug-and-play’ system. However, a limitation in the current design is that the core element of the promoter and the 5′ untranslated region (UTR) are not independent components within the TU. Therefore, a simple improvement would be to add a Sall restriction endonuclease site between the core element and 5′UTR. This endonuclease is co-functional with the existing enzymes and the overhang post-digestion is non-homologous to other sticky ends, in turn, avoiding mismatch ligation. This improvement would allow for more effective manipulation of gene expression within a TU by changing both the proximal and core elements of a promoter (Ede et al., 2016; Juven-Gershon and Kadonaga, 2010). Furthermore, active research is ongoing within the group towards design of synthetic and optimal 5′UTRs to enhance mRNA stability, as well as, for enhancing recombinant protein expression by influencing translation rates (Petersen et al.; Zucchelli et al., 2016).

In this project large libraries of synthetic promoters and transcription terminators had to be cloned into TUs, this highlighted that a ‘plug-and-play’ system using digestion-ligation cloning is still time consuming. Moreover, the direction towards fully synthetic expression cassettes including synthetic UTRs and signal sequence peptides, requires a HT means to assemble TUs and facilitate a design driven approach. Therefore, a rapid cloning technique is necessary to assemble a TU from a library of multiple genetic components. One such approach is to use Golden Gate assembly by using an alternative Type IIs restriction endonuclease such as BbsI or Esp3I. The method would be identical to assembling a MGEV, where a series of genetic components are selected and constructed together by *in silico* designed linkers to regulate order and orientation within a TU (Engler et al., 2008). An alternative approach is to use Gibson assembly by flanking the same components with unique nucleotide sequences (UNSeqs) and use sequence homology driven order and orientation controlled cloning. The TU would be constructed using an isothermal reaction combining the function of a restriction endonuclease, T5 exonuclease, a polymerase and T4 DNA ligase (Torella et al., 2014a).

6.2.2. Further characterisation of gene expression dynamics within a multi-gene expression vector

Current data has shown promoter squelching within the synthetic promoter library was observed by flow cytometry-based quantification. However, co-expression of multiple fluorescent proteins can invoke a metabolic burden on the host cell, whereas promoter squelching is an effect occurring transcriptionally, both of which would exhibit gene repression. Therefore, flow cytometry does not distinguish between metabolic burden and promoter squelching. As a result, performing absolute quantification by qPCR would measure mRNA copies as a function of transcriptional activity of the synthetic promoters. This assay would quantify the effect of self- and cross-interaction of promoters and in turn assist in understanding the transcriptional dynamics within a CHO cell.

The library of 27 MGEVs constructed in this study titrates express by utilising a low, medium and high strength synthetic promoter in every combination within a polycistronic cassette. However, the positional-mediated transcriptional interference was calculated using the same promoter strength in each position only. Therefore, to comprehensively understand the dynamics of positional effect, quantifying the variability across the entire library of 27 MGEVs would be beneficial. This could be achieved by transfecting the MGEV and co-transfecting the respective TUs as SGVs at the same gene copies, as a control. Thus, directly comparing the same combination of synthetic promoters and recombinant proteins but only differentiating between the formation of a polycistronic cassette against independent TUs on separate plasmids to quantify the positional-mediated transcriptional interference.

Additionally, the library of 27 MGEVs determines the gene expression stoichiometric ratios of using a 5RPU, 40RPU and 80RPU synthetic promoter. Further quantification can be performed using the 10RPU, 20RPU, 60RPU and 100RPU (the 10RPU and 100RPU have since been *de novo* synthesised) to expand the database of quantified gene expression stoichiometric ratios within a MGEV. This expands the database of available gene expression stoichiometric ratios to guide multi-gene engineering. Furthermore, these ratios account for the dynamic nature of transcriptional interference.

6.2.3. Achieving modularisation within a polycistronic cassette of a multi-gene expression vector

Preliminary screens within the MGEV assisted in determining the raw functionality of the transcription terminator library and deducing essential features within the design space. The data indicated that majority of the library negatively impacted on the gene expression of both *SEAP* and *CL*, indicating the terminators were non-functional in enhancing gene expression by efficient transcription termination and mitigating transcriptional interference (Proudfoot, 2016; West and Proudfoot, 2009). However, more quantifiable data could be generated by performing additional experiments to comprehensively characterise the terminator library in terms of dissecting relative impact on *SEAP* and *CL* expression within the bicistronic cassette. For example, using a single gene control of *SEAP* and *CL* that are co-transfected at identical gene copies to the MGEV would accurately quantify the positional effect. Ergo, the level of *SEAP* and *CL* expression with the addition of a transcription terminator variant would determine the effect on transcriptional interference. Additionally, a single TU expressing *SEAP* only with the terminator variant downstream should be measured, to further understand the impact on 3' end processing and gene regulation.

An alternative assay to measure *SEAP* and *CL* expression is using absolute quantification of mRNA copies by qPCR. The assay would allow direct comparison of *SEAP* and *CL* expression within a MGEV or in SGVs, and the output would be a function of transcriptional activity. As such, the level of downstream *CL* expression could infer the rate of transcription termination. In addition, a nuclear run-on assay provides an alternative means to measure termination efficiency by quantifying the population of RNA pol II across a coding DNA strand. This is observed by hybridisation of different probes to southern blots and band intensity is quantified by densitometry. Similarly, a northern blot can be used to measure the length of the mRNA transcript of interest by using a sequence specific hybridisation probe. The length of the mRNA transcript and density of the band would be indicators of RNA pol II dissociation and transcription termination.

The outcome of the above experiments coupled with the screening of transcription terminators with the MGEV would narrow the design space and assist in designing a

2nd generation library to enhance gene expression and provide modularisation of TUs by improving transcription termination efficiency.

Another approach for achieving modularisation of MGEVs is by using an insulator element. An insulator element is defined as a non-coding regulatory DNA sequence that protects genes from undesired signals by functioning as barrier and an enhancer blocker (Liao et al., 2018; West et al., 2002). Examples of insulators identified are *gypsy* in *Drosophila* and 5' HS4 upstream of the chicken β globin locus. Both of these elements have been proven to insulate in the genome (Yusufzai and Felsenfeld, 2004). Additionally, a core element within the cHS4 element have been implemented to block promoter interference (Uchida et al., 2013). However, synthetic versions of insulator elements have not been designed and tested to replicate the same functionality. Therefore, the demand for synthetic elements which are non-native to the host genome but provide modulation between TUs within a MGEV is warranted.

The application of both the synthetic transcription terminator and insulator elements would assist in achieving predictable gene expression stoichiometric ratios driven by the transcriptional activity of the synthetic promoters by omitting the effects of transcriptional interference.

6.2.4. Validating the application of multi-gene expression vectors for stable CHO cell engineering

Previous studies have shown optimising the dose of light chain (*LC*) and heavy chain (*HC*) genes of a DTE mAb can increase transient recombinant expression (Pybus et al., 2014). More recently, data generated by Dr. Joe Cartwright (within the lab group) has demonstrated that transiently expressing additional *LC* and *XBP-Is* into a stable pool cell line producing a DTE mAb has increased titres by 312% (Cartwright et al., 2020). Indicating that an optimal *LC* and *HC* ratio of a mAb and supplemented with recombinant *XBP-Is* can improve expression levels within CHO cells. Therefore, using the database of gene expression stoichiometric ratios as a guidance to identify a combination of synthetic promoters and positions within a MGEV to perform the same engineering strategy and generate stably expressing CHO cell lines. However, to achieve precise stoichiometry, a pre-requisite screening of the recombinant genes of interest is required to characterise the gene expression dynamics within a MGEV. The

application of MedImmune's TI cell system will control the number of MGEV gene copies integrated. Therefore, the resulting data would indicate the influence of transcriptional power of the promoters to effectively stably engineer the CHO cell for biopharmaceutical production.

Another approach to further validate the functionality of the MGEV as a synthetic biology multi-gene engineering tool for CHO cells, is by phenotypically modifying the host cell. For example, a high throughput (HT) screen of accessory genes may identify 3 genes that in synergy at the optimal dose would successfully modify the CHO cell for a favourable phenotype. By using a combination of synthetic promoters to titrate the genes of interest, terminators/insulators to modulate the TUs to achieve predictable stoichiometric ratios and stably engineering the CHO cell by using MedImmune's TI system, a phenotype which increases mAb expression can be generated. This newly multi-gene engineered cell line would be extensively tested by performing fed-batch evaluation to measure growth characteristics, mAb expression and verifying successful stoichiometric engineering by qPCR quantification. Furthermore, the stability of the engineered cell line would be investigated by testing the cell line after 60 generations.

The MGEV system provides the capacity to effectively engineer CHO cell factories for biopharmaceutical production by combining synthetic genetic components to control gene expression with multi-gene engineering. This was demonstrated from the *XBP-1s* engineering strategy in this study. However, a difference in enhanced mAb expression was observed between previously generated transient data and stable expression. This is hypothesised to be a consequence of substantial self-regulation of the CHO cell during stable cell line generation and selection, leading to subpopulations with recombinant *XBP-1s* expression being repressed from the expected levels (dictated by the strength of the promoter) to survive the potential apoptotic effect from overexpression of *XBP-1s* (Becker et al., 2010). This indicates that a different design space is required for transient and stable cell line engineering. Consequently, Cartwright et al discusses a paradigm shift occurring in the approach to screening and implementing CHO cell multi-gene engineering (Cartwright et al., 2020). For example, the first step to perform multi-gene engineering in biopharmaceutical production is to identify an optimal *LC* and *HC* expression ratios to optimise mAb

expression (Pybus et al., 2014). This stoichiometric ratio can be achieved stably by using synthetic promoters within a MGEV context. To further enhance mAb expression, particularly DTE mAbs, a transient HT screen of various accessory genes (part of the UPR or secretion pathway) is performed to determine an optimal combination and stoichiometry of an accessory gene subset. This combination and gene stoichiometry can be stably implemented by targeted integration into a mAb producing stable pool using synthetic promoters within a MGEV. Therefore, the MGEV system can be used to execute complex multi-gene engineering strategies at a defined stoichiometry.

References

Agrawal, V., and Bal, M. (2012). Strategies for Rapid Production of Therapeutic Proteins in Mammalian Cells. *Bioprocess Int.* *10*, 32–48.

Ajikumar, P.K., Xiao, W.-H., Tyo, K.E.J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., and Stephanopoulos, G. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* *330*, 70–74.

Aker, M., Tubb, J., Groth, A.C., Bukovsky, A.A., Bell, A.C., Felsenfeld, G., Kiem, H.-P., Stamatoyannopoulos, G., and Emery, D.W. (2007). Extended core sequences from the cHS4 insulator are necessary for protecting retroviral vectors from silencing position effects. *Hum. Gene Ther.* *18*, 333–343.

Andreou, A.I., and Nakayama, N. (2018). Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLOS ONE* *13*, e0189892.

Bandaranayake, A.D., and Almo, S.C. (2014). Recent advances in mammalian protein production. *FEBS Lett.* *588*, 253–260.

Barrett, L.W., Fletcher, S., and Wilton, S.D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* *69*, 3613–3634.

Bayat, H., Hossienzadeh, S., Pourmaleki, E., Ahani, R., and Rahimpour, A. (2018). Evaluation of different vector design strategies for the expression of recombinant monoclonal antibody in CHO cells. *Prep. Biochem. Biotechnol.* *48*, 160–164.

Beck, A., Cochet, O., and Wurch, T. (2010). GlycoFi's technology to control the glycosylation of recombinant therapeutic proteins. *Expert Opin. Drug Discov.* *5*, 95–111.

Becker, E., Florin, L., Pfizenmaier, K., and Kaufmann, H. (2008). An XBP-1 dependent bottle-neck in production of IgG subtype antibodies in chemically defined serum-free Chinese hamster ovary (CHO) fed-batch processes. *J. Biotechnol.* *135*, 217–223.

Becker, E., Florin, L., Pfizenmaier, K., and Kaufmann, H. (2010). Evaluation of a combinatorial cell engineering approach to overcome apoptotic effects in XBP-1(s) expressing cells. *J. Biotechnol.* *146*, 198–206.

Berlec, A., and Strukelj, B. (2013). Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J. Ind. Microbiol. Biotechnol.* *40*, 257–274.

Bernard, H.-U., Krämmer, G., and Röwekamp, W.G. (1985). Construction of a fusion gene that confers resistance against hygromycin B to mammalian cells in culture. *Exp. Cell Res.* *158*, 237–243.

References

- Birch, J.R., and Racher, A.J. (2006). Antibody production. *Adv. Drug Deliv. Rev.* *58*, 671–685.
- Bird, G., Fong, N., Gatlin, J.C., Farabaugh, S., and Bentley, D.L. (2005). Ribozyme Cleavage Reveals Connections between mRNA Release from the Site of Transcription and Pre-mRNA Processing. *Mol. Cell* *20*, 747–758.
- Blazeck, J., Garg, R., Reed, B., and Alper, H.S. (2012). Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnol. Bioeng.* *109*, 2884–2895.
- Brown, A.J., and James, D.C. (2016). Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnol. Adv.* *34*, 492–503.
- Brown, A.J., and James, D.C. (2017). Constructing Strong Cell Type-Specific Promoters Through Informed Design. In *Mammalian Synthetic Promoters*, D. Gould, ed. (New York, NY: Springer), pp. 131–145.
- 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- κ B, 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 silico design of context-responsive mammalian promoters with user-defined functionality. *Nucleic Acids Res.* *45*, 10906–10919.
- Brown, A.J., Gibson, S., Hatton, D., and James, D.C. (2018). Transcriptome-Based Identification of the Optimal Reference CHO Genes for Normalisation of qPCR Data. *Biotechnol. J.* *13*, 1700259.
- Butler, M., and Spearman, M. (2014). The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr. Opin. Biotechnol.* *30*, 107–112.
- Bywater, M.J., Pearson, R.B., McArthur, G.A., and Hannan, R.D. (2013). Dysregulation of the basal RNA polymerase transcription apparatus in cancer. *Nat. Rev. Cancer* *13*, 299–314.
- Cahill, M.A., Ernst, W.H., Janknecht, R., and Nordheim, A. (1994). Regulatory squelching. *FEBS Lett.* *344*, 105–108.
- 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- α has increased levels of transient protein expression. *Biotechnol. Prog.* *29*, 697–706.
- Callen, B.P., Shearwin, K.E., and Egan, J.B. (2004). Transcriptional Interference between Convergent Promoters Caused by Elongation over the Promoter. *Mol. Cell* *14*, 647–656.

- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* *38*, 626–635.
- Cartwright, J.F., Arnall, C.L., Patel, Y.D., Barber, N.O.W., Lovelady, C.S., Rosignoli, G., Harris, C.L., Dunn, S., Field, R.P., Dean, G., et al. (2020). A platform for context-specific genetic engineering of recombinant protein production by CHO cells. *J. Biotechnol.* *312*, 11–22.
- Casteleijn, M.G., Urtti, A., and Sarkhel, S. (2013). Expression without boundaries: cell-free protein synthesis in pharmaceutical research. *Int. J. Pharm.* *440*, 39–47.
- Chao, R., Yuan, Y., and Zhao, H. (2014). Recent advances in DNA assembly technologies. *FEMS Yeast Res.*
- Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* *12*, 326–328.
- Chen, Y.-J., Liu, P., Nielsen, A.A.K., Brophy, J.A.N., Clancy, K., Peterson, T., and Voigt, C.A. (2013). Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* *10*, 659–664.
- Cheng, J.K., Morse, N.J., Wagner, J.M., Tucker, S.K., and Alper, H.S. (2019). Design and Evaluation of Synthetic Terminators for Regulating Mammalian Cell Transgene Expression. *ACS Synth. Biol.* *8*, 1263–1275.
- Cho, S.W., Kim, S., Kim, J.M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* *31*, 230–232.
- Chua, K.J., Kwok, W.C., Aggarwal, N., Sun, T., and Chang, M.W. (2017). Designer probiotics for the prevention and treatment of human diseases. *Curr. Opin. Chem. Biol.* *40*, 8–16.
- Cobb, R.E., Ning, J.C., and Zhao, H. (2014). DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* *41*, 469–477.
- Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* *11*, 2755–2766.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* *339*, 819–823.
- Coralli, C., Cemazar, M., Kanthou, C., Tozer, G.M., and Dachs, G.U. (2001). Limitations of the Reporter Green Fluorescent Protein under Simulated Tumor Conditions. *Cancer Res.* *61*, 4784–4790.

- Corless, S., and Gilbert, N. (2017). Investigating DNA supercoiling in eukaryotic genomes. *Brief. Funct. Genomics* *16*, 379–389.
- Costa, R.A., Elisa Rodrigues, M., Henriques, M., Azeredo, J., and Oliveira, R. (2010). Guidelines to cell engineering for monoclonal antibody production. *Eur. J. Pharm. Biopharm. Off. J. Arbeitsgemeinschaft Für Pharm. Verfahrenstechnik EV* *74*, 127–138.
- Curran, K.A., Morse, N.J., Markham, K.A., Wagman, A.M., Gupta, A., and Alper, H.S. (2015). Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth. Biol.* *4*, 824–832.
- Curtin, J.A., Dane, A.P., Swanson, A., Alexander, I.E., and Ginn, S.L. (2008). Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct. *Gene Ther.* *15*, 384–390.
- 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.
- Darrah, P.A., Patel, D.T., Luca, P.M.D., Lindsay, R.W.B., Davey, D.F., Flynn, B.J., Hoff, S.T., Andersen, P., Reed, S.G., Morris, S.L., et al. (2007). Multifunctional TH1 cells define a correlate of vaccine-mediated protection against *Leishmania major*. *Nat. Med.* *13*, 843–850.
- D. Craggs, T. (2009). Green fluorescent protein : structure, folding and chromophore maturation. *Chem. Soc. Rev.* *38*, 2865–2875.
- De Jesus, M.J., and Wurm, F.M. (2013). Scale-up and predictability in process development with suspension cultures of mammalian cells for recombinant protein manufacture: comments on a trend reversal. *Pharm. Bioprocess.* *1*, 333–335.
- Deans, T.L., Cantor, C.R., and Collins, J.J. (2007). A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* *130*, 363–372.
- Dekker, L., and Polizzi, K.M. (2017). Sense and sensitivity in bioprocessing—detecting cellular metabolites with biosensors. *Curr. Opin. Chem. Biol.* *40*, 31–36.
- Dobhoff-Dier, O., and Bliem, R. (1999). Quality control and assurance from the development to the production of biopharmaceuticals. *Trends Biotechnol.* *17*, 266–270.
- Dower, K., Kuperwasser, N., Merrikh, H., and Rosbash, M. (2004). A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. *RNA* *10*, 1888–1899.
- Duan, F.F., Liu, J.H., and March, J.C. (2015). Engineered Commensal Bacteria Reprogram Intestinal Cells Into Glucose-Responsive Insulin-Secreting Cells for the Treatment of Diabetes. *Diabetes* *64*, 1794–1803.

References

- Duportet, X., Wroblewska, L., Guye, P., Li, Y., Eyquem, J., Rieders, J., Rimchala, T., Batt, G., and Weiss, R. (2014). A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* gku1082.
- Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* 32, 127–139.
- Ede, C., Chen, X., Lin, M.-Y., and Chen, Y.Y. (2016). Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. *ACS Synth. Biol.* 5, 395–404.
- Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* 14, 496–506.
- Engler, C., and Marillonnet, S. (2011). Generation of families of construct variants using golden gate shuffling. *Methods Mol. Biol. Clifton NJ* 729, 167–181.
- Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3, e3647.
- Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One* 4, e5553.
- Eszterhas, S.K., Bouhassira, E.E., Martin, D.I.K., and Fiering, S. (2002). Transcriptional Interference by Independently Regulated Genes Occurs in Any Relative Arrangement of the Genes and Is Influenced by Chromosomal Integration Position. *Mol. Cell. Biol.* 22, 469–479.
- Fan, L., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M., and Frye, C.C. (2012). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnol. Bioeng.* 109, 1007–1015.
- Fan, L., Frye, C.C., and Racher, A.J. (2013). The use of glutamine synthetase as a selection marker: recent advances in Chinese hamster ovary cell line generation processes. *Pharm. Bioprocess.* 1, 487–502.
- Ferreira, J.P., Peacock, R.W.S., Lawhorn, I.E.B., and Wang, C.L. (2011). Modulating ectopic gene expression levels by using retroviral vectors equipped with synthetic promoters. *Syst. Synth. Biol.* 5, 131–138.
- 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.
- Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol. Cell* 60, 256–267.

References

- Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W.A., Lowe, A., et al. (2018). A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res.* *46*, 4072–4086.
- 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.
- Gasanov, N.B., Toshchakov, S.V., Georgiev, P.G., and Maksimenko, O.G. (2015). The Use of Transcription Terminators to Generate Transgenic Lines of Chinese Hamster Ovary Cells (CHO) with Stable and High Level of Reporter Gene Expression. *Acta Naturae* *7*, 74–80.
- Geisse, S. (2009). Reflections on more than 10 years of TGE approaches. *Protein Expr. Purif.* *64*, 99–107.
- Glick, B.R. (1995). Metabolic load and heterologous gene expression. *Biotechnol. Adv.* *13*, 247–261.
- Goh, J.B., and Ng, S.K. (2018). Impact of host cell line choice on glycan profile. *Crit. Rev. Biotechnol.* *38*, 851–867.
- Gomez, B.P., Riggins, R.B., Shajahan, A.N., Klimach, U., Wang, A., Crawford, A.C., Zhu, Y., Zwart, A., Wang, M., and Clarke, R. (2007). Human X-box binding protein-1 confers both estrogen independence and antiestrogen resistance in breast cancer cell lines. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* *21*, 4013–4027.
- Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause Sites Promote Transcriptional Termination of Mammalian RNA Polymerase II. *Mol. Cell. Biol.* *26*, 3986–3996.
- 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*, 26.
- Guye, P., Li, Y., Wroblewska, L., Duportet, X., and Weiss, R. (2013). Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.* *41*, e156.
- Halleran, A.D., Swaminathan, A., and Murray, R.M. (2018). Single Day Construction of Multigene Circuits with 3G Assembly. *ACS Synth. Biol.* *7*, 1477–1480.
- Hasegawa, K., and Nakatsuji, N. (2002). Insulators prevent transcriptional interference between two promoters in a double gene construct for transgenesis. *FEBS Lett.* *520*, 47–52.
- Hesse, F., and Wagner, R. (2000). Developments and improvements in the manufacturing of human therapeutics with mammalian cell cultures. *Trends Biotechnol.* *18*, 173–180.
- Hornstein, B.D., Roman, D., Arévalo-Soliz, L.M., Engevik, M.A., and Zechiedrich, L. (2016). Effects of Circular DNA Length on Transfection Efficiency by Electroporation into HeLa Cells. *PLOS ONE* *11*, e0167537.

References

- 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.
- Hunter, M., Yuan, P., Vavilala, D., and Fox, M. (2019). Optimization of Protein Expression in Mammalian Cells. *Curr. Protoc. Protein Sci.* *95*, e77.
- Jäger, R., Bertrand, M.J.M., Gorman, A.M., Vandenabeele, P., and Samali, A. (2012). The unfolded protein response at the crossroads of cellular life and death during endoplasmic reticulum stress. *Biol. Cell* *104*, 259–270.
- Jasin, M., and Rothstein, R. (2013). Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harb. Perspect. Biol.* *5*, a012740.
- Jayapal, K.R., Wlaschin, K.F., Hu, W.-S., and Yap, M.G.S. (2007). Recombinant protein therapeutics from CHO cells - 20 years and counting. *Chem. Eng. Prog.* *103*, 40–47.
- Jenkins, N., Murphy, L., and Tyther, R. (2008). Post-translational Modifications of Recombinant Proteins: Significance for Biopharmaceuticals. *Mol. Biotechnol.* *39*, 113–118.
- Jewett, M.C., and Ellis, T. (2017). Editorial overview: Synthetic biology: Frontiers in synthetic biology. *Curr. Opin. Chem. Biol.* *40*, A1–A3.
- Jimenez, R.M., Polanco, J.A., and Lupták, A. (2015). Chemistry and biology of self-cleaving ribozymes. *Trends Biochem. Sci.* *40*, 648–661.
- 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.*
- 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.*
- Johnson, I.S. (1983). Human insulin from recombinant DNA technology. *Science* *219*, 632–637.
- Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* *339*, 225–229.
- Kao, F.T., and Puck, T.T. (1968). Genetics of somatic mammalian cells, VII. Induction and isolation of nutritional mutants in Chinese hamster cells. *Proc. Natl. Acad. Sci.* *60*, 1275–1281.
- Kemmer, C., Gitzinger, M., Daoud-El Baba, M., Djonov, V., Stelling, J., and Fussenegger, M. (2010). Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Nat. Biotechnol.* *28*, 355–360.
- Kesik-Brodacka, M. (2017). Progress in biopharmaceutical development. *Biotechnol. Appl. Biochem.*

- Khalil, A.S., and Collins, J.J. (2010). Synthetic biology: applications come of age. *Nat. Rev. Genet.* *11*, 367–379.
- Kim, M., O’Callaghan, P.M., Droms, K.A., and James, D.C. (2011). A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol. Bioeng.* *108*, 2434–2446.
- Krasowska, J., Olasek, M., Bzowska, A., Clark, P.L., and Wielgus-Kutrowska, B. (2010). The comparison of aggregation and folding of enhanced green fluorescent protein (EGFP) by spectroscopic studies.
- Kriz, A., Schmid, K., Baumgartner, N., Ziegler, U., Berger, I., Ballmer-Hofer, K., and Berger, P. (2010). A plasmid-based multigene expression system for mammalian cells. *Nat. Commun.* *1*, 120.
- Kühn, R., and Torres, R.M. (2002). Cre/loxP recombination system and gene targeting. *Methods Mol. Biol. Clifton NJ* *180*, 175–204.
- Lalonde, M.-E., and Durocher, Y. (2017). Therapeutic glycoprotein production in mammalian cells. *J. Biotechnol.* *251*, 128–140.
- Lanza, A.M., Kim, D.S., and Alper, H.S. (2013). Evaluating the influence of selection markers on obtaining selected pools and stable cell lines in human cells. *Biotechnol. J.* *8*, 811–821.
- Le Fourn, V., Girod, P.-A., Buceta, M., Regamey, A., and Mermoud, N. (2014). CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metab. Eng.* *21*, 91–102.
- Lee, G.M., Varma, A., and Palsson, B.O. (1991). Application of Population Balance Model to the Loss of Hybridoma Antibody Productivity. *Biotechnol. Prog.* *7*, 72–75.
- 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.* *113*, 2518–2523.
- Lee, J.S., Park, J.H., Ha, T.K., Samoudi, M., Lewis, N., Palsson, B.Ø., Kildegaard, H.F., and Lee, G.M. (2018). Revealing key determinants of clonal variation in transgene expression in recombinant CHO cells using targeted genome editing. *ACS Synth. Biol.*
- Lesueur, L.L., Mir, L.M., and André, F.M. (2016). Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro. *Mol. Ther. Nucleic Acids* *5*, e291.
- Levy, L., Anavy, L., Solomon, O., Cohen, R., Brunwasser-Meirom, M., Ohayon, S., Atar, O., Goldberg, S., Yakhini, Z., and Amit, R. (2017). A Synthetic Oligo Library and Sequencing Approach Reveals an Insulation Mechanism Encoded within Bacterial σ 54 Promoters. *Cell Rep.* *21*, 845–858.

References

- Lewis, C.J.T., Pan, T., and Kalsotra, A. (2017). RNA modifications and structures cooperate to guide RNA–protein interactions. *Nat. Rev. Mol. Cell Biol.* *18*, 202–210.
- Ley, D., Seresht, A.K., Engmark, M., Magdenoska, O., Nielsen, K.F., Kildegaard, H.F., and Andersen, M.R. (2015). Multi-omic profiling -of EPO-producing Chinese hamster ovary cell panel reveals metabolic adaptation to heterologous protein production. *Biotechnol. Bioeng.* *112*, 2373–2387.
- Li, F., Vijayasankaran, N., Shen, A.Y., Kiss, R., and Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. *MAbs* *2*, 466–479.
- Li, W., Li, W., Laishram, R.S., Hoque, M., Ji, Z., Tian, B., and Anderson, R.A. (2017). Distinct regulation of alternative polyadenylation and gene expression by nuclear poly(A) polymerases. *Nucleic Acids Res.* *45*, 8930–8942.
- Liao, W., Liu, B., Chang, C.-C., Lin, L.-J., Lin, C., Chen, B.-S., and Xie, Z. (2018). Functional Characterization of Insulation Effect for Synthetic Gene Circuits in Mammalian Cells. *ACS Synth. Biol.* *7*, 412–418.
- Libri, D. (2015). Endless Quarrels at the End of Genes. *Mol. Cell* *60*, 192–194.
- Lienert, F., Lohmueller, J.J., Garg, A., and Silver, P.A. (2014). Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat. Rev. Mol. Cell Biol.* *15*, 95–107.
- Liu, Y., Li, Q., Zheng, P., Zhang, Z., Liu, Y., Sun, C., Cao, G., Zhou, W., Wang, X., Zhang, D., et al. (2015). Developing a high-throughput screening method for threonine overproduction based on an artificial promoter. *Microb. Cell Factories* *14*, 121.
- Ma, J., and Wang, M.D. (2016). DNA supercoiling during transcription. *Biophys. Rev.* *8*, 75–87.
- Maehara, K., Hasegawa, T., Xiao, H., Takeuchi, A., Abe, R., and Isobe, K. (1999). Cooperative interaction of NF-kappaB and C/EBP binding sites is necessary for manganese superoxide dismutase gene transcription mediated by lipopolysaccharide and interferon-gamma. *FEBS Lett.* *449*, 115–119.
- Mahameed, M., and Tirosh, B. (2017). Engineering CHO cells with an oncogenic KIT improves cells growth, resilience to stress, and productivity. *Biotechnol. Bioeng.* *114*, 2560–2570.
- Mahr, R., Boeselager, R.F. von, Wiechert, J., and Frunzke, J. (2016). Screening of an *Escherichia coli* promoter library for a phenylalanine biosensor. *Appl. Microbiol. Biotechnol.* *100*, 6739–6753.
- Martínez, J.L., Liu, L., Petranovic, D., and Nielsen, J. (2012). Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Curr. Opin. Biotechnol.* *23*, 965–971.
- Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M.J., et al. (2019). mRNA structure regulates

protein expression through changes in functional half-life. *Proc. Natl. Acad. Sci.* *116*, 24075–24083.

Nissim, L., and Bar-Ziv, R.H. (2010). A tunable dual-promoter integrator for targeting of cancer cells. *Mol. Syst. Biol.* *6*, 444.

Nojima, T., Dienstbier, M., Murphy, S., Proudfoot, N.J., and Dye, M.J. (2013). Definition of RNA Polymerase II CoTC Terminator Elements in the Human Genome. *Cell Rep.* *3*, 1080–1092.

Nunes, N.M., Li, W., Tian, B., and Furger, A. (2010). A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO J.* *29*, 1523–1536.

Palmer, A.C., Egan, J.B., and Shearwin, K.E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription* *2*, 9–14.

Pandhal, J., Desai, P., Walpole, C., Doroudi, L., Malyshev, D., and Wright, P.C. (2012). Systematic metabolic engineering for improvement of glycosylation efficiency in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* *419*, 472–476.

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

Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., and Gersbach, C.A. (2013). Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods* *10*, 239–242.

Petersen, S.D., Zhang, J., Lee, J.S., Jakočiūnas, T., Grav, L.M., Kildegaard, H.F., Keasling, J.D., and Jensen, M.K. Modular 5'-UTR hexamers for context-independent tuning of protein expression in eukaryotes. *Nucleic Acids Res.*

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* *16*, 190–202.

Prashad, K., and Mehra, S. (2015). Dynamics of unfolded protein response in recombinant CHO cells. *Cytotechnology* *67*, 237–254.

Proudfoot, N. (1991). Poly(A) signals. *Cell* *64*, 671–674.

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

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* *352*, aad9926.

Puck, T.T., Cieciura, S.J., and Robinson, A. (1958). Genetics of Somatic Mammalian Cells: Iii. Long-Term Cultivation of Euploid Cells from Human and Animal Subjects. *J. Exp. Med.* *108*, 945–956.

References

- 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.
- Rebatchouk, D., Daraselina, N., and Narita, J.O. (1996). NOMAD: a versatile strategy for in vitro DNA manipulation applied to promoter analysis and vector design. *Proc. Natl. Acad. Sci.* *93*, 10891–10896.
- Ross, J. (1995). mRNA stability in mammalian cells. *Microbiol. Rev.* *59*, 423–450.
- Rössger, K., Charpin-El-Hamri, G., and Fussenegger, M. (2014). Bile acid-controlled transgene expression in mammalian cells and mice. *Metab. Eng.* *21*, 81–90.
- Ruder, W.C., Lu, T., and Collins, J.J. (2011). Synthetic Biology Moving into the Clinic. *Science* *333*, 1248–1252.
- Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A., and Orzaez, D. (2011). GoldenBraid: An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules. *PLOS ONE* *6*, e21622.
- 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.
- Schmidt, S.F., Larsen, B.D., Loft, A., and Mandrup, S. (2016). Cofactor squelching: Artifact or fact? *BioEssays* *38*, 618–626.
- Seghezzi, N., Amar, P., Koebmann, B., Jensen, P.R., and Virolle, M.-J. (2011). The construction of a library of synthetic promoters revealed some specific features of strong *Streptomyces* promoters. *Appl. Microbiol. Biotechnol.* *90*, 615–623.
- Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. *Science* *322*, 1849–1851.
- 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.
- Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. *Nat. Methods* *2*, 905–909.
- Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference – a crash course. *Trends Genet. TIG* *21*, 339–345.
- Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells Dayt. Ohio* *27*, 543–549.

- Spadiut, O., Capone, S., Krainer, F., Glieder, A., and Herwig, C. (2014). Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol.* *32*, 54–60.
- Stadlmayr, G., Mecklenbräuker, A., Rothmüller, M., Maurer, M., Sauer, M., Mattanovich, D., and Gasser, B. (2010). Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. *J. Biotechnol.* *150*, 519–529.
- Stinski, M.F., and Isomura, H. (2008). Role of the cytomegalovirus major immediate early enhancer in acute infection and reactivation from latency. *Med. Microbiol. Immunol. (Berl.)* *197*, 223–231.
- Tian, J., and Andreadis, S.T. (2009). Independent and high-level dual-gene expression in adult stem-progenitor cells from a single lentiviral vector. *Gene Ther.* *16*, 874–884.
- Tian, P., Wang, J., Shen, X., Rey, J.F., Yuan, Q., and Yan, Y. (2017). Fundamental CRISPR-Cas9 tools and current applications in microbial systems. *Synth. Syst. Biotechnol.* *2*, 219–225.
- Tigges, M., and Fussenegger, M. (2006). Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. *Metab. Eng.* *8*, 264–272.
- Torella, J.P., Lienert, F., Boehm, C.R., Chen, J.-H., Way, J.C., and Silver, P.A. (2014a). Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications. *Nat. Protoc.* *9*, 2075–2089.
- Torella, J.P., Boehm, C.R., Lienert, F., Chen, J.-H., Way, J.C., and Silver, P.A. (2014b). Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acids Res.* *42*, 681–689.
- Uchida, N., Hanawa, H., Yamamoto, M., and Shimada, T. (2013). The Chicken Hypersensitivity Site 4 Core Insulator Blocks Promoter Interference in Lentiviral Vectors. *Hum. Gene Ther. Methods* *24*, 117–124.
- User Guide - Gateway® Technology Gateway® Technology.
- Viotti, M., Nowotschin, S., and Hadjantonakis, A.-K. (2011). Afp::mCherry, a red fluorescent transgenic reporter of the mouse visceral endoderm. *Genes. N. Y. N* *2000* *49*, 124–133.
- Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* *15*, 69–81.
- Walter, P., and Ron, D. (2011). The unfolded protein response: from stress pathway to homeostatic regulation. *Science* *334*, 1081–1086.
- Wang, X., Li, P., and Gutenkunst, R.N. (2017). Systematic effects of mRNA secondary structure on gene expression and molecular function in budding yeast. *BioRxiv* 138792.

References

- Wang, Y., Xing, P., Cui, W., Wang, W., Cui, Y., Ying, G., Wang, X., and Li, B. (2015). Acute Endoplasmic Reticulum Stress-Independent Unconventional Splicing of XBP1 mRNA in the Nucleus of Mammalian Cells. *Int. J. Mol. Sci.* *16*, 13302–13321.
- Weber, W., and Fussenegger, M. (2012). Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet.* *13*, 21–35.
- Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PloS One* *6*, e16765.
- Wei, P., Yang, Y., Li, T., Ding, Q., and Sun, H. (2015). A engineered Bifidobacterium longum secreting a bioactive penetratin-Glucagon-like peptide 1 fusion protein enhances Glucagon-like peptide 1 absorption in the intestine. *J. Microbiol. Biotechnol.*
- Wei, W., Pelechano, V., Järvelin, A.I., and Steinmetz, L.M. (2011). Functional consequences of bidirectional promoters. *Trends Genet. TIG* *27*, 267–276.
- West, N.R. (2014). Development of a Tunable Mammalian Protein Expression System and an Investigation of Promoter Interference in Three Promoters Often Utilized in the Production of Biopharmaceuticals.
- West, S., and Proudfoot, N.J. (2009). Transcriptional Termination Enhances Protein Expression in Human Cells. *Mol. Cell* *33*, 354–364.
- West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. *Genes Dev.* *16*, 271–288.
- West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* *432*, 522–525.
- West, S., Zaret, K., and Proudfoot, N.J. (2006). Transcriptional Termination Sequences in the Mouse Serum Albumin Gene. *RNA* *12*, 655–665.
- West, S., Proudfoot, N.J., and Dye, M.J. (2008). Molecular Dissection of Mammalian RNA Polymerase II Transcriptional Termination. *Mol. Cell* *29*, 600–610.
- White, E., Kamieniarz-Gdula, K., Dye, M.J., and Proudfoot, N.J. (2013). AT-rich sequence elements promote nascent transcript cleavage leading to RNA polymerase II termination. *Nucleic Acids Res.* *41*, 1797–1806.
- Wong, D., Teixeira, A., Oikonomopoulos, S., Humburg, P., Lone, I.N., Saliba, D., Siggers, T., Bulyk, M., Angelov, D., Dimitrov, S., et al. (2011). Extensive characterization of NF-κB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. *Genome Biol.* *12*, R70.
- Wurm, F.M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* *22*, 1393–1398.

References

- Yahata, K., Maeshima, K., Sone, T., Ando, T., Okabe, M., Imamoto, N., and Imamoto, F. (2007). cHS4 insulator-mediated alleviation of promoter interference during cell-based expression of tandemly associated transgenes. *J. Mol. Biol.* *374*, 580–590.
- Ye, H., Baba, M.D.-E., Peng, R.-W., and Fussenegger, M. (2011). A Synthetic Optogenetic Transcription Device Enhances Blood-Glucose Homeostasis in Mice. *Science* *332*, 1565–1568.
- Yim, S.S., An, S.J., Kang, M., Lee, J., and Jeong, K.J. (2013). Isolation of fully synthetic promoters for high-level gene expression in *Corynebacterium glutamicum*. *Biotechnol. Bioeng.* *110*, 2959–2969.
- Yusufi, F.N.K., Lakshmanan, M., Ho, Y.S., Loo, B.L.W., Ariyaratne, P., Yang, Y., Ng, S.K., Tan, T.R.M., Yeo, H.C., Lim, H.L., et al. (2017). Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. *Cell Syst.* *4*, 530-542.e6.
- Yusufzai, T.M., and Felsenfeld, G. (2004). The 5'-HS4 chicken β -globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc. Natl. Acad. Sci.* *101*, 8620–8624.
- Zeng, L., Zampetaki, A., Margariti, A., Pepe, A.E., Alam, S., Martin, D., Xiao, Q., Wang, W., Jin, Z.-G., Cockerill, G., et al. (2009). Sustained activation of XBP1 splicing leads to endothelial apoptosis and atherosclerosis development in response to disturbed flow. *Proc. Natl. Acad. Sci.* *106*, 8326–8331.
- Zhang, H., Rigo, F., and Martinson, H.G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Mol. Cell* *59*, 437–448.
- Zhu, J. (2012). Mammalian cell protein expression for biopharmaceutical production. *Biotechnol. Adv.* *30*, 1158–1170.
- Zimmerman, E.S., Heibeck, T.H., Gill, A., Li, X., Murray, C.J., Madlansacay, M.R., Tran, C., Uter, N.T., Yin, G., Rivers, P.J., et al. (2014). Production of Site-Specific Antibody–Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System. *Bioconjug. Chem.* *25*, 351–361.
- Zucchelli, S., Patrucco, L., Persichetti, F., Gustincich, S., and Cotella, D. (2016). Engineering Translation in Mammalian Cell Factories to Increase Protein Yield: The Unexpected Use of Long Non-Coding SINEUP RNAs. *Comput. Struct. Biotechnol. J.* *14*, 404–410.

Appendix

Appendix A

Table 6-1: Cost and service analysis of de novo synthesis of transcription units by third party suppliers.

A table summarising the cost, time and service provision of various *de novo* gene synthesis providers within the industry for the construction of individual TUs or multiple TUs combined together to form a polycistronic cassette of various sizes.

Provider	Type of Synthesis	Cost (+VAT)	Final Cost	Duration of Synthesis	Sequence Check post-Synthesis
Genscript	Polycistronic - 3738 bp	\$1495.20/£984.56	\$1345.68/£886.00 (Not including shipping)	24-28 business days	Yes
	Multiple single TUs – Sequence 1: 1227bp; Sequence 2: 1323bp; Sequence 3: 1188bp	\$1307.55/£860.80	\$1177.46/£775.25 (Not including shipping)	20-24 business days	
DNA 2.0	Polycistronic - 15kb	£12300 - 14850 (£0.82 to £0.99 per base pair)	-	2-3 months	Yes
GeneArt	Polycistronic - 3738 bp	£1,150.13	-	36 business days	Yes
	Multiple single TUs – Sequence 1: 1227bp; Sequence 2: 1323bp; Sequence 3: 1188bp	£917.69	-	19 business days	
Liverpool GeneMill	Polycistronic- 15kb	£0.35 per base pair - £5250	-	35 business days	Yes
	Polycistronic - 3738 bp	£0.27 per base pair - £1009.26	£1,079.26	25 - 30 business days	
	Multiple single TUs – Sequence 1: 1227bp; Sequence 2: 1323bp; Sequence 3: 1188bp	£0.25 per base pair - £934.50	£1,004.50		

Table 6-2: Coding DNA sequence for *eGFP*, *mCherry* and *tagBFP*.

The table shows the nucleotide sequence for *eGFP*, *mCherry* and *tagBFP* which was *de novo* synthesised and codon optimised for *Cricetulus griseus* by GeneArt. These CDSes were used for all fluorescent protein expression work performed in the PhD.

Fluorescent Protein	Nucleotide Sequence (5' – 3')
<i>eGFP</i>	ATGGTGTCCAAGGGCGAGGAACTGTTCACCGGCGTGGTGCCCATCCTGG TGGAAC TGGACGGCGACGTGAACGGCCACAAGTTCTCCGTGTCTGGCGA GGGCGAAGGGCGACGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC ACCACCGGCAAGCTGCCCGTGGCCTACCCTCGTGACAACCCTGA CCTACGGCGTGCAGTGCTTCTCCAGATACCCCGACCACATGAAGCAGCA CGATTTCTTCAAGTCCGCCATGCCCGAGGGCTACGTGCAGGAACGGACC ATCTTCTTCAAGGACGACGGCAACTACAAGACCAGAGCCGAAGTGAAGT TCGAGGGCGACACCCTGGTGAACCGGATCGAGCTGAAGGGCATCGAATT CAAAGAGGATGGCAACATCCTGGGCCACAAGCTGGAGTACAACACTACAAC TCCCACAATGTGTATATCATGGCCGACAAGCAGAAAAACGGCATCAAAG TGAACTTCAAGATCCGGCACAACATCGAGGACGGCTCCGTGCAGCTGGC CGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTG CCTGACAACCACTACCTGTCCACCCAGTCCGCCCTGTCCAAGGACCCCAA CGAGAAGCGGGACCACATGGTGTGCTGGAATTTGTGACCGCCGCTGGC ATCACCTGGGCATGGACGAGCTGTATAAGTGA
<i>mCherry</i>	ATGGTGTCCAAGGGCGAAGAGGACAACATGGCCATCATCAAAGAGTTCA TGCGGTTCAAGGTGCACATGGAAGGCTCCGTGAACGGCCACGAGTTCGA GATCGAGGGCGAGGGCGAAGGCAGACCCTACGAGGGCACCCAGACCGC CAAGCTGAAAGTGACCAAGGGCGGACCCCTGCCCTTCGCCTGGGACATC CTGTCCCCCAGTTTATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGC CGACATCCCCGACTACCTGAAGCTGAGCTTCCCCGAGGGCTTCAAGTGG GAGAGAGTGATGAACTTCGAGGACGGCGGCGTGGTGACAGTGACCCAG GACAGTCTCTGCAGGACGGCGAGTTCATCTACAAAGTGAAGCTGCGGG GCACCAACTTCCCTTCCGACGGCCCCGTGATGCAGAAAAAGACAATGGG CTGGGAGGCCTCCTCCGAGCGGATGTACCCTGAGGATGGCGCCCTGAAG GGCGAGATCAAGCAGCGGCTGAAGCTGAAGGATGGCGGCCACTACGAC GCCGAAGTGA AAAACCACTACAAGGCCAAGAAACCCGTGCAGCTGCCTG GCGCCTACAACGTGAACATCAAGCTGGACATCACCTCCCACAACGAGGA CTACACCATCGTGGAACAGTACGAGCGGGCCGAGGGCAGACACTCCACC GGCGGCATGGACGAGCTGTACAAGTGA
<i>tagBFP</i>	ATGTCCGAGCTGATCAAAGAAAACATGCACATGAAGCTGTATATGGAAG GCACCGTGGACAACCACCACTTCAAGTGCACCTCCGAGGGCGAGGGCAA GCCCTACGAGGGCACCCAGACCATGCGGATCAAGGTGGTGAAGGCGG CCCTCTGCCCTTCGCCTTCGATATCCTGGCCACCTCCTTTCTGTACGGCTC CAAGACCTTCATCAACCACACCCAGGGCATCCCCGATTTCTTCAAGCAG AGCTTCCCCGAGGGCTTACCTGGGAGAGAGTGACCACCTACGAGGACG GCGGCGTGCTGACCGCTACCCAGGACACCTCTCTGCAGGACGGCTGCCT GATCTACAACGTGAAGATCCGGGGCGTGAACCTCACCTCCAACGGCCCC GTGATGCAGAAGAAAACCTGGGCTGGGAGGCCTTACCAGGACACTGT ACCCTGCCGACGGCGGCCTGGAAGGCAGAAACGACATGGCCCTGAAGCT CGTGGGAGGCAGCCACCTGATCGCCAACATCAAGACCACCTACAGATCC AAGAAGCCCCGAAGAACCTGAAGATGCCCGGCGTGTACTACGTGGACT ACCGGCTGGAACGGATCAAAGAGGCCAACAACGAGACATACGTGGAAC AGCACGAGGTGGCCGTGGCCCCGTACTGCGACCTGCCTTCAAAGCTGGG CCACAAGCTGAACTGA

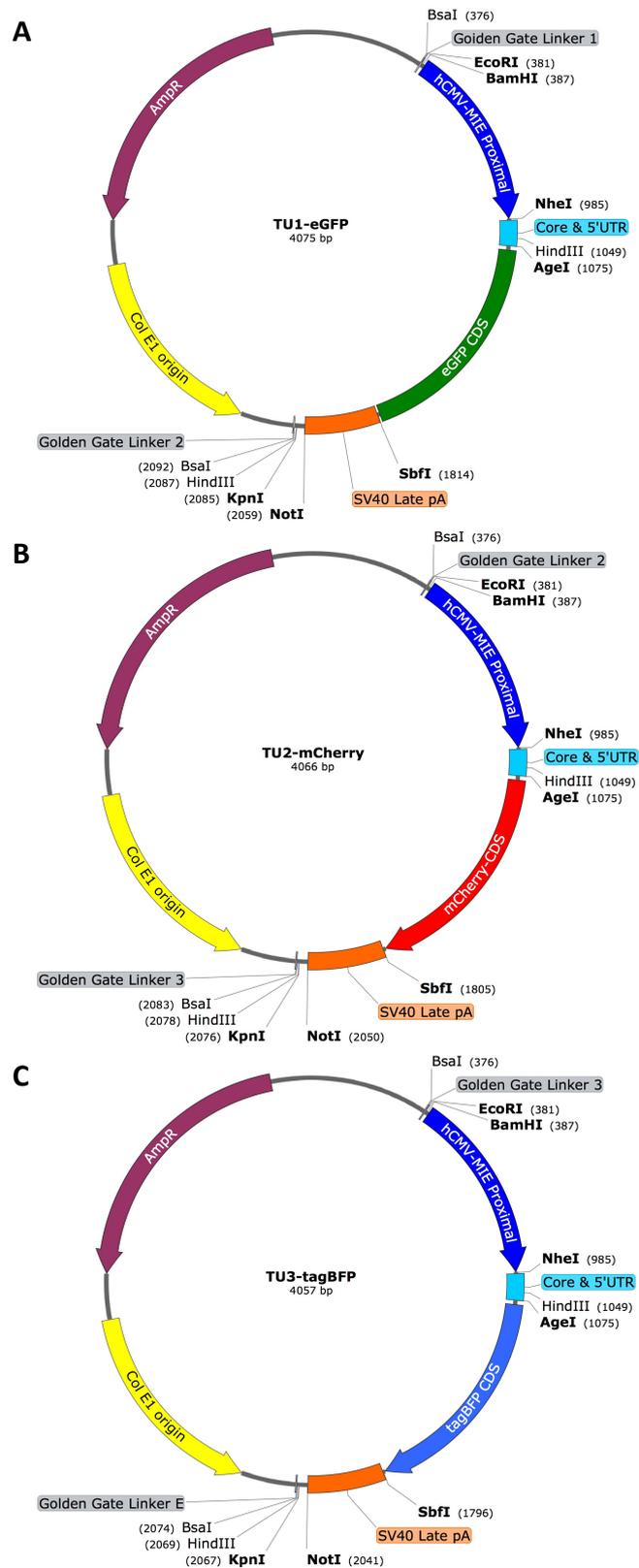


Figure A-1: Vector maps showing the transcription units encoding for *eGFP*, *mCherry* and *tagBFP*.

TU1-eGFP (A), TU2-mCherry (B) and TU3-tagBFP (C) represents the TU encoding for *eGFP*, *mCherry* and *tagBFP* in position 1, 2 and 3 of the polycistronic cassette respectively. The vector maps show the annotations of the individual transcription units *de novo* synthesised and housed within an Ampicillin resistant vector for amplification. These plasmids were the inserts combined with the pExp-Vec-GG to construct the MGEVs by golden gate cloning.

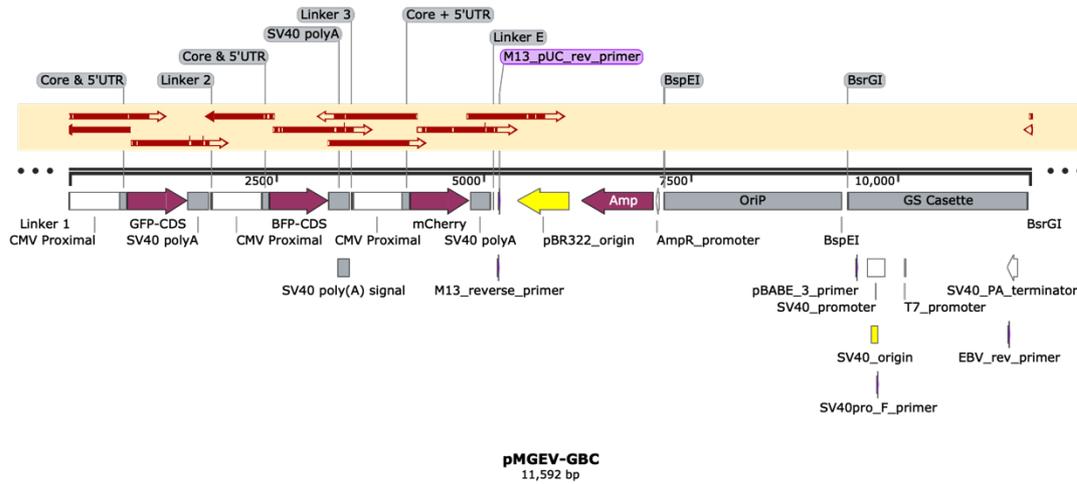


Figure A-2: Map showing successful DNA sequencing of pMGEV-GBC.

The map summarising the layout and elements within the pMGEV-GBC plasmid. The plasmid was sequence verified using 9 different primers to obtain coverage over the polycistronic cassette comprising of the three TUs encoding *eGFP*, *tagBFP* and *mCherry* respectively. The red arrows above the polycistronic cassette highlight successful alignment of each sequence contig (generated from the 9 different primers) against the pMGEV-GBC reference sequence. This highlighted successful assembly of the MGEV with the correct order of genes within the polycistronic cassette.

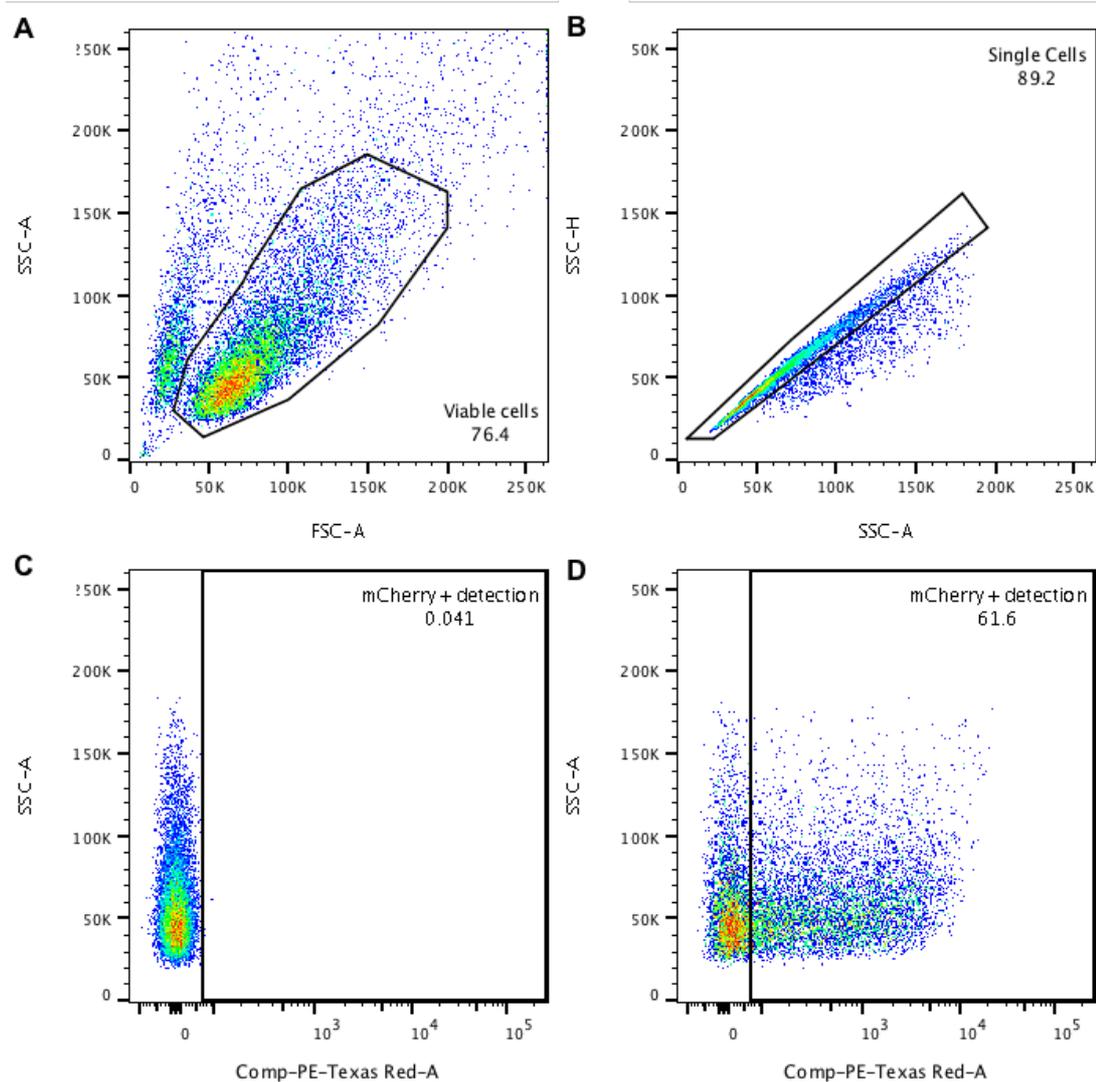


Figure A-3: Gating strategy for singular expression *mCherry* using the BD LSR Fortessa flow cytometer.

Figure shows the method of gating used to isolate and analyse the transfected cell population which is expressing a single fluorescent protein. This example demonstrates identification of transfectants successfully expressing *mCherry*. The population was identified by gating for viable cells (A), followed by gating for single cells/event and exclusion of doublets (B) and finally distinguishing between autofluorescence of the parental cell and positive detection of fluorescent protein using a mock transfected (C) and successful transfectants expressing the fluorescent protein (D).

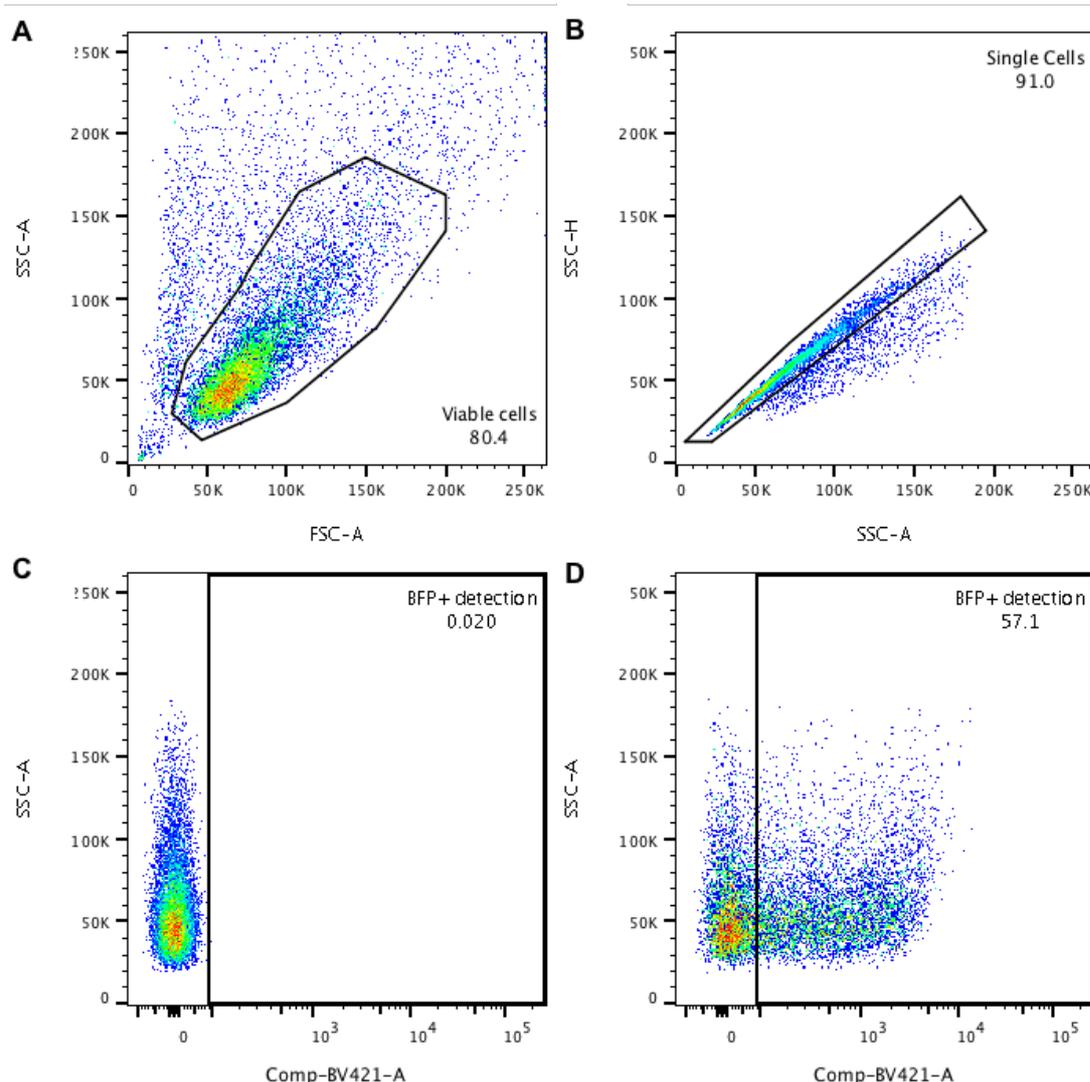


Figure A-4: Gating strategy for singular expression *tagBFP* using the BD LSR Fortessa flow cytometer.

Figure shows the method of gating used to isolate and analyse the transfected cell population which is expressing a single fluorescent protein. This example demonstrates identification of transfectants successfully expressing *tagBFP*. The population was identified by gating for viable cells (A), followed by gating for single cells/event and exclusion of doublets (B) and finally distinguishing between autofluorescence of the parental cell and positive detection of fluorescent protein using a mock transfected (C) and successful transfectants expressing the fluorescent protein (D).

Table 6-3: Ordinary one-way ANOVA – Viable Cell Concentration.

Table summarising the results of an ordinary one-way ANOVA using a Dunnett's multiple comparisons statistical test. This is to compare viable cell concentration between the mock and different MGEV transfection conditions for statistical significance to analyse optimal transfection conditions.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant ?	Summary	Adjusted P Value
Mock vs. MGEV-GBC (400 ng)	0.085	-0.2403 to 0.4103	No	ns	0.9285
Mock vs. MGEV-GBC (600 ng)	0.245	-0.08029 to 0.5703	No	ns	0.1757
Mock vs. MGEV-GBC (800 ng)	0.355	0.02971 to 0.6803	Yes	*	0.0307
Mock vs. MGEV-CBG (400 ng)	0.225	-0.1003 to 0.5503	No	ns	0.2346
Mock vs. MGEV-CBG (600 ng)	0.128	-0.197 to 0.4536	No	ns	0.7182
Mock vs. MGEV-CBG (800 ng)	0.265	-0.06029 to 0.5903	No	ns	0.1301

Table 6-4: Ordinary one-way ANOVA – Viability.

Table summarising the results of an ordinary one-way ANOVA using a Dunnett's multiple comparisons statistical test. This is to compare cell viability between the mock and different MGEV transfection conditions for statistical significance to analyse optimal transfection conditions.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant ?	Summary	Adjusted P Value
Mock vs. MGEV-GBC (400 ng)	15.87	4.54 to 27.19	Yes	**	0.0058
Mock vs. MGEV-GBC (600 ng)	26.9	15.57 to 38.23	Yes	****	0.0001
Mock vs. MGEV-GBC (800 ng)	36.77	25.44 to 48.09	Yes	****	0.0001
Mock vs. MGEV-CBG (400 ng)	13.1	1.774 to 24.43	Yes	*	0.0216
Mock vs. MGEV-CBG (600 ng)	8.967	-2.36 to 20.29	No	ns	0.1458
Mock vs. MGEV-CBG (800 ng)	15.4	4.074 to 26.73	Yes	**	0.0072

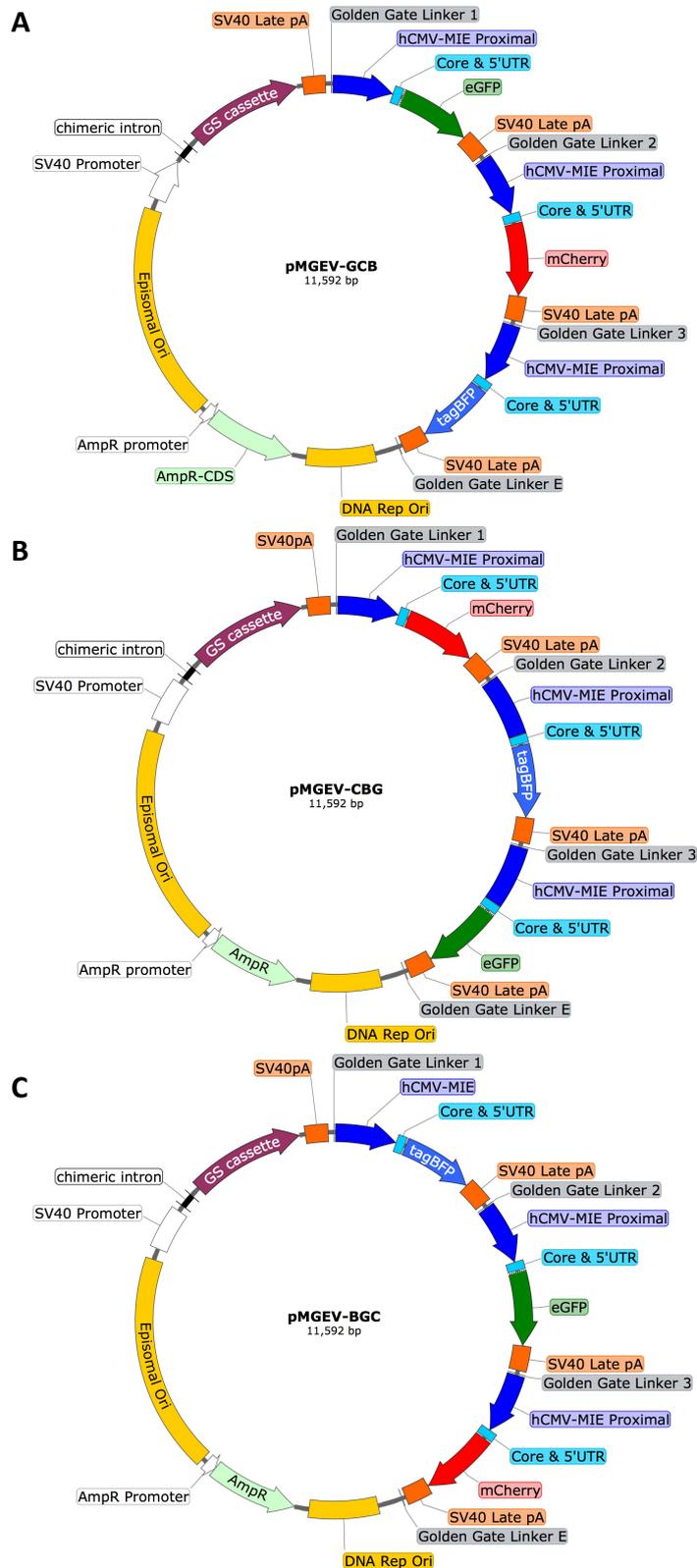


Figure A-5: Vector maps showing multi-gene expression vector variants constructed.
A- pMGEV-GCB shows the polycistronic cassette encoding for *eGFP*, *mCherry* and *tagBFP* in position 1,2 and 3 respectively, **B-** pMGEV-CBG shows the polycistronic cassette encoding for *mCherry*, *tagBFP* and *eGFP* in position 1,2 and 3 respectively, **C-** pMGEV-BGC shows the polycistronic cassette encoding for *tagBFP*, *eGFP* and *mCherry* in position 1,2 and 3 respectively. The vector maps show the annotations and all additional genetic elements within the different MGEVs constructed and tested by transient expression to characterise gene expression dynamics within a MGEV.

Appendix B**Table 6-5: List of primers used for PCR modification of synthetic promoter library.**

The table summarised the list of primers, the synthetic promoters they target, the primer sequence, melting temperature and GC content.

Synthetic Promoter	Primer Name	Primer Sequence (5' – 3')	Melting Temp. (°C)	GC Content (%)
5RPU.1	5RPU.1_FW	CGTGGATCCTATAGGAAG GTCTTACCGGA	61.7	51.7
	5RPU.1_RV	CGCATAGCTAGCCATACT AACTTCCGGTAATAG	60.6	45.5
5RPU.2	5RPU.2_FW	CGAGGATCCAGGGTCCT GATATGGGG	62.0	57.7
	5RPU.2_RV	CGCATAGCTAGCTGATCA AATTGCGCAAAAAA	61.6	40.6
10RPU.1	10RPU.1_FW	CGCGGATCCTTACCGGA AGTTAAGGTC	62.3	55.6
	10RPU.1_RV	CGCATAGCTAGCAACTTC CGGTAATGATC	60.6	48.3
	10RPU.1_FW Long	CGCGGATCCTTACCGGA AGTTAAGGTCTTTTG	63.1	50.0
	10RPU.1_RV Long	CGCATAGCTAGCAACTTC CGGTAATGATCAAATTG	61.8	42.9
10RPU.2	10RPU.2_FW	CGTGGATCCTGGGGCGG GGAGTATACGA	67.7	64.3
	10RPU.2_RV	CGCATTGCTAGCTCCCCG CCCCACTAAG	67.7	64.3
	10RPU.2_FW Long	CGTGGATCCTGGGGCGG GGAGTATACGACCTATT	68.1	58.8
	10RPU.2_RV Long	CGCATTGCTAGCTCCCCG CCCCACTAAGGCATACTA	69.4	58.3
20RPU.1	20RPU.1_FW	CGCGGATCCTTACCGGA AGTTGACC	63.2	60.0
	20RPU.1_RV	CGCTTAGCTAGCATTGCT GTGTCATCC	61.4	51.9
20RPU.2	20RPU.2_FW	CGTGGATCCTGGGGCGG GGACAGATT	67.9	65.4
	20RPU.2_RV	CCGATTGCTAGCTCCCCG CCCCATAGGTC	68.3	65.5
40RPU.1	40RPU.1_FW	CGCGGATCCTTACCGGA AGTTCCTTAG	62.1	55.6
	40RPU.1_RV	CGCATAGCTAGCAACTTC CGGTAACTAAG	60.3	48.3
40RPU.2	40RPU.2_FW	CGTGGATCCTTTTGC GCAATTTATAGGT	59.5	42.9
	40RPU.2_RV	CCGATAGCTAGCATTGCT GTGTCATGTTA	60.2	44.8

Appendix

Synthetic Promoter	Primer Name	Primer Sequence (5' – 3')	Melting Temp. (°C)	GC Content (%)
60RPU.1	60RPU.1_FW	CGCGGATCCTTACCGG AAGTTCTAGAC	61.8	55.6
	60RPU.1_RV	CGCTTAGCTAGCATTG CTGTGTCATCAT	60.8	46.4
60RPU.2	60RPU.2_FW	CGTGGATCCTGGGGCG GGGAGTTAGATG	67.0	64.3
	60RPU.2_RV	CGCATAGCTAGCTGGA AAGTCCCATGATCAA CTCC	64.8	48.6
80RPU.1	80RPU.1_FW	CGTGGATCCTGGGGCG GGGAAGTATGATGA	67.3	60.0
	80RPU.1_RV	CGCTTAGCTAGCATTG CTGTGTCATAATCTGT GGAAAG	63.6	44.7
80RPU.2	80RPU.2_FW	CGTGGATCCTGGGACT TTCCAAGTATGTG	62.0	51.7
	80RPU.2_RV	CTCATAGCTAGCATTG CTGTGTCATGTCTAG	59.9	45.2
100RPU.1	100RPU.1_FW	CGCGGATCCTGGGACT TTCCACCTTAGAT	64.5	55.2
	100RPU.1_RV	CGCATAGCTAGCTGGA AAGTCCCATGATCA	62.9	50.0
	100RPU.1_FW_ Long	CGCGGATCCTGGGACT TTCCACCTTAGATGAC AC	66.3	55.9
	100RPU.1_RV_ Long	CGCATAGCTAGCTGGA AAGTCCCATGATCATC TTCTCA	64.9	47.4
100RPU.2	100RPU.2_FW	CGCGGATCCTGGGACT TTCCACTAGACATGA	65.3	54.8
	100RPU.2_RV	CGCATAGCTAGCTGGA AAGTCCCAGACCT	64.5	55.2
	100RPU.2_FW_ Long	CGCGGATCCTGGGACT TTCCACTAGACATGAC AC	66.3	55.9
	100RPU.2_RV_ Long	CGCATAGCTAGCTGGA AAGTCCCAGACCTTTC TTCT	65.5	50.0

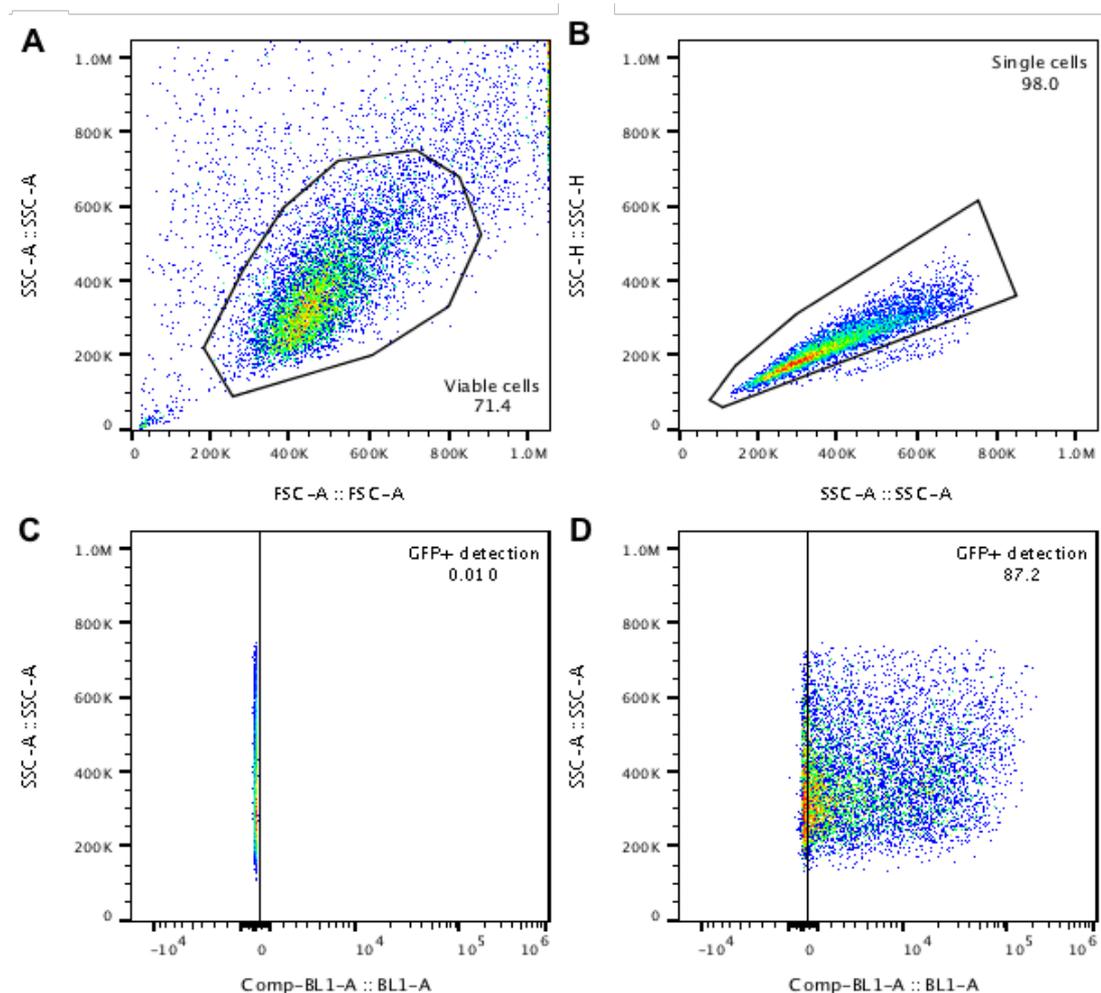


Figure B-1: Gating strategy for singular expression of *eGFP* using the Attune NxT flow cytometer. Figure shows the method of gating used to isolate and analyse the transfected cell population which is expressing a single fluorescent protein. This example demonstrates identification of transfectants successfully expressing *eGFP*. The population was identified by gating for viable cells (A), followed by gating for single cells/event and exclusion of doublets (B), and finally exclusion of cellular autofluorescence within the BL1 channel (C), and successful expression of *eGFP* within transfected cells (D).

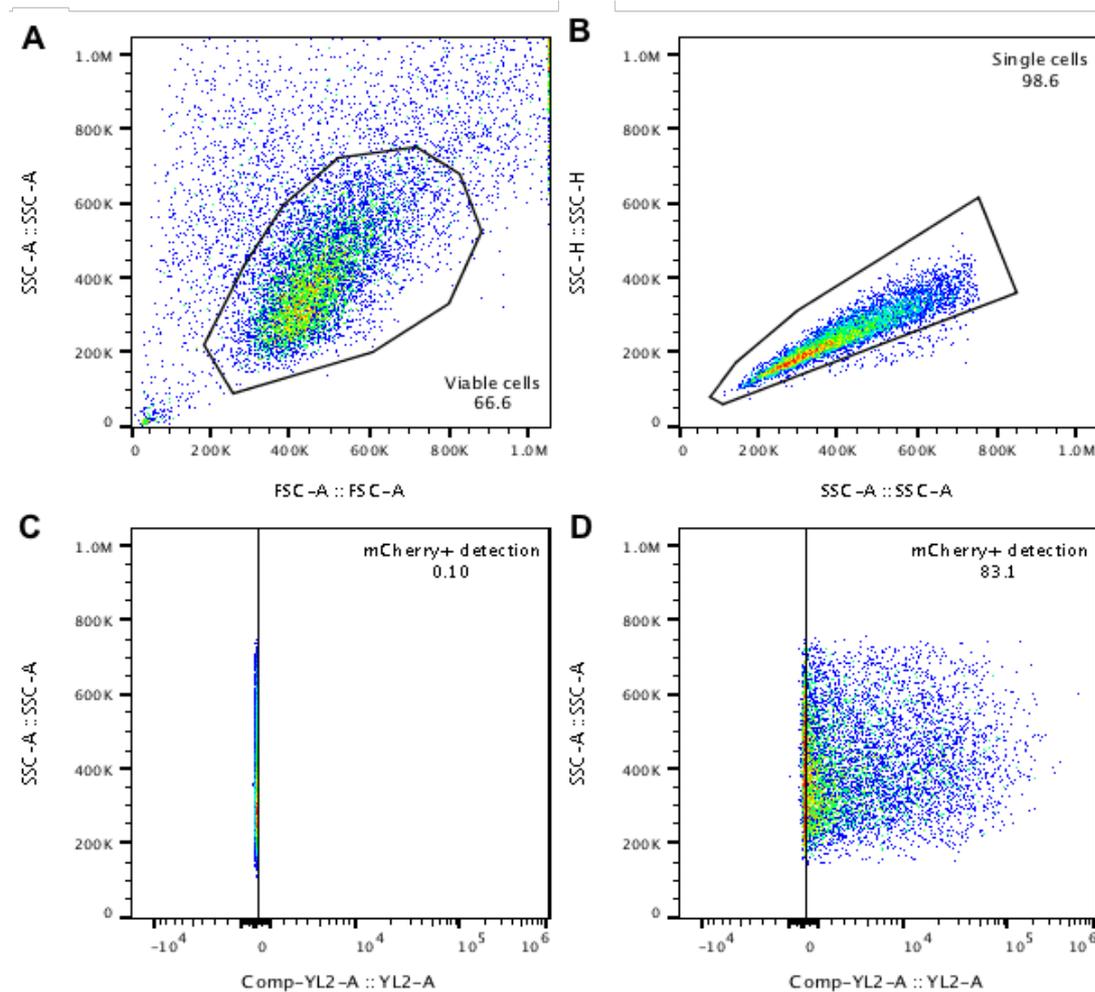


Figure B-2: Gating strategy for singular expression of *mCherry* using the Attune NxT flow cytometer.

Figure shows the method of gating used to isolate and analyse the transfected cell population which is expressing a single fluorescent protein. This example demonstrates identification of transfectants successfully expressing *mCherry*. The population was identified by gating for viable cells (A), followed by gating for single cells/event and exclusion of doublets (B), and finally exclusion of cellular autofluorescence within the YL2 channel (C), and successful expression of *mCherry* within transfected cells (D).

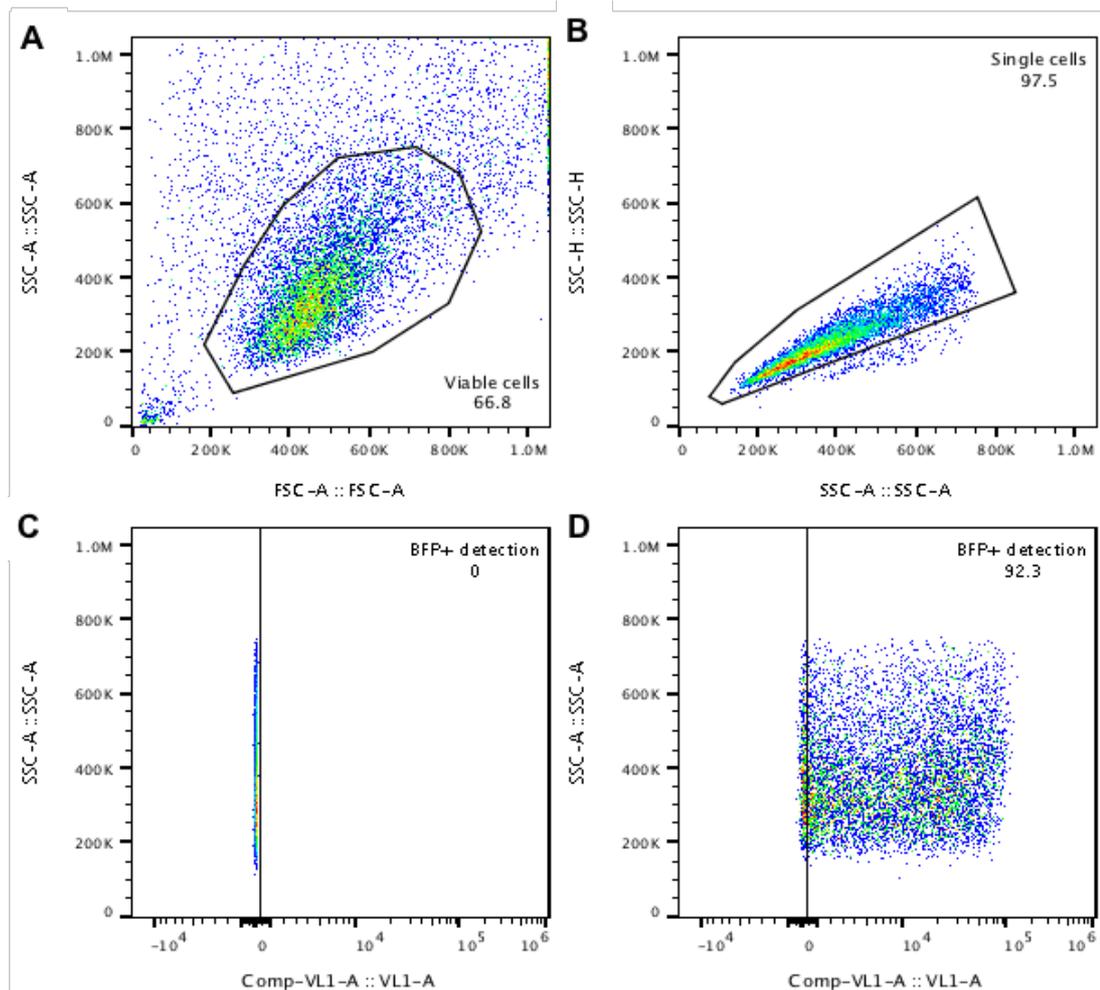


Figure B-3: Gating strategy for singular expression of *tagBFP* using the Attune NxT flow cytometer.

Figure shows the method of gating used to isolate and analyse the transfected cell population which is expressing a single fluorescent protein. This example demonstrates identification of transfectants successfully expressing *tagBFP*. The population was identified by gating for viable cells (A), followed by gating for single cells/event and exclusion of doublets (B), and finally exclusion of cellular autofluorescence within the VL1 channel (C), and successful expression of *tagBFP* within transfected cells (D).

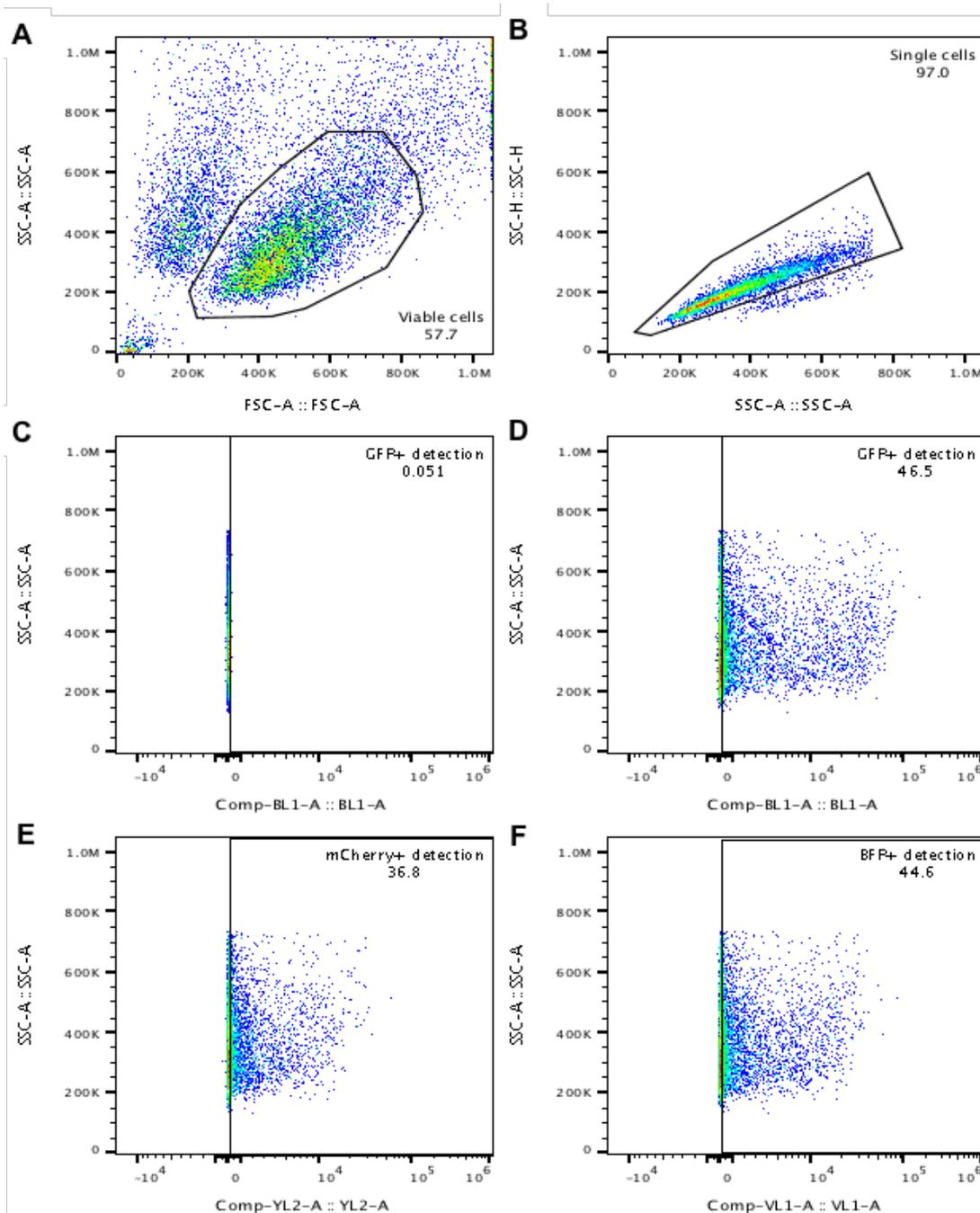


Figure B-4: Gating strategy for expression of multiple fluorescent using the Attune NxT flow cytometer.

Figure shows the method of gating used to isolate and analyse the transfected cell population which is expressing a multiple fluorescent protein. The population is identified by gating for viable cells (A), followed by gating for single cells/event and exclusion of doublets (B), exclusion of cellular autofluorescence using mock transfected cells within the BL1 channel and same was performed for the VL1 and YL2 channels (C), and finally successful expression of *eGFP* (D), *mCherry* (E) and *tagBFP* (F) within transfected cells respectively.

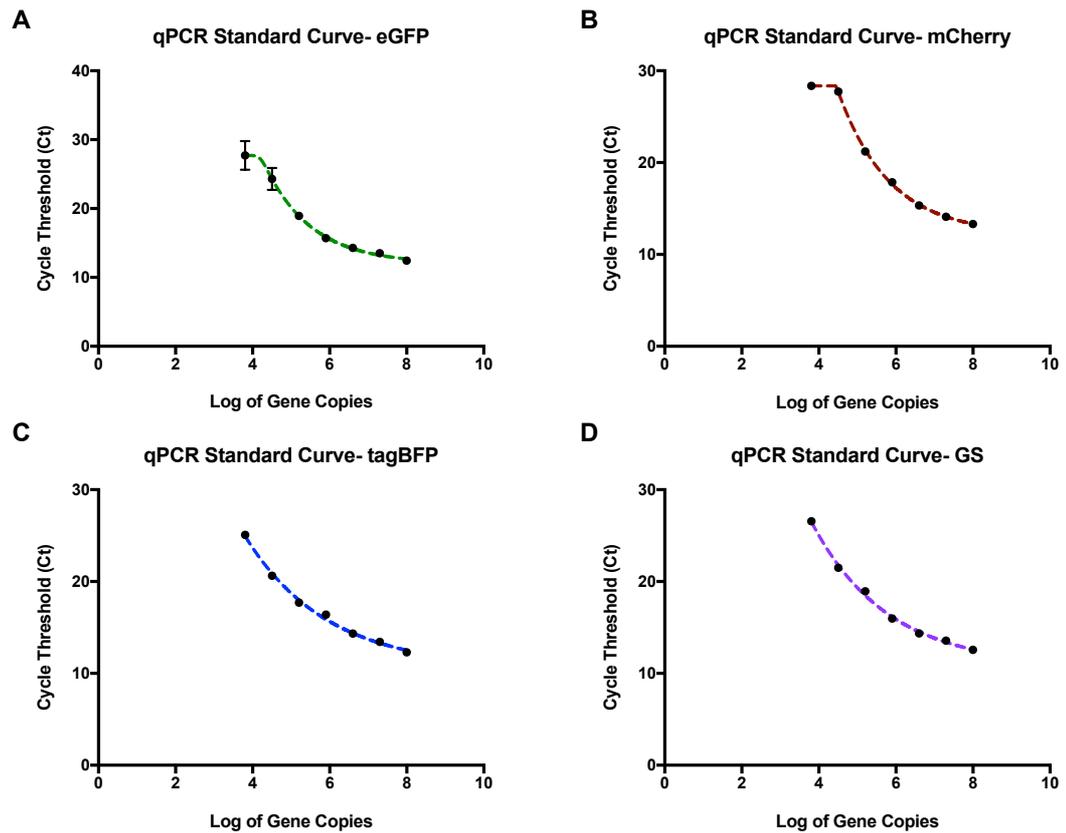


Figure B-5: qPCR Standard curves for *eGFP*, *mCherry*, *tagBFP* and glutamine synthetase for mRNA copy quantification.

eGFP (A), *mCherry* (B), *tagBFP* (C) and *GS* (D) standard curves derived by testing 10×10^8 to 6.4×10^3 gene copies (as log of gene copies on the x axis) of linearised DNA template by qPCR and measuring the Ct values (y axis). An asymmetrical sigmoidal relationship was observed as shown by the trendlines in each plot.

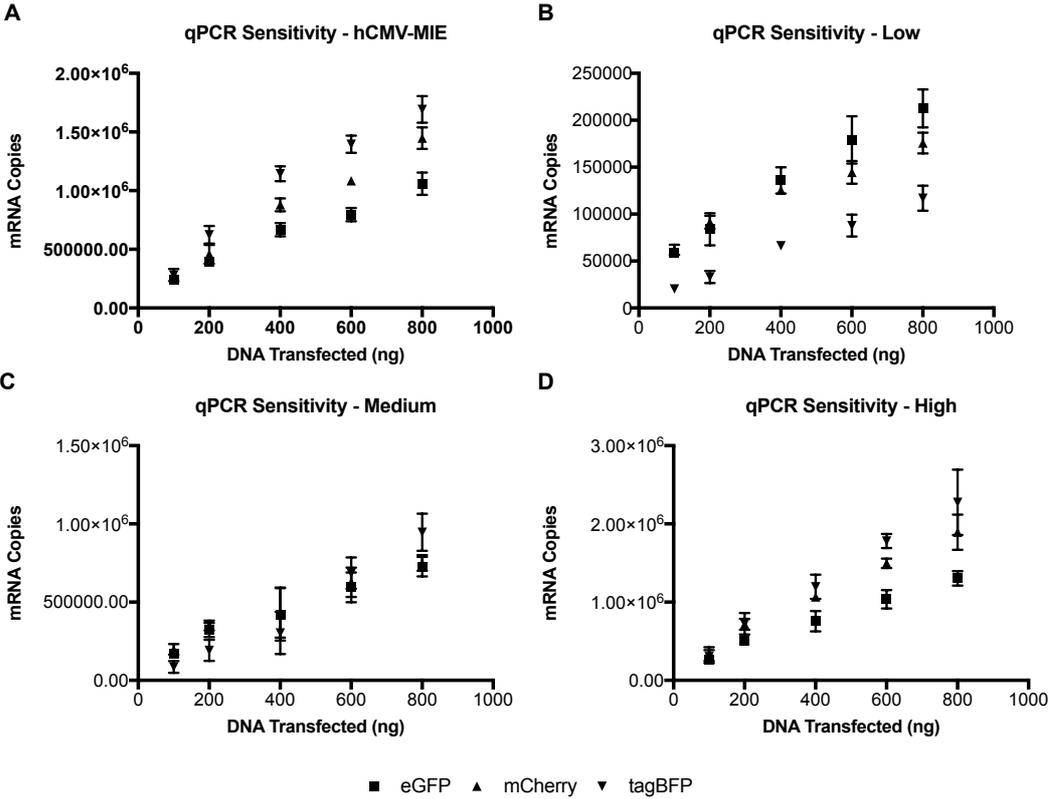


Figure B-6: Measuring dynamic range of qPCR by quantified mRNA copies. The sensitivity and dynamic range of qPCR detection was measured by titrating 100 to 800 ng of total plasmid DNA which comprised of equimolar *eGFP*, *mCherry* and *tagBFP* plasmids driven by either a hCMV-MIE (A) promoter, low (B), medium (C) and high (D) strength synthetic promoter. The absolute mRNA copies were plotted against the amount of DNA transfected to demonstrate the relationship. The error bars represent the SEM of biological triplicate.

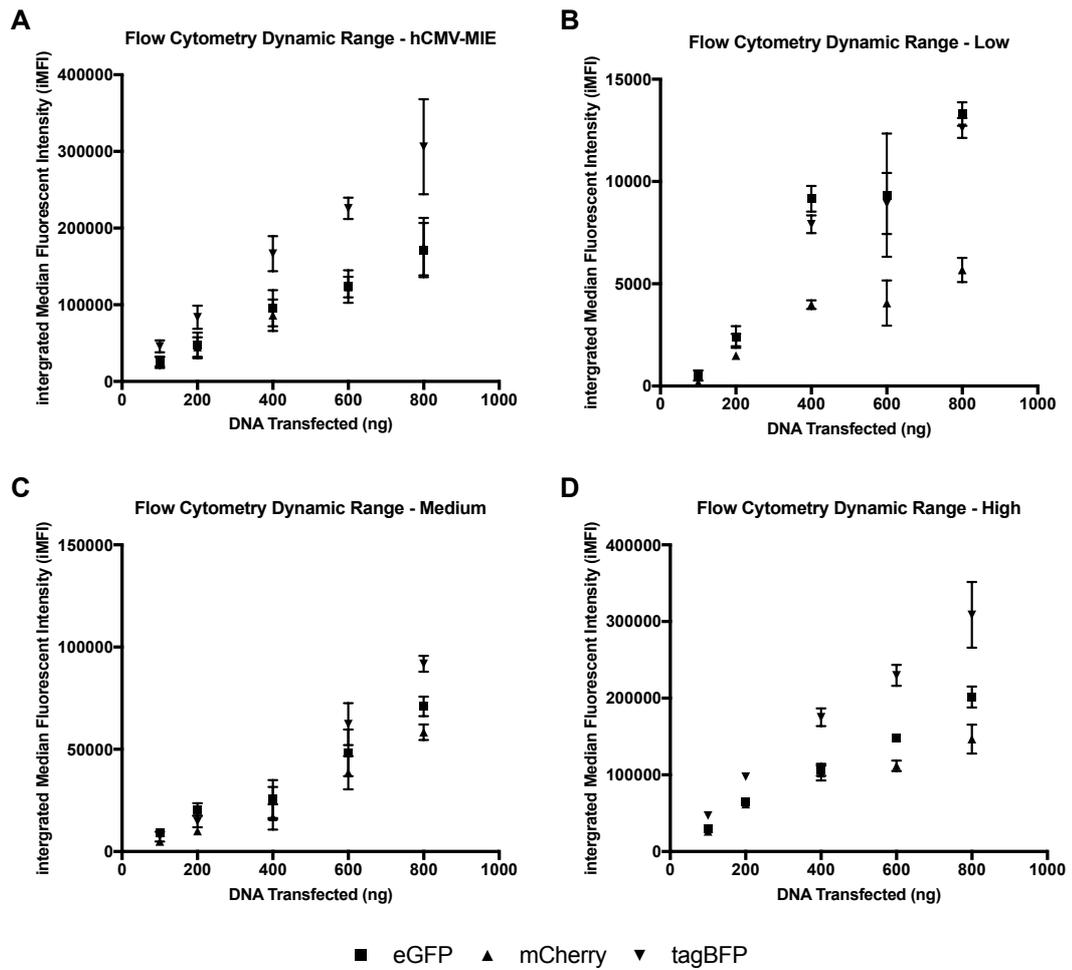


Figure B-7: Measuring dynamic range of flow cytometry by quantified integrated median fluorescent intensity.

The sensitivity and dynamic range of flow cytometry based detection was measured by titrating 100 to 800 ng of total plasmid DNA which comprised of equimolar *eGFP*, *mCherry* and *tagBFP* plasmids driven by either a hCMV-MIE (A) promoter, low (B), medium (C) and high (D) strength synthetic promoter. The iMFI were plotted against the amount of DNA transfected to demonstrate the relationship. The error bars represent the SEM of biological triplicate.

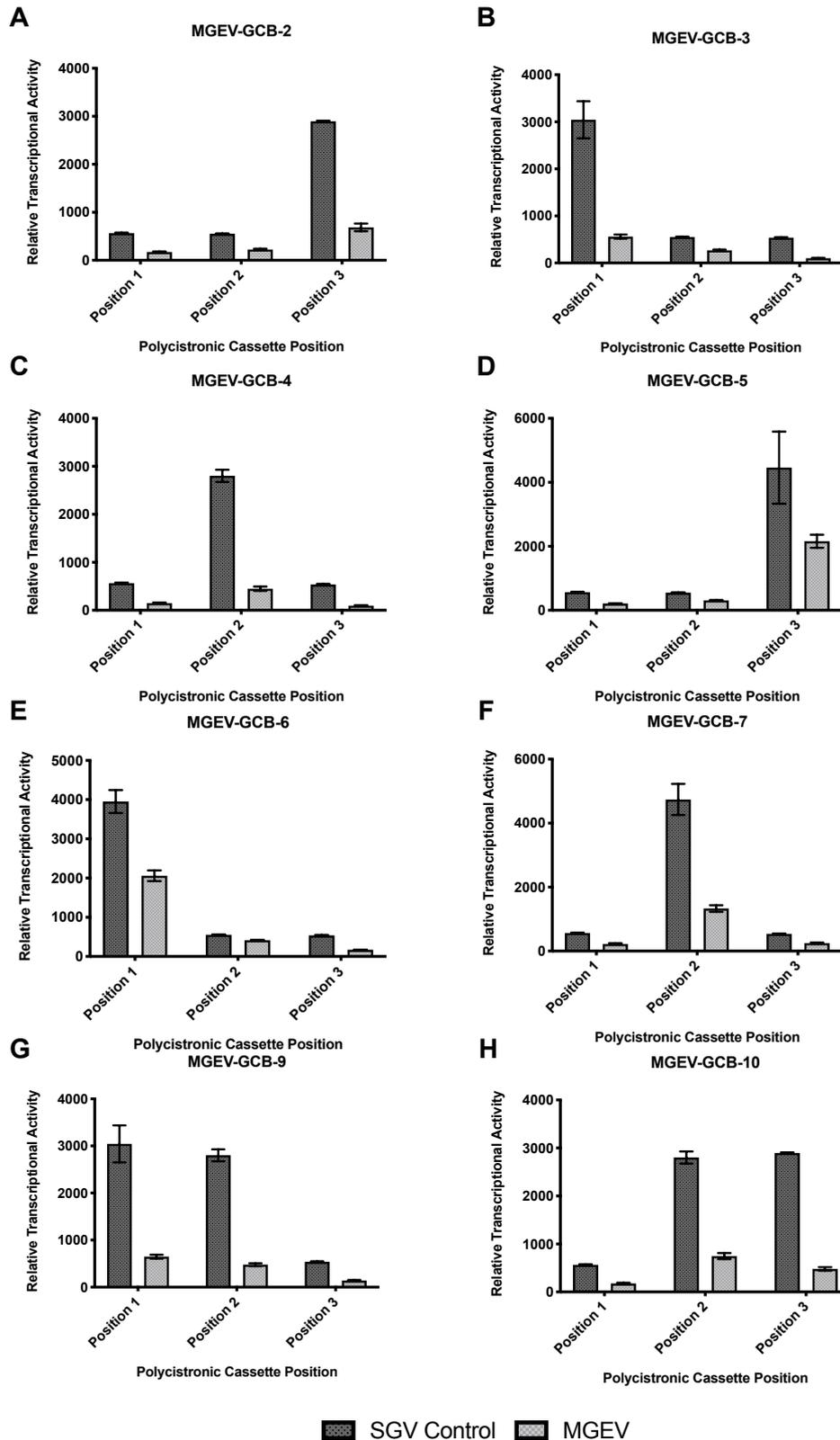


Figure B-8: Quantifying positional effects within the MGEV library -1.

Graph showing the predicted and observed RTA of position 1, 2 and 3 within a polycistronic cassette of a MGEV. The SGV control RTA represents normalised mRNA copies of co-transfected SGVs expressing *eGFP*, *mCherry* and *tagBFP* when regulated by a low, medium and high strength synthetic promoter. The observed RTA is the normalised mRNA copies of the same fluorescent proteins within a MGEV regulated by same synthetic promoter variants. The error bars represent the SEM of biological triplicate. **A-H-** represents the RTAs for MGEV-GCB-2, 3, 4, 5, 6, 7, 9 and 10 respectively.

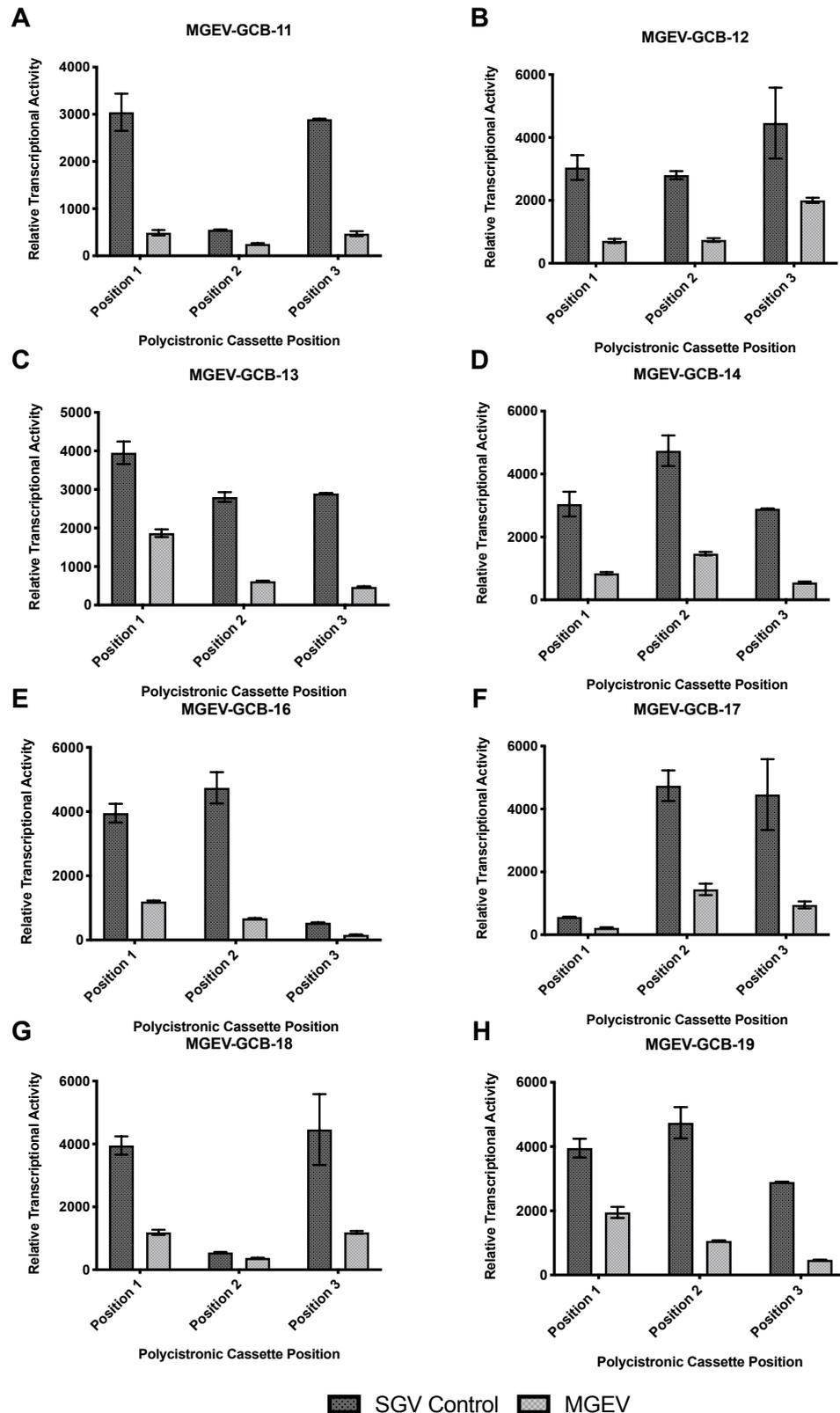


Figure B-9: Quantifying positional effects within the MGEV library -2.

Graph showing the predicted and observed RTA of position 1, 2 and 3 within a polycistronic cassette of a MGEV. The SGV control RTA represents normalised mRNA copies of co-transfected SGVs expressing *eGFP*, *mCherry* and *tagBFP* when regulated by a low, medium and high strength synthetic promoter. The observed RTA is the normalised mRNA copies of the same fluorescent proteins within a MGEV regulated by same synthetic promoter variants. The error bars represent the SEM of biological triplicate. A-H- represents the RTAs for MGEV-GCB-11, 12, 13, 14, 16, 17, 18 and 19 respectively.

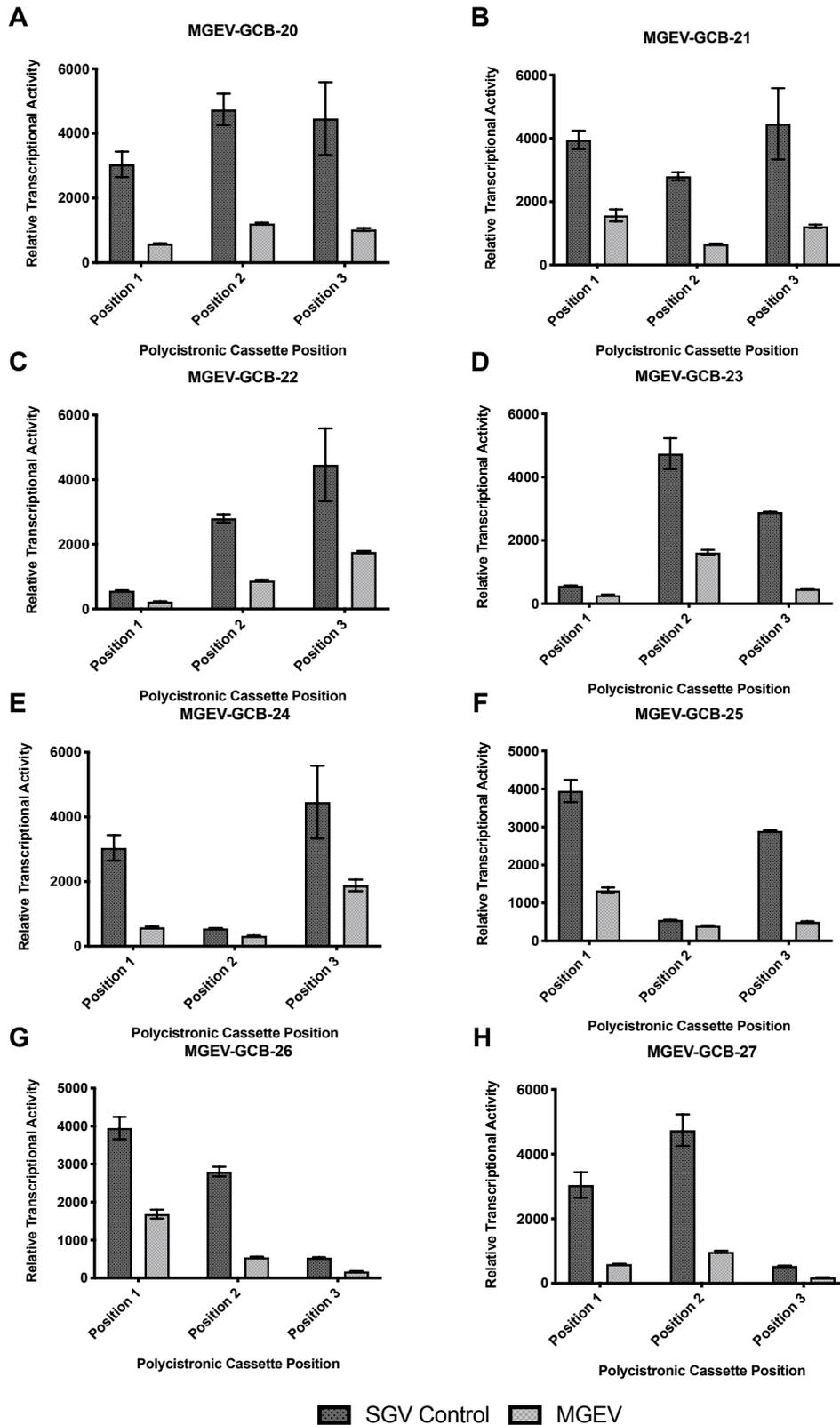


Figure B-10: Quantifying positional effects within the MGEV library -3.

Graph showing the predicted and observed RTA of position 1, 2 and 3 within a polycistronic cassette of a MGEV. The SGV control RTA represents normalised mRNA copies of co-transfected SGVs expressing *eGFP*, *mCherry* and *tagBFP* when regulated by a low, medium and high strength synthetic promoter. The observed RTA is the normalised mRNA copies of the same fluorescent proteins within a MGEV regulated by same synthetic promoter variants. The error bars represent the SEM of biological triplicate. **A-H**- represents the RTAs for MGEV-GCB-20, 21, 22, 23, 24, 25, 26 and 27 respectively.

Appendix C

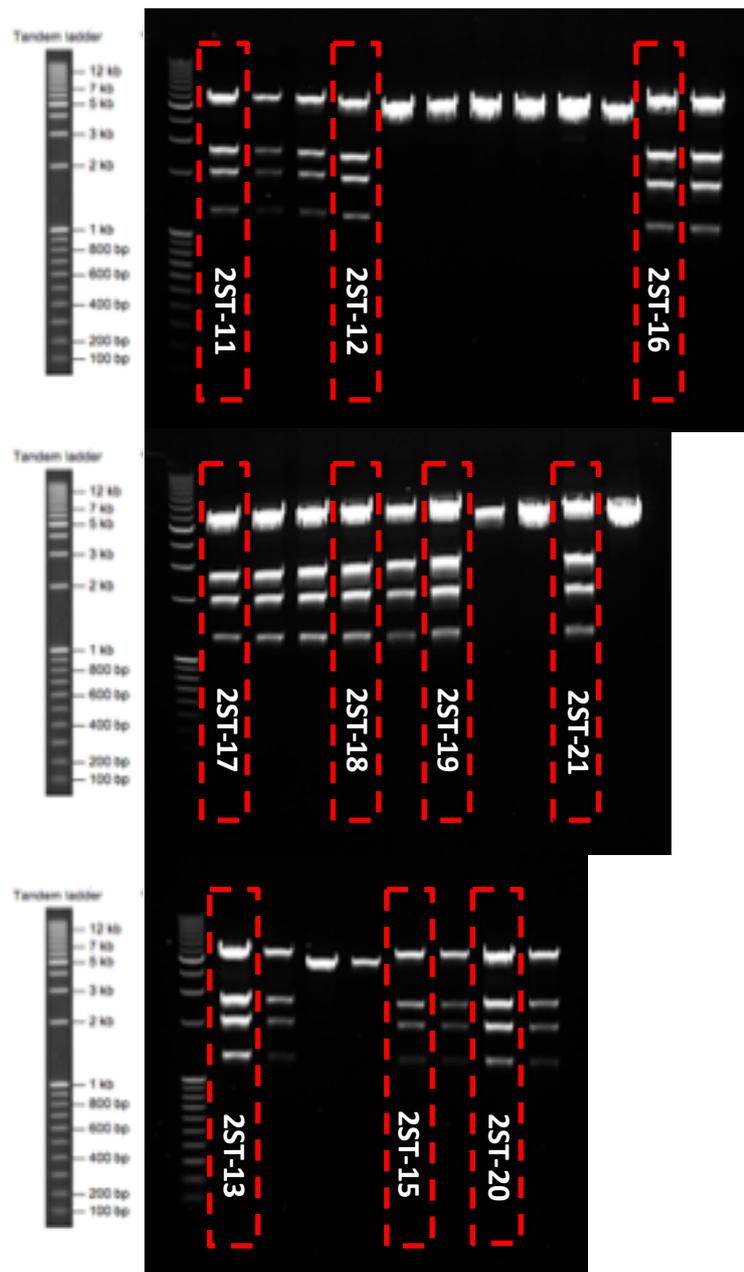


Figure C-1: Restriction digest colony screen of multi-gene expression vector variants containing the synthetic co-transcriptional cleavage elements.

A 1% agarose gel showing the results of a restriction digest colony screen using *AgeI* endonuclease. The image indicates the successful construction of the MGEV variants comprising of the synthetic CoTC elements. The lanes highlighted with the red box are the positive colonies based on fragment sizes and progressed to DNA amplification and testing.

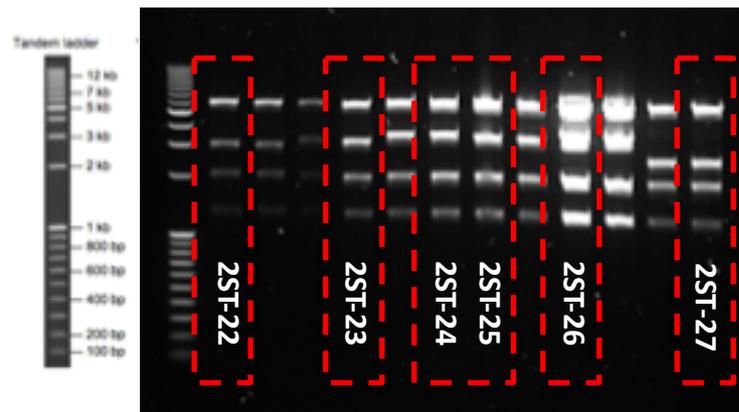


Figure C-2: Restriction digest colony screen of multi-gene expression vector variants containing the synthetic pause and ribozyme elements.

A 1% agarose gel showing the results of a restriction digest colony screen using *AgeI* endonuclease. The image indicates the successful construction of the MGEV variants comprising of the synthetic pause and ribozyme elements. The lanes highlighted with the red box are the positive colonies based on fragment sizes and progressed to DNA amplification and testing.

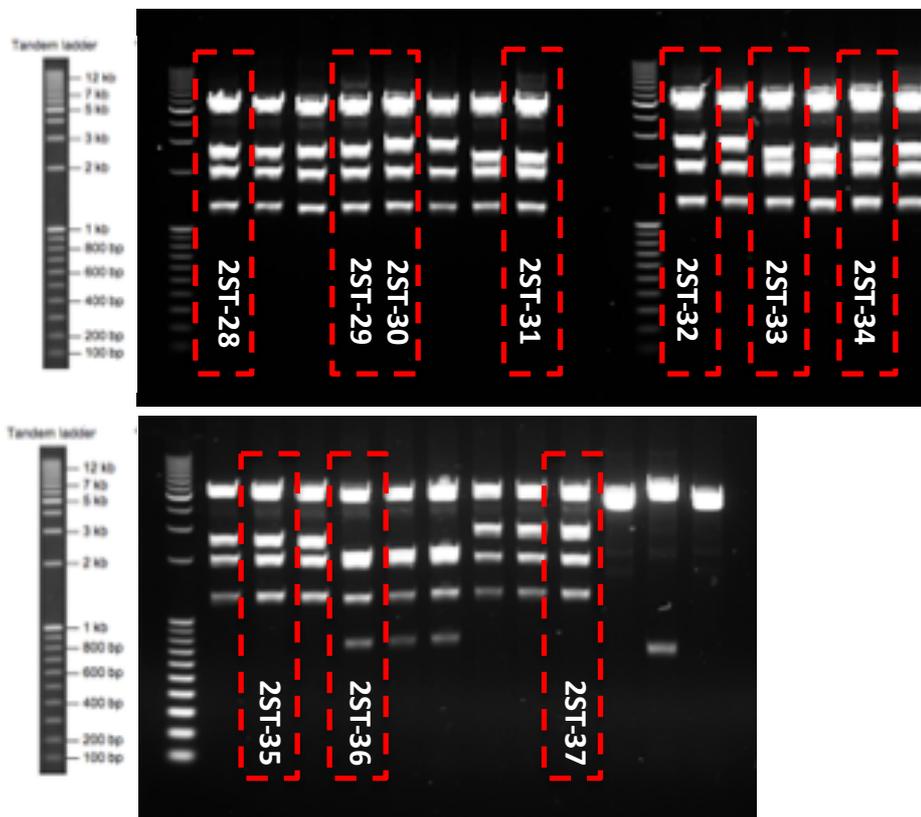


Figure C-3: Restriction digest colony screen of multi-gene expression vector variants containing the CHO-K1 polyadenylation and co-transcriptional cleavage homologues, and commercially available polyadenylation elements.

A 1% agarose gel showing the results of a restriction digest colony screen using *AgeI* endonuclease. The image indicates the successful construction of the MGEV variants comprising of the CHO-K1 pA and CoTC homologues, and the commercially available pA variants. The lanes highlighted with the red box are the positive colonies based on fragment sizes and progressed to DNA amplification and testing.

Appendix

Table 6-6: Transcription termination library.

The table shows the nucleotide sequence of the entire *in silico* designed transcription termination library, including the commercially available pA, synthetic pA, synthetic CoTC, synthetic pause, synthetic ribozyme and CHO-K1 pA and CoTC homologous variants.

Sequence Name	Sequence (5' - 3')	Sequence Length
HSVtk_pA	CCTGCAGGAACACGGAAGGAGACAATACCGGAAGG AACCCGCGCTATGACGGCAATAAAAAGACAGAATAAA AACCCACGGGTGTTGGGTGTTTTGCATAAACCCGG GGTTCGGTCCC GCGGCCGC	126
HIV-1_pA	CCTGCAGGCTGGAAGGGCTAATTCCTCCCAAAGAA GACAAGATATCCTTGATCTGTGGATCTACCACACAC AAGGCTACTTCCCTGATTAGCAGAACTACACACCAG GGCCAGGGGTGAGATATCCACTGACCTTTGGATGGT GCTACAAGCTAGTACCAGTTGAGCCAGATAAGATAG AAGAGGCCAATAAAGGAGAGAACCAGCCTTGTTAC ACCCTGTGAGCCTGCATGGGATGGATGACCCGGAGA GAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAG CATTTCATCACGTGGCCCGAGAGCTGCATCCGGAGT ACTTCAAGAAGCTGCTGACATCGAGCTTGCTACAAGG GACTTTCCGCTGGGGACTTTCCAGGGAGGCCTGGCC TGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCT GCATATAAGCAGCTGCTTTTTGCCTGTACTGCGGCC GC	472
Adeno_L1_pA	CCTGCAGGCCTTCGCCCCAGGCTGGGGAGAATGTTTT AAAAAAAAAAAAAAAAAAGCATGATGCAAATAAAA AACGCACCAAGGCCATGGCACCAGCGTTGGTTTTTC TTGTATTGCGGCCGC	123
Lambda_LC_pA	CCTGCAGGGTGACCAACCCTCACCCCCACCACGGG AGACTAGAGCTGCAGGATCCCAGGGGAGGGGTCTCT CCTCCACCCCAAGGCATCAAGCCCTTCTCCCTGCAC TCAATAAACCCCTCAATAAATATTCTCATTGTCAATCA GAAATCTTGTATTTATCTCATTTTTTCTTTTCTCACATA TAATTCCTAGCCTTTTCTGGGTTCTCAATTTGTGGTG GAAGCTTCCCTCTAGAGGCGGCCGC	246
hGH_pA	CCTGCAGGGGCATCCCTGTGACCCCTCCCAGTGCCT CTCCTGGCCCTGGAAGTTGCCACTCCAGTGCCACCA GCCTTGTCTCAATAAATTAAGTTGCATCATTTTGTCT TGACTAGGTGTCCTTCTATAATATTAGGGGTGGAGG GGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGAC AACCTGTAGGGCCTGCGGGTCTATTGGGAACCAAG CTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAAT CTCCGCTCCTGGGTTCAAGCGATTCTCCTGCCTCAG CCTCCCGAGTTGTTGGGATTCCAGGCATGCATGACCA GGCTCAGCTAATTTTTGTTTTTTGGTAGAGACGGGG TTTACCATAATTGGCCAGGCTGGTCTCCAACCTCTAA TCTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGC TGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGT CCTTCTGATTTTAAATAACTATAACCAGCAGGAGGA CGTCCAGACACAGCATAGGCTACCTGGCCATGCCCA ACCGGTGGGACATTTGAGTTGCTTGCTTGGCACTGTC CTTCATGCGGCCGC	602
bGH_pA	CCTGCAGGCTGTGCCTTCTAGTTGCCAGCCATCTGTT GTTTGGCCCTCCCCGTGCCTTTCCTTGACCCTGGAAG GTGCCACTCCCACTGTCCTTTCCTAATAAATGAGGA AATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATT CTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA GGATTGGGAAGACAATAGCAGGCATGCTGGGGATGC GGTGGCTCTATGGCGGCCGC	241

Appendix

Sequence Name	Sequence (5' - 3')	Sequence Length
chef1 α _pA	CCTGCAGGATATTACCCCTAACACCTGCCACCCCAGTC TTAATCAGTGGTGGGAAGAACGGTCTCAGAACTGTTTGT CTCAATTGGCCATTTAAGTTTAATAGTGAAAGACTGGT TAATGATAACAATGCATCGGAAAACCTTCAGGAGGAA AGGAGAATGTTTTGTGGAACATTTTTGTGTGTGGCA GTTTTAAGTTATTAGTTTTCAAAAATCAGTACTTTTTAAT GAAACAACCTTGACCAAAAATCTGTACAGAAATTTG AGACCCATTAAAATACAAGTTTAATGAGAAGTCTGTC TCTGTTAATGCTGAAGTCATTAAGTGTCTTAGCTTA GCAAGGTATGTGGATGCCCATTTGTGTTCCAAGGGATT GGACTGTTTCATCAGGACCCAGAGCTGAGTTTCAAGGG CTCAAGAGATGGCTTATTACCTGTGGGTGCTTGAAGG TTCTGGTTGGGACAAAATTAGGAATGTTTTTGGCAGACA TGGTGACTACCTTCATCTGGGTGAGTTCAGTTGATTTG TCTTGAGCCTTGGGGTTTACACAAGTAAATGACATCA TACAGTTAGTGTATTGTTAGTGAATATTAATATATGAG GCAGGCTTTGCTCTAGCAATTTTGC GGCCGC	698
Synth_pA_1	CCTGCAGGTGTAGAGGTTTTACTTGCTTTAAAAAACCT CCCACACCTCCCCCTGAACCTGAAACATAAAAATGAAT GCAATTGTTGTTGTTAACTGTTTATTGCAGCTTATAAT GGTTACAATAAAGCAATAGCATCACAAATTTACAA ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGT GCGGCCGC	198
Synth_pA_2	CCTGCAGGTGTAATAAAAAATAAAGTTGTGGTTTGTG CGGCCGC	44
Synth_pA_3	CCTGCAGGTGTAATAAAGCAATAGCATCACAAATTT CACAAATAAAGTTGTGGTTTGTGCGGCCGC	67
Synth_pA_4	CCTGCAGGTGTAATAAAAAGACAGATAAAGTTGTG GTTTGTGCGGCCGC	51
Synth_pA_5	CCTGCAGGTGTAATAAAGCAATAGCATCACAAATTT CACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTG GTTTGTGCGGCCGC	89
Synth_pA_6	CCTGCAGGAATAAAGCATTTTTTTCACTGCATTCTAGT TGTGGTTTGTGCGGCCGC	56
Synth_pA_7	CCTGCAGGAATAAAGCAATAGCATCACAAATTTACAA AATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTT GTGTTGTGGTTTGTGCGGCCGC	97
Synth_pA_8	CCTGCAGGAATAAAGCAATAGCATCACAAATTTACAA AATAAAGCATTTTTTTCACTGCATTCTAGGGGGGTGGG GTGGGGGCGGCCGC	89
Synth_pA_9	CCTGCAGGTGTAATAAAGCAATAGCATCACAAATTT CACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTG GTTTGTAAATAAATGAGGAAATTGCATCGCATTGTCTG AGTAGGTGCATTCTATTCTGGGGGTGGGGTGGGGG CGGCCGC	157
Synth_pA_10	CCTGCAGGATCATAATCAGCCATACCACATTTACAG AGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCC CTGAACCTGAAACATAAAAATGAATGCAATTGTTGTTGT TAACTGTTTATTGCAGCTTATAATGGTTACAATAA GCAATAGCATCACAAATTTACAAATAAAGCATTTTTT TCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAA TGTATCTTATCATGTCTGGATCGCGGCCGC	257
Synth_CoTC_1	GCGGCCGCAAATTGAAATTCAAATTGAAATTCAAATT GAAATTGAAATTCAAATTCAAATTTGGTACC	67
Synth_CoTC_2	GCGGCCGCAAATTGAAATTCAAATTGAAATTCAAATT CAAATTGAAATTCAAATTGATAAAAAATAGGTACC	72

Appendix

Sequence Name	Sequence (5' - 3')	Sequence Length
Synth_CoTC_3	GCGGCCGCAAATTGCGCACAAAATTCACGAAATTG ACTGGAAATTGCGAGGAAATTCTGTCAAATTTGAGTT CAAATTTGCCGGAAATTCGCTTGAATTTGGTACC	108
Synth_CoTC_4	GCGGCCGCAAATTCGCACTCGCACATAAAATTTGGCAC TCATATCACCTAAATTCGCACTCCACGTTCCGAAATTTG GCACTCATATCACCTAAATTTGCACTCAGTTCCGCCACAA ATTTGGCACTAGTTCCGCCACAAATTCGCACTGCCGGGA ATAAAATTTGGCACTGCTTGCCAATAATTTGGTACC	185
Synth_CoTC_5	GCGGCCGCAAATTTGCGCACAAATTTGCATATAAATTTCC CACGAAATTTGCATATAAATTTGAGTTCAAATTTGCTGAA GAAATTTGCCGGAAATTCGCTTGAATAAAAATAGGTA CC	113
Synth_CoTC_6	GCGGCCGCAAATTTGCGCACATATTGGAGTAAATTTCCA TATCACCTGCGAGAAATTTCCACGTTCCGAAATTTAAAT TGCATATCACCTGCGAGAAATTTGTGTCAGACATCCGGT AAATTTCTGAAGGGCATTCCAAAATTTGCCGGGAATA TCTACAAATTTGGCTTGCCAATAGACTATAAAAATAG GTACC	192
Synth_CoTC_7	GCGGCCGCAATTAATTAAGAATTAATTAACAATTAATTT AAGAATTAATTAACAATTAATTAAGAATTAATTAAGA ATTAATTAAGAATTAATTAAGAATTTGGTACC	107
Synth_CoTC_8	GCGGCCGCAATTAATTAAGCGCACAAATTAATTAACCA TATAATTAATTAAGCCACGAATTAATTAAGCGCTAAAT TAATTAAGCATATAATTAATTAAGAGTTCAATTAATTA AGGCCGGAATTAATTAACGCTTGAATTTGGTACC	147
Synth_CoTC_9	GCGGCCGCAATTAATTAAGATATTGGAGTGCATAAT TAATTAACCATATCACCTGCGAGAAATTAATTAACCCAC GTTCCGAAATTAATTAATTAACCATATCACCTGCGAGA ATTAATTAAGCTGTGACACATCCGGT AATTAATTAAGTGAAGGGCATTCCAAAATTAATTAAC GCCGGGAATATCTACAATTAATTAAGGCTTGCCAATA GACTAATTTGGTACC	229
Synth_CoTC_10	GCGGCCGCAATTAATTAATTATAGAATTAATTAATTAT AGAATTAATTAATTATACAATTAATTAATTATACAATT AATTAATTATAGAATTAATTAATTATACAATTAATTA TTATAGAATTAATTAATTATAGAATTTGGTACC	147
Synth_CoTC_11	GCGGCCGCAATTAATTAATTATACCGCACAAATTAATTA ATTATAGCATATAATTAATTAATTATACCCACGAATTA ATTAATTATACATATCAATTAATTAATTATAGAGTTCA ATTAATTAATTATAGGCCGGAATTAATTAATTATAGGC TTGAATTAATTAATTATACCATGGAATTTGGTACC	187
Synth_CoTC_12	GCGGCCGCAATTAATTAATTATAGCGCACATATTGGA GTAATTAATTAATTATACCGCACATATTGGAGTAATTA ATTAATTATACCCACGTTCCGAAATTAATTAATTAAT TATAGCATATCACCTGCGAGAAATTAATTAATTATAGAG TTCGCCACTCGAAATTAATTAATTATACTGAAGGGCAT TCCAAAATTAATTAATTATAGGCCGGGAATATCTAAAT TAATTAATTATAGGCTTGCCAATAGACAATTTGGTACC	265

Appendix

Sequence Name	Sequence (5' - 3')	Sequence Length
Synth_Pause_1	<p> GCGGCCGCTTCGATAGACTGATAGGGGATGCAGTATATC CCTGGATAACAATAGACGCACAGGTTGGAATCCTAAGTG AAGTCGCGCGTCCGAACCCAGCTCTATTTTAGAGGTCAT GGTTCTGGTGCCCGCGAGCCGCGGAACCGATTAGGGG CATGTACAACAATATTTATTAGTCATCTTTCAGACACAA TCTCCCAGCTCACTGGTATATAGTTCCTGCTATAATTAG CCTCCCTCATAAGTTGCACCTGGCCTTGGGGGAGGGGG AGGCCAGAATGACTGGCCTTGGGGGAGGGGGAGGCCA GAATGACTGGCCTTGGGGGAGGGGGAGGCCAGAATGAC TGGCCTTGGGGGAGGGGGAGGCCAGAATGAGGTACC </p>	380
Synth_Pause_2	<p> GCGGCCGCTTCGATAGACTGATAGGGGATGCAGTATATC CCTGGATAACAATAGACGCACAGGTTGGAATCCTAAGTG AAGTCGCGCGTCCGAACCCAGCTCTATTTTAGAGGTCAT GGTTCTGGTGCCCGCGAGCCGCGGAACCGATTAGGGG CATGTACAACAATATTTATTAGTCATCTTTCAGACACAA TCTCCCAGCTCACTGGTATATAGTTCCTGCTATAATTAG CCTCCCTCATAAGTTGCACATGCGAGATCGATAGACTGA TAGGGGATGCAGTATATCCCTGGATAACAATAGACGCAC AGGTTGGAATCCTAAGTGAAGTCGCGCGTCCGAACCCA GCTCTATTTTAGAGGTCATGGGTTCTGGTGCCCGCGAGC CGCGGAACCGATTAGGGGCATGTACAACAATATTTATT AGTCATCTTTCAGACACAATCTCCCAGCTCACTGGTATA TAGTTCCTGCTATAATTAGCCTCCCTCATAAGTTGCACC TGGCCTTGGGGGAGGGGGAGGCCAGAATGACTGGCCTT GGGGGAGGGGGAGGCCAGAATGACTGGCCTTGGGGGA GGGGGAGGCCAGAATGACTGGCCTTGGGGGAGGGGGA GGCCAGAATGAGGTACC </p>	630
Synth_Pause_3	<p> GCGGCCGCTTCGATAGACTGATAGGGGATGCAGTATATC CCTGGATAACAATAGACGCACAGGTTGGAATCCTAAGTG AAGTCGCGCGTCCGAACCCAGCTCTATTTTAGAGGTCAT GGTTCTGGTGCCCGCGAGCCGCGGAACCGATTAGGGG CATGTACAACAATATTTATTAGTCATCTTTCAGACACAA TCTCCCAGCTCACTGGTATATAGTTCCTGCTATAATTAG CCTCCCTCATAAGTTGCACATGCGAGATCGATAGACTGA TAGGGGATGCAGTATATCCCTGGATAACAATAGACGCAC AGGTTGGAATCCTAAGTGAAGTCGCGCGTCCGAACCCA GCTCTATTTTAGAGGTCATGGGTTCTGGTGCCCGCGAGC CGCGGAACCGATTAGGGGCATGTACAACAATATTTATT AGTCATCTTTCAGACACAATCTCCCAGCTCACTGGTATA TAGTTCCTGCTATAATTAGCCTCCCTCATAAGTTGCACA TGCAGATCGATAGACTGATAGGGGATGCAGTATATCC CTGGATAACAATAGACGCACAGGTTGGAATCCTAAGTGA AGTCGCGCGTCCGAACCCAGCTCTATTTTAGAGGTCATG GGTTCTGGTGCCCGCGAGCCGCGGAACCGATTAGGGGC ATGTACAACAATATTTATTAGTCATCTTTCAGACACAAT CTCCCAGCTCACTGGTATATAGTTCCTGCTATAATTAGC CTCCCTCATAAGTTGCACCTGGCCTTGGGGGAGGGGGA GGCCAGAATGACTGGCCTTGGGGGAGGGGGAGGCCAG AATGACTGGCCTTGGGGGAGGGGGAGGCCAGAATGACT GGCCTTGGGGGAGGGGGAGGCCAGAATGAGGTACC </p>	880

Appendix

Sequence Name	Sequence (5' - 3')	Sequence Length
Synth_Pause_4	<p>GCGGCCGCTCGATAGACTGATAGGGGATGCAGTATA TCCCTGGATACAATAGACGCACAGGTTGGAATCCTA AGTGAAGTCGCGCGTCCGAACCCAGCTCTATTTTAG AGGTCATGGGTTCTGGTGCCCGCGAGCCGCGGAACC GATTAGGGGCATGTACAACAATATTTATTAGTCATCT TTCAGACACAATCTCCCAGCTCACTGGTATATAGTTC CTGCTATAATTAGCCTCCCTCATAAGTTGCACCTGGC CTTGGGGGAGGGGGAGGCCAGAATGACTGGCCTTGG GGGAGGGGGAGGCCAGAATGACTGGCCTTGGGGGA GGGGGAGGCCAGAATGACTGGCCTTGGGGGAGGGG GAGGCCAGAATGACTGGCCTTGGGGGAGGGGGAGG CCAGAATGACTGGCCTTGGGGGAGGGGGAGGCCAG AATGACTGGCCTTGGGGGAGGGGGAGGCCAGAATG ACTGGCCTTGGGGGAGGGGGAGGCCAGAATGAGGT ACC</p>	504
Synth_Pause_5	<p>GCGGCCGCTCGATAGACTGATAGGGGATGCAGTATA TCCCTGGATACAATAGACGCACAGGTTGGAATCCTA AGTGAAGTCGCGCGTCCGAACCCAGCTCTATTTTAG AGGTCATGGGTTCTGGTGCCCGCGAGCCGCGGAACC GATTAGGGGCATGTACAACAATATTTATTAGTCATCT TTCAGACACAATCTCCCAGCTCACTGGTATATAGTTC CTGCTATAATTAGCCTCCCTCATAAGTTGCACATGCG AGATCGATAGACTGATAGGGGATGCAGTATATCCCT GGATAACAATAGACGCACAGGTTGGAATCCTAAGTGA AGTCGCGCGTCCGAACCCAGCTCTATTTTAGAGGTCA TGGGTTCTGGTGCCCGCGAGCCGCGGAACCGATTAG GGGCATGTACAACAATATTTATTAGTCATCTTTCAGA CACAATCTCCCAGCTCACTGGTATATAGTTCCTGCTA TAATTAGCCTCCCTCATAAGTTGCACCTGGCCTTGG GGAGGGGGAGGCCAGAATGACTGGCCTTGGGGGAG GGGGAGGCCAGAATGACTGGCCTTGGGGGAGGGGG AGGCCAGAATGACTGGCCTTGGGGGAGGGGGAGGC CAGAATGACTGGCCTTGGGGGAGGGGGAGGCCAGA ATGACTGGCCTTGGGGGAGGGGGAGGCCAGAATGAC TGGCCTTGGGGGAGGGGGAGGCCAGAATGACTGGCC TTGGGGGAGGGGGAGGCCAGAATGAGGTACC</p>	754

Appendix

Sequence Name	Sequence (5' - 3')	Sequence Length
Synth_Pause_6	<p>GCGGCCGCTCGATAGACTGATAGGGGATGCAGTATA TCCCTGGATACAATAGACGCACAGGTTGGAATCCTA AGTGAAGTCGCGCGTCCGAACCCAGCTCTATTTT AGGTCATGGGTTCTGGTGCCCGGAGCCGCGGAACC GATTAGGGGCATGTACAACAATATTTATTAGTCATCT TTCAGACACAATCTCCCAGCTCACTGGTATATAGTTC CTGCTATAATTAGCCTCCCTCATAAGTTGCACATGCG AGATCGATAGACTGATAGGGGATGCAGTATATCCCT GGATACAATAGACGCACAGGTTGGAATCCTAAGTGA AGTCGCGCGTCCGAACCCAGCTCTATTTTAGAGGTCA TGGGTTCTGGTGCCCGGAGCCGCGGAACCGATTAG GGGCATGTACAACAATATTTATTAGTCATCTTTCAGA CACAATCTCCCAGCTCACTGGTATATAGTTCCTGCTA TAATTAGCCTCCCTCATAAGTTGCACATGCGAGATCG ATAGACTGATAGGGGATGCAGTATATCCCTGGATAC AATAGACGCACAGGTTGGAATCCTAAGTGAAGTCGC GCGTCCGAACCCAGCTCTATTTTAGAGGTCATGGGTT CTGGTGCCCGGAGCCGCGGAACCGATTAGGGGCAT GTACAACAATATTTATTAGTCATCTTTCAGACACAAT CTCCCAGCTCACTGGTATATAGTTCCTGCTATAATTA GCCTCCCTCATAAGTTGCACCTGGCCTTGGGGAGG GGGAGGCCAGAATGACTGGCCTTGGGGAGGGGGA GGCCAGAATGACTGGCCTTGGGGAGGGGGAGGCC AGAATGACTGGCCTTGGGGAGGGGGAGGCCAGAA TGACTGGCCTTGGGGAGGGGGAGGCCAGAATGACT GGCCTTGGGGAGGGGGAGGCCAGAATGACTGGCCT TGGGGAGGGGGAGGCCAGAATGACTGGCCTTGGG GGAGGGGGAGGCCAGAATGAGGTACC</p>	1004
Synth_Ribo_1	<p>GCGGCCGCCCTGTCACCGGATGTGTTTTCCGGTCTGA TGAGTCCGTGAGGACCAACAGGGGTACC</p>	53
ActB_pA	<p>CCTGCAGGTGTACATTTTTTTTCTTTTTTAAGTCATTC CAAGTACCCATGAGATGGCTACAGGAAGTCCCTCAC CCTCCAAAAGCCATCCCATTCCTAGAAAGAGGAT GGCTGAGTCCATGCCCTGAGTCCACACCGGGGAGGT GACAGCATTGCTTCTGTGTA AATTATGTA CTGCAAAA ATTTTTTAAATCTTCCGCCTTAATACTTCTTTTGT TTATTTTTGAATGGTCAGCCATCGTGGCCCTTTTTTT TTTTTTTTTTTTTCCCCCAACTTGATGTATGAAGG CTTTTGGTCTCCCTGGGAGGGGTTGAGGTGTTGAGG CAGCCAGGGCTTGCCTGTACACTGACTTGAGACCAG TTAATAAA GTGCACACCTTACAAACA GTGCTGCTTG TTTGTGGCTTTGCTAGATTCTGGGTAGCAGCGGGGGA GGGGTCACTATTACCTTGTCTCCAAGAGGTTCTAGG GTGGTCTGGGGCGGCCGC</p>	500
S100a6_pA	<p>CCTGCAGGGCACGGTCTCTACCCACACCTGCAGC TCCTTGTCTTTCCCTCTGCAGCCTCTTAAACTGCTCCT CTTACGCCCTGGCCCTTCTCTTCTCATGGGTGGAT TCTTCCAGTAGAGAAATAAAGCCCTTCCCCCTTCC ATGTGTTGGTTTTGAGGTGGTTTGTCTCCGTTGGCTG AGTCAGGGGAGAACAGACAGACATTTTGAAGCATT AGCCTCAGGTCACACACAGGTGGCCTGTGGGTGCAG GGGGTGGACTTTGCGGCCGC</p>	278

Appendix

Sequence Name	Sequence (5' - 3')	Sequence Length
Gadph_pA	<p>CCTGCAGGGAAGCCCACCTGGACCATCCACCCAGC AAGGACTCGAGCAAGAGGGAGGCCCTGGCTGCTGAGC AGTCCCTGTCCAATAACCCACACCGATCATCTCCCT CACAGTGTCCATCCCAGACCCCAAGAATAAGGAGGGG CTTAGGGAGCCCTACTCTCTTGAATACCATCAATAAAG TTCCTGACCCATCTTCCCTGGCCTTTCAATGTAAGG TTGGGGGAGGGAGGCTGTGACCTAGCAAGGTTGGGAA TCCTCTGTGTCACCTTTCAAACAGGGCACTAGCCACAT GCCAGCCAGGTTTCTGTCTGACAGATGAAAATTC ACCTAAGGTGCTTTGGTGCTGGGAGGAGTGGGGGTTG GCGGCCGC</p>	383
ActB_CoTC	<p>GCGGCCGCAAAATTCATATACAGTCACAAGGGCTGGA GACATGGCTTAGTTCTTACAGGCTAAGGCTCAAGACC AAAAGCACGTGGAGAGCAGAGGCCATGTGGTCTCACT AATGTAGACAACAAAATTCAAAGGCCAGATGTGGACA CCTGTCATCCCAGCTCTGAGGCAAGCCTGAGCCAGCT TGGACTTCAAGATGAGACCCATATTTTAAGGGAATGAA TCTCATAAATATTGTCTCAGAGATGAAGGCAGGAAGG TGAGTACAAAGCCATCCCTAGCTATTCATTGAGTTTGC CTGGTATACATGAGACACTGCCTCAAACAAAAGTCCC TAAACAAATAAACACCTCAGGCAAGAACAAAACCTA AGAAATTTGATTTTTGCCCACTCCAAACAAGCTGGGC ATGTTGGTGCACAACTTTAACCTCCAGTTGTCTGGAAGC AGAGGCAGGTGGATGTGTGAGTTTGAAGTTCAGCCTGG TTACTGTGTGAGATCCTTGACAGGCAGAGCTTCATAG AGAGACCATCTTGAAAAAGAGACCAAAGCAAAGGTA ATAGGTACC</p>	568
S100a6_CoTC	<p>GCGGCCGCATTAATTTGTAAGCTGCTCTAAAGATGAA CTTCCAGGCAGTGAGCTGGAAGAAGCGAGTTAGACAG AAATTTATTGTTGGTGGGGGATGGTGTCTGAAATCCTT TAGACTGTGTCCTCCCTTTTTTTGAGACAGGGTTTT ATATAGCCCAGGTTGGCTCAGAATTCTGCCTCGTGGGA TCAACCTACTGAGCTATATCCCCAAGTCTTAAACTAGT GAGGTCAAACCACCTATCAGAGGGGTTGCCTAAGAT CATCGGAAAACACAAGTATTTCACTGAGATTTCATAA CAGTAGCAAAATTAACGGTGTGAAGCAGCAGTGAATA AATTTTATGATTGGGGGACACCACAACATGAGAATCT GTGTCCAAGGGTCATAGAATTAGGTACC</p>	402

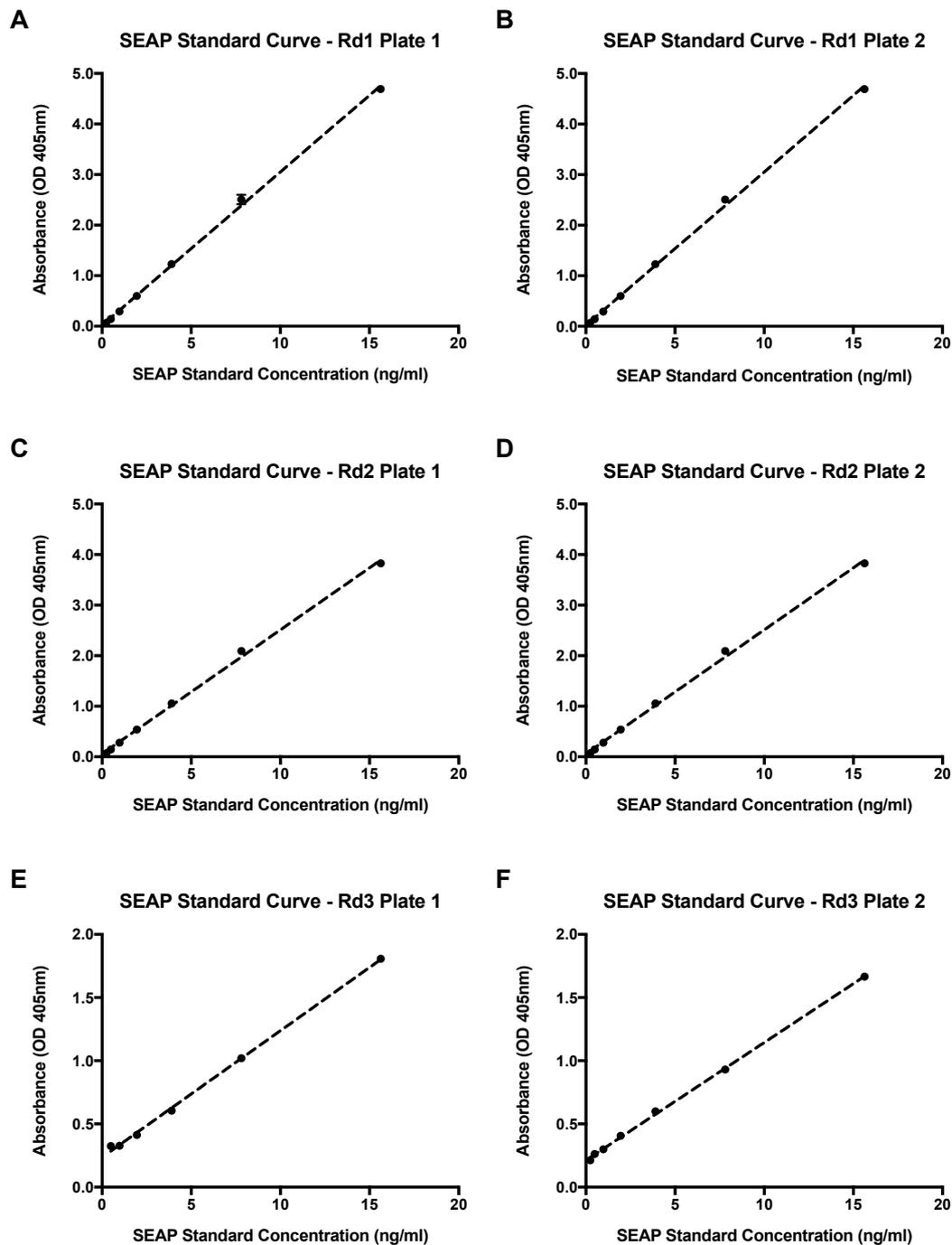


Figure C-4: Secreted alkaline phosphatase standard curves.

A-F- shows the *SEAP* standard curves generated per assay plate using the human placental alkaline phosphatase provided within the SensoLyte™ pNPP *SEAP* reporter gene assay colourmetric kit at concentrations ranging from 15.63 to 0.24 ng/ml. The y-axis shows the absorbance detected at 405nm. The error bars represent the SEM of technical duplicates. The dotted line represents the trendline which by linear regression was performed to calculate the concentration of *SEAP* expressed during screening of the terminator library within the bicistronic cassette.