

Developing a lactate-inducible transgene expression system for use in Chinese hamster ovary cells

Charles Motraghi

Imperial College London
Department of Life Sciences
Imperial College Centre for Synthetic Biology

Submitted for the Degree of Doctor in Philosophy

Declaration of originality

I declare that this thesis represents my own work except where referenced.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

The accumulation of lactate during cultivation of mammalian cells for biopharmaceutical production is a longstanding issue affecting glycosylation quality and productivity. Many approaches exist to mitigate its impact, either through the replacement of glucose with slowly metabolised sugars, dynamic feeding strategies, or host cell engineering. The manipulation of genes in this latter approach is constitutive and may suboptimally respond to cellular needs.

The LldR proteins from *Corynebacterium glutamicum* and *Pseudomonas aeruginosa* have been used in this project to create a lactate-inducible transgene expression system, which can be used subsequently to dynamically drive expression of proteins previously targeted to mitigate the accumulation of lactate. Expression and purification of these LldR proteins, fused to transcriptional effector domains in various orientations and with fusion linkers in certain cases, allowed *in vitro* characterisation and optimisation of the constituent parts of the inducible system. This provided crucial information in some cases about the need to use a flexible linker between LldR and a VP64 transactivation domain. *In vivo* experimentation of these optimised systems showed significant levels of induction in response to 20 mM lactate, with a 3.46-fold decrease in expression seen for one construct. Some preliminary work was also carried out with Cas9-VPR, which was shown to be able to upregulate transiently transfected genes up to 1.6-fold. In the future, this will be a useful tool for upregulating multiple previously identified targets, as well as helping to find new beneficial targets.

The general approach outlined here for the development of this lactate-inducible transgene expression system will be appropriate for any other such project where the ligand of interest is a central metabolite. Inherently weaker induction might be a feature of such a system, given the presence of the inducer at low and relatively benign or neutral concentrations throughout a period of interest; testing an unoptimised system in mammalian cells may return little or no detectable induction signal and therefore it will not be straightforward to optimise such a system solely through the use of *in vivo* experimentation. *In vitro* characterisation of the inducible system components, as performed here, can provide essential feedback regarding the impact of effector domain fusion and operator design on the DNA-binding affinity of the biosensor prior to *in vivo* testing.

The work in this thesis will allow the future exploration of dynamically regulated host cell engineering designed to combat the lactate accumulation phenotype.

Acknowledgements

Thank you to my supervisors, Karen Polizzi and Cleo Kontoravdi. I'm grateful for having been given the opportunity to undertake this project with you both. In particular, I want to thank Karen for all of the guidance and support she gave me throughout, as well as her diligence in helping me write this thesis.

The people in the Polizzi Lab and the Synthetic Biology Lab were a true pleasure to work with all of these years - thank you to all of you for your help and camaraderie. Many of you contributed to this project. Rochelle Aw helped me in all things, including how to perform RT-qPCR; James Arpino showed me how to use the FPLC machine and gave me general support with protein expression, as did Kate Royle and Kealan Exley; Antony Constantinou gave me a crash course in mammalian cell culture; Joyce Riangrungrroj helped me with protein expression and cloning, and was a general sounding board; Will Shaw and Olivier Borkowski trained me on the various flow cytometers; and George Pothoulakis showed me how to use the fluorescence microscope. I am also grateful towards my Progress Review Panel, John Heap and Mark Isalan, who were able to give me such useful advice at an early stage of this project. Thank you also to the two students I supervised, Abisola Fafolu and Yifan Zhou, for all of their work. I worked with Becky Wilson and Kay Penicud during my three month placement in the Imperial Corporate Partnerships team, and I couldn't have asked for better supervisors there.

Thank you to all of my family, of course. Mary, Nooshie, Nadia and Sara took care of me since I moved to London in 2010 (sorry for stealing so much of your tupperware). Jerry guarded me from postmen and planes while I was writing up. I am profoundly grateful to my parents, who have supported me from the beginning and to whom I dedicate this thesis.

Finally, thank you to Grace.

Table of contents

Declaration of originality.....	2
Copyright Declaration	2
Abstract.....	3
Acknowledgements.....	4
List of figures.....	9
List of tables.....	13
Abbreviations	14
1 Introduction.....	17
1.1 Biopharmaceuticals.....	17
1.2 Expression hosts.....	19
1.2.1 <i>Escherichia coli</i>	20
1.2.2 Yeast.....	22
1.2.3 Mammalian cells	23
1.2.3.1 Scale up.....	25
1.2.3.2 Productivity loss through genetic instability, promoter silencing and stochastic gene expression	27
1.2.3.3 Accumulation of the toxic metabolites lactate and ammonia	28
1.2.3.3.1 Lactate metabolism in mammalian cells.....	30
1.3 Combatting lactate accumulation.....	32
1.3.1 Growth media formulation.....	32
1.3.2 Feeding strategies and metabolite monitoring	33
1.3.3 Host cell engineering	35
1.3.3.1 Constitutive and dynamic approaches to host cell engineering.....	37
1.3.3.1.1 Lactate-inducible host cell engineering	38
1.3.3.2 Manipulation of endogenous gene expression as a host cell engineering strategy	39
1.3.3.2.1 Cas9-VPR as a convenient and versatile genome targeting tool ...	40
1.4 Biosensor design and testing for use in inducible transgene expression systems	42
1.4.1 What is a biosensor?	42
1.4.2 One-component systems and their use as mammalian transgene expression systems	42
1.4.3 Lactate-binding proteins suitable for use in a mammalian transgene expression system	47
1.4.4 The LldR operators; designing an operator for use with a one-component system..	48
1.4.5 Inducible transgene expression system design	49
1.4.5.1 De-repression by steric hindrance	50

1.4.5.2	De-repression by heterochromatin recruitment	54
1.4.5.3	Transactivation	54
1.4.6	Protein fusion design	54
1.4.7	Testing one component systems in mammalian systems.....	55
1.4.8	Promoters used.....	57
1.5	Project aims	57
2	Materials and methods	59
2.1	Plasmids	59
2.2	Molecular biology techniques.....	69
2.2.1	Restriction enzyme digestion	69
2.2.2	Insert generation via annealed short oligonucleotides.....	69
2.2.3	High-fidelity PCR with Phusion polymerase.....	69
2.2.4	Gibson assembly	70
2.2.5	Agarose gel electrophoresis, imaging and purification	70
2.2.6	DNA ligation	71
2.2.7	Preparation of chemically competent <i>Escherichia coli</i> cells; transformation of <i>Escherichia coli</i>	71
2.2.8	Colony PCR with Taq polymerase	71
2.2.9	Plasmid DNA isolation and purification	72
2.2.10	Golden Gate approach to design, construction and screening of guide RNA expression plasmids for use with Cas9-VPR	72
2.2.11	Assessment of RNA quality by bleach gel	74
2.2.12	cDNA preparation	74
2.2.13	Quantitative Reverse Transcription PCR (qRT-PCR)	75
2.3	Mammalian cell handling	75
2.3.1	Mammalian cell culture	76
2.3.2	Chemical transfection of mammalian cells.....	76
2.3.3	Fluorescence microscopy	77
2.3.4	Flow cytometry.....	77
2.3.5	RNA isolation from CHO-S cells	77
2.4	Techniques relating to <i>in vitro</i> assays.....	78
2.4.1	Protein expression and purification.....	78
2.4.2	Gel shift assays.....	79
3	Design and <i>in vivo</i> testing of a lactate-inducible transgene expression system in CHO cells.....	80
3.1	Chapter aims.....	80
3.2	Chapter summary	80

3.3	Results	80
3.3.1	Summary of experiments performed in this chapter	80
3.3.2	Use of MXS chaining	83
3.3.3	Transient transfections into CHO-S, flow cytometry	84
3.3.4	Transfection of constructs for visualisation of LldR nuclear localisation	87
3.4	Summary of results; discussion	89
4	<i>In vitro</i> optimisation of lactate biosensor components	90
4.1	Chapter aims.....	90
4.2	Chapter summary	90
4.3	Background.....	90
4.3.1	General outline of protein expression in <i>Escherichia coli</i>	91
4.3.1.1	Isolation and cloning of LldR and fusion variant constructs.....	91
4.3.1.2	Codon usage; transformation of LldR constructs into Rosetta2 cells	92
4.4	Results	92
4.4.1	Induction of protein expression; cell lysis; protein solubility testing	93
4.4.1.1	Induction and lysis optimisation of constructs containing LldR.....	94
4.4.2	Protein purification with nickel resin affinity via fast protein liquid chromatography..	97
4.4.3	Densitometry for determination of protein purity and extent of protein-DNA binding	100
4.4.4	Description of the EMSA experiments	101
4.4.4.1	Testing the affinity of LldR (<i>Corynebacterium glutamicum</i>) to two single operator repeats	102
4.4.4.2	Effect of inter-operator spacing on binding affinity.....	105
4.4.4.3	Determination of the impact of effector domain orientation on DNA binding affinity	107
4.4.4.3.1	Variants of the <i>Corynebacterium glutamicum</i> LldR protein	108
4.4.4.3.2	<i>Pseudomonas aeruginosa</i> variants	109
4.4.4.4	Effector domain fusion impact on lactate unbinding capability.	110
4.5	Summary of results; discussion	118
5	Testing of optimised <i>in vivo</i> constructs in CHO cells.....	120
5.1	Chapter aims.....	120
5.2	Chapter summary	120
5.3	Cloning strategy for generation of optimised <i>in vivo</i> constructs in CHO cells.....	120
5.4	Transient transfections of plasmids containing LldR variants and response elements into CHO-S	122
5.4.1	Comparison regarding the impact of operator number	129
5.4.2	Comparison regarding the use of LldR from either <i>Corynebacterium glutamicum</i> or <i>Pseudomonas aeruginosa</i>	133

5.4.3	Comparison regarding the placement of operators upstream or downstream of a promoter.....	135
5.4.4	Comparison regarding the use of the minCMV or YB_TATA minimal promoters ...	136
5.4.5	Comparison regarding co-transfections of two plasmids against the transfection of a single, larger plasmid	137
5.5	Summary and discussion.....	139
6	Use of active Cas9-VPR to upregulate endogenous genes	142
6.1	Chapter aims.....	142
6.2	Chapter summary	142
6.3	Results	142
6.3.1	Constructing guide RNA-expressing plasmids for use with Cas9-VPR	142
6.3.2	Preliminary testing of the Cas9-VPR system in CHO-S.....	145
6.3.3	Targeting genes involved in lactate metabolism with Cas9-VPR.....	146
6.4	Summary of results; discussion	148
7	Discussion and future work	151
7.1	Summary of the results.....	151
7.2	Future work.....	152
8	References	158
9	Appendices.....	194
9.1	All biopharmaceuticals approved for sale by the Europeans Medicines Agency, from 1990 to July 2018.....	194
9.2	Bacterial strains used.....	206
9.3	Constructs used in chapter 3; experimental results from chapter 3.....	207
9.4	Solubility testing of LldR and variants by SDS-PAGE.....	217
9.5	Chromatograms showing purification profiles of LldR and variants	219
9.6	Evaluation of purification fraction content by SDS-PAGE.....	226
9.7	Oligonucleotides used in EMSA experiments	230
9.8	EMSA experiments of LldR and variants with cgl2917	231
9.9	Constructs used in chapter 5	235
9.10	Summary of results from the transfection experiments in section 5.	239
9.11	Constructs used in section 6.....	247
9.12	Table of primer pairs used in RT-qPCR.....	249
9.13	All DNA sequences used	250
9.14	Reprint permissions	256

List of figures

Figure 1.1. Production hosts used in the manufacture of biopharmaceuticals approved by the European Medicines Agency up to July 2018	19
Figure 1.2. Number of biopharmaceutical approvals by the European Medicines Agency up to July 2018, by production host.....	21
Figure 1.3. Variation in lactate accumulation for a single cell line across different production runs (Le et al., 2012).....	30
Figure 1.4. Overview of carbon metabolism in mammalian cells.....	31
Figure 1.5. Alternative carbon sources such as galactose can encourage lactate consumption in the later stages of mammalian cell cultivation (Altamirano et al., 2006).....	33
Figure 1.6. Different glucose feeding strategies can lead to different lactate accumulation outcomes	34
Figure 1.7. The use of host cell engineering to encourage cells to reliably shift to lactate consumption in the later stages of a production run.....	36
Figure 1.8. Transcription regulation by steric hindrance and de-repression.....	51
Figure 1.9. Transcription regulation by steric hindrance and heterochromatin recruitment.....	52
Figure 1.10. Transcription regulation by transactivation	53
Figure 2.1. Depiction of the overall process required for the design, construction and screening of guide RNA-expressing constructs	73
Figure 3.1. The principle of the MXS chaining cloning method (Sladitschek & Neveu, 2015).....	85
Figure 3.2. Flow cytometry gating strategy.....	86
Figure 3.3. Conditions sorted by experiment number.....	87
Figure 3.4. Subcellular localisation of LldR-NLS (fused to GFP/mAzamiGreen) in adherent CHO cells	88
Figure 3.5. Subcellular localisation of LldR (fused to GFP/mAzamiGreen) in adherent CHO cells.....	89
Figure 4.1. SDS-PAGE analysis of the solubility of LldR (<i>Corynebacterium glutamicum</i>) by itself and with the KRAB domain fused to either its N- or C-terminus.....	94
Figure 4.2. Testing the use of lower concentrations of IPTG to boost solubility of LldR (<i>Corynebacterium glutamicum</i>) fused to KRAB with SDS-PAGE.....	96
Figure 4.3. Chromatogram produced during the purification of LldR (<i>Corynebacterium glutamicum</i>)	98
Figure 4.4. SDS-PAGE analysis of the fractions collected during the purification of LldR (<i>Corynebacterium glutamicum</i>).....	99
Figure 4.5. Using densitometry to determine the purity of LldR (<i>Corynebacterium glutamicum</i>)	100
Figure 4.6. Two different operators whose affinity for LldR was tested	103
Figure 4.7. EMSA showing cgl2917 incubated with increasing concentrations of LldR (<i>Corynebacterium glutamicum</i>)	104

Figure 4.8. EMSA showing 1xOperator incubated with increasing concentrations of LldR (<i>Corynebacterium glutamicum</i>).....	104
Figure 4.9. Analysis of the binding of LldR (<i>Corynebacterium glutamicum</i>) to two similar operators, 1xOperator and cgl2917	105
Figure 4.10. Two double operator variants where the spacing between the operators is different ...	106
Figure 4.11. EMSA showing 2xOperatorA incubated with increasing concentrations of LldR (<i>Corynebacterium glutamicum</i>).....	106
Figure 4.12. EMSA showing 2xOperatorB incubated with increasing concentrations of LldR (<i>Corynebacterium glutamicum</i>).....	107
Figure 4.13. Analysis of the binding of LldR (<i>Corynebacterium glutamicum</i>) to 2xOperatorA and 2xOperatorB.....	107
Figure 4.14. Analysis of the binding of LldR (<i>Corynebacterium</i>) and fusion variants to cgl2917.....	109
Figure 4.15. Analysis of the binding of LldR (<i>Pseudomonas aeruginosa</i>) and fusion variants to cgl2917.	110
Figure 4.16. EMSA showing the incubation of cgl2917 with a 4-fold molar excess of LldR and 10-fold molar excess LldR-VP64 (both <i>Corynebacterium glutamicum</i>) and either 0, 5, 20 or 40 mM of lactate	111
Figure 4.17. EMSA showing the incubation of cgl2917 with a 20-fold molar excess of LldR and LldR-VP64 (both from <i>Corynebacterium glutamicum</i>) and either 0, 20, 40 or 80 mM of lactate	112
Figure 4.18. EMSA from Georgi et al. (2008) showing unbinding of LldR (<i>Corynebacterium glutamicum</i>) from a fragment containing the operator sequence in the presence of lactate	113
Figure 4.19. EMSA of LldR, LldR-VP64 and LldR-15aa-VP64 (all from <i>Corynebacterium glutamicum</i>) incubated with 0, 20 and 200 mM of lactate	114
Figure 4.20. EMSA showing the incubation of LldR, LldR-VP64 and LldR-15aa-VP64 (all from <i>Pseudomonas aeruginosa</i>) with cgl2917 in the presence of 0, 20 and 200 mM of lactate.....	115
Figure 4.21. EMSA from Gao et al. (2012) showing the unbinding of LldR (<i>Pseudomonas aeruginosa</i>) from its operator in the presence of lactate.....	116
Figure 4.22. The cgl2917 operator tested in this <i>in vitro</i> optimisation chapter (panel a) and its use in subsequent <i>in vivo</i> experiments (panel b).	117
Figure 5.1. Flow chart showing the overall cloning strategy employed in this section.....	121
Figure 5.2. Conditions sorted by the average fold-change in geometric mean of the GFP signal.....	124
Figure 5.3. The configurations demonstrating the highest levels of induction in response to lactate	128
Figure 5.4. Configurations demonstrating significant levels of induction in multiple independent experiments	129
Figure 5.5. Comparison of configurations with either 1 or 2 upstream operators	131
Figure 5.6. Comparison of configurations with either 1 or 2 downstream operators	132

Figure 5.7. Comparison of configurations with either 1 or 2 operator repeats flanking both sides of a promoter.....	133
Figure 5.8. Comparison of configurations with upstream operators using LldR from either <i>Corynebacterium glutamicum</i> or <i>Pseudomonas aeruginosa</i>	134
Figure 5.9. Comparison of configurations with downstream or flanking operators using LldR from either <i>Corynebacterium glutamicum</i> or <i>Pseudomonas aeruginosa</i>	135
Figure 5.10. Comparison of configurations using either minCMV or YB_TATA	137
Figure 5.11. Comparison of configurations with upstream operators transfected either with separate constructs or on a single construct	138
Figure 5.12. Comparison of configurations with downstream or flanking operators transfected either with separate constructs or on a single construct.....	139
Figure 6.1. Location of guide RNA targets against the minCMV-mAzamiGreen plasmid.....	144
Figure 6.2. Testing the ability of Cas9-VPR to manipulate gene expression in a transient transfection experiment	146
Figure 6.3. Testing the ability of Cas9-VPR to manipulate transcript levels of MCT1	147
Figure 6.4. The Casilio system uses orthogonal pairs of Pumilio RNA-binding domains and cognate RNA octamer sequences to achieve simultaneous transactivation and transrepression at different genomic loci (Cheng et al., 2016)	149
Figure 9.1. Initial solubility testing of LldR, LldR-KRAB and KRAB-LldR (all from <i>Corynebacterium glutamicum</i>) with SDS-PAGE	217
Figure 9.2. Solubility testing of LldR-VP64 and VP64-LldR (all from <i>Corynebacterium glutamicum</i>) with SDS-PAGE	217
Figure 9.3. Solubility testing of LldR, LldR-VP64 and LldR-15aa-VP64 (all from <i>Pseudomonas aeruginosa</i>) with SDS-PAGE	218
Figure 9.4. Chromatogram produced during the purification of LldR-KRAB (<i>Corynebacterium glutamicum</i>)	219
Figure 9.5. Chromatogram produced during the purification of LldR-VP64 (<i>Corynebacterium glutamicum</i>)	220
Figure 9.6. Chromatogram produced during the purification of VP64-LldR (<i>Corynebacterium glutamicum</i>)	221
Figure 9.7. Chromatogram produced during the purification of LldR-15aa-VP64 (<i>Corynebacterium glutamicum</i>)	222
Figure 9.8. Chromatogram produced during the purification of LldR (<i>Pseudomonas aeruginosa</i>)....	223
Figure 9.9. Chromatogram produced during the purification of LldR-VP64 (<i>Pseudomonas aeruginosa</i>)	224
Figure 9.10. Chromatogram produced during the purification of LldR-15aa-VP64 (<i>Pseudomonas aeruginosa</i>).....	225

Figure 9.11. SDS-PAGE analysis of the fractions produced during the purification of LldR (<i>Corynebacterium glutamicum</i>).....	226
Figure 9.12. SDS-PAGE analysis of the fractions produced during the purification of LldR-KRAB (<i>Corynebacterium glutamicum</i>).....	226
Figure 9.13. SDS-PAGE analysis of the fractions produced during the purification of LldR-VP64 (<i>Corynebacterium glutamicum</i>).....	227
Figure 9.14. SDS-PAGE analysis of the fractions produced during the purification of VP64-LldR (<i>Corynebacterium glutamicum</i>).....	227
Figure 9.15. SDS-PAGE analysis of the fractions produced during the purification of LldR-15aa-VP64 (<i>Corynebacterium glutamicum</i>).....	228
Figure 9.16. SDS-PAGE analysis of the fractions produced during the purification of LldR (<i>Pseudomonas aeruginosa</i>).....	228
Figure 9.17. SDS-PAGE analysis of the fractions produced during the purification of LldR-VP64 (<i>Pseudomonas aeruginosa</i>).....	229
Figure 9.18. SDS-PAGE analysis of the fractions produced during the purification of LldR-15aa-VP64 (<i>Pseudomonas aeruginosa</i>).....	229
Figure 9.19. EMSA showing cgl2917 incubated with increasing concentrations of LldR (<i>Corynebacterium glutamicum</i>).....	231
Figure 9.20. EMSA showing cgl2917 incubated with increasing concentrations of LldR-VP64 (<i>Corynebacterium glutamicum</i>).....	231
Figure 9.21. EMSA showing cgl2917 incubated with increasing concentrations of VP64-LldR (<i>Corynebacterium glutamicum</i>).....	232
Figure 9.22. EMSA showing cgl2917 incubated with increasing concentrations of LldR-15aa-VP64 (<i>Corynebacterium glutamicum</i>).....	232
Figure 9.23. EMSA showing cgl2917 incubated with increasing concentrations of LldR (<i>Pseudomonas aeruginosa</i>).....	233
Figure 9.24. EMSA showing cgl2917 incubated with increasing concentrations of LldR-VP64 (<i>Pseudomonas aeruginosa</i>).....	233
Figure 9.25. EMSA showing cgl2917 incubated with increasing concentrations of LldR-15aa-VP64 (<i>Pseudomonas aeruginosa</i>).....	234
Figure 9.26. The impact of lactate on the GFP signal in mAzamiGreen-expressing controls	246

List of tables

Table 1.1. Prominent examples of biopharmaceuticals (Phillippidis, 2018).....	18
Table 1.2. Bacterial transcriptional regulators that have been employed in mammalian inducible transgene expression systems	44
Table 1.3. Dissociation constants of bacterial transcriptional regulators that have previously been used in mammalian inducible transgene expression systems.....	49
Table 2.1. Plasmids used to express various LldR fusion constructs in <i>Escherichia coli</i> , prior to purification.....	59
Table 2.2. Optimised single transient constructs	61
Table 2.3. Optimised double transient constructs, which express both a fusion variant of LldR and a variant of a response element.....	64
Table 3.1. All protein-operator-promoter construct combinations that were initially tested, alongside the configuration type, fold-induction and p-value calculated through the use of a Student's t-test (one-tailed, two-sampled, equal variance) to determine whether the uninduced and induced populations were significantly different.....	82
Table 4.1. Frequency of rare codons found in LldR biosensor components	92
Table 4.2. Estimated purity of LldR fusion variants isolated from <i>Escherichia coli</i>	101
Table 4.3. Summary of fold molar excesses (FME) of LldR variants required to bind 50% of the available cgl2917 operator	118
Table 5.1. Variables employed in the <i>in vivo</i> testing of an optimised lactate-inducible transgene expression system	122
Table 5.2. Detailed summary of results from the transient transfection experiments conducted with optimized <i>in vivo</i> constructs	124
Table 6.1. Description of the guide RNAs designed against the minCMV-mAzamiGreen construct.	143
Table 9.1. Biopharmaceuticals approved for sale by the EMA from 1990 to July 2018	194
Table 9.2. Bacterial strains used for cloning and protein expression	206
Table 9.3. Constructs used in chapter 3	207
Table 9.4. Results from the transient transfections of the initial designs of the lactate-inducible system	211
Table 9.5. Double-stranded oligonucleotides used in EMSA experiments as part of <i>in vitro</i> optimisation	230
Table 9.6. List of constructs designed for <i>in vivo</i> testing of an optimized lactate-inducible transgene expression system	235
Table 9.8. Detailed summary of results from the transfection experiments conducted with optimized <i>in vivo</i> constructs	239
Table 9.9. Constructs used to test the ability of Cas9-VPR to upregulate genes in CHO cells	247
Table 9.10. Primer pairs used to quantify relative transcript levels in CHO.....	249

Abbreviations

15aa	flexible protein linker domain consisting of (GGGGS) ₃
A549	human lung epithelial lung carcinoma cell line
acetyl-CoA	acetyl coenzyme-A
AP-1	activator protein-1
ATP	adenosine triphosphate
bGpA	rabbit β-globin polyadenylation signal
BHK	baby hamster kidney cell line
C2C12	mouse myoblast cell line
CA IX	carbonic anhydrase IX
CD-CHO	chemically defined Chinese hamster ovary cell media
cDNA	complementary deoxyribonucleic acid
CHO	Chinese hamster ovary
CHO-S	suspension Chinese hamster ovary cell line
CMV	cytomegalovirus promoter
COS-7	African green monkey kidney cell line
CT	cycle threshold
DHFR	dihydrofolate reductase
DMSO	dimethylsulphoxide
dNTP	deoxyribonucleoside triphosphates
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
EMA	European Medicines Agency
EMSA	electrophoretic mobility shift assay
Fc	fragment crystallisable region
FME	fold molar excess
FRET	Forster resonance energy transfer
GFP	green fluorescent protein
gRNA	guide ribonucleic acid
HaCat	human skin keratinocyte cell line
HDAC4	histone deacetylase 4

HEK293-T	human embryonic kidney cell line
HeLa	human cervical adenocarcinoma cell line
HepG2	human liver carcinoma cell line
HIF1- α	hypoxia-inducible factor 1- α
hMSC-TERT	human telomerase-immortalised mesenchymal stem cell line
HT-1080	human fibrosarcoma cell line
Huh7	human liver carcinoma cell line
HUVEC	primary human umbilical vein endothelial cell line
IgG	immunoglobulin G
IPTG	isopropyl β -D-1-thiogalactopyranoside
K562	human chronic myelogenous leukemia cell line
kDa	kilodaltons
KRAB	Krüppel associated box domain
LDHA	lactate dehydrogenase A
MCF7	human mammary gland adenocarcinoma cell line
MCT1	monocarboxylate transporter 1
minCMV	minimal cytomegalovirus promoter
MTX	methotrexate
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NDRG3	N-MYC downstream-regulated gene-3 protein
Ni-NTA	nickel-nitrilotriacetic acid
NICE	new <i>in vivo</i> construct experiment
NIH/3T3	mouse fibroblast cell line
NLS	nuclear localisation signal
NS0	non-secreting murine cell line
OCS	one-component system
OD ₆₀₀	optical density at the 600 nm wavelength
P2A	porcine 2A self-cleaving peptide
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PEG	polyethylene glycol
PTM	post-translational modifications
PYC2	pyruvate carboxylase 2
qRT-PCR	quantitative reverse transcription polymerase chain reaction

R	correlation coefficient
RD	human muscle rhabdomyosarcoma cell line
RFP	red fluorescent protein
RNAi	RNA interference
SCB1	<i>Streptomyces coelicolor</i> butyrolactone 1
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SELEX	systematic evolution of ligands by exponential enrichment
sfGFP	superfolding green fluorescent protein
siRNA	small interfering RNA
Sp2/0	mouse spleen cell line
SV40	simian virus 40 polyadenylation signal
TALE	transcription activator-like effector
Tm	melting temperature
TSS	transcription start site
VP64	tetrad of herpes simplex viral protein 16
VPR	tripartite transcriptional activator consisting of VP64, p65 and Rta
$x g$	multiples of the local gravitational constant
ZF	zinc finger

1 Introduction

1.1 Biopharmaceuticals

Biopharmaceuticals have been described as a one of the most exciting breakthroughs of modern medicine (An, 2010), an accolade afforded to them on the basis of their widespread usage to treat a variety of diseases such as cancer, diabetes, heart disease, infertility and arthritis, as well as many others. Biopharmaceuticals encompass a broad range of drugs, including antibodies, hormones, growth factors, blood factors and vaccines (Aggarwal, 2014), with examples shown in table 1.1. They are polymers of peptides or nucleotides (in the case of some vaccines) and are typically made in a biological production host, distinguishing them from the majority of so-called small-molecule drugs which are normally derived from chemical synthesis. Reflecting their value as medicines, biopharmaceuticals account for a large and increasing proportion of global drug sales, with \$217 billion sold in 2016, a 360% increase from 2002 (Lindsay, 2017). Monoclonal antibodies account for approximately 66% of these sales, with the approval of approximately 6 to 9 such drugs expected per year in the near-future (Kesik-Brodacka, 2017). The large sales figures for these drugs reflects not only the number of effective biopharmaceuticals available, but also the high prices that are charged for them in an effort to recoup their development and production costs.

There are numerous factors contributing to the superiority of biopharmaceuticals over small-molecule drugs, stemming from the fundamental fact that as proteins they can interact far more specifically within a cellular system, or they can precisely replace any absent or defective protein (Leader et al., 2008). This specificity of action means that biopharmaceuticals are much less likely to interfere with normal biological processes and so give rise to the multiple side-effects that are typical of many small-molecule drugs. Biopharmaceuticals can also be modified by chemical or genetic means in order to enhance their efficacy. The post-expression covalent attachment of PEG molecules to therapeutic proteins (known as PEGylation) can increase their solubility and clearance times, meaning that the PEGylated interferon alfa-2a will have a 12-fold greater serum half-life compared to a non-PEGylated variant, for example (Lai et al., 2006). Genetic modifications to existing therapeutic proteins can, for instance, replace hydrophobic residues and reduce aggregation (Courtois et al., 2015), or add new glycosylation sites to boost serum half-life (Kiss et al., 2010). Additionally, new

biopharmaceuticals tend to be approved by regulatory bodies for sale more rapidly than small-molecule drugs (Aitken & Kleinrock, 2017, Reichert, 2003).

Table 1.1. Prominent examples of biopharmaceuticals (Phillippidis, 2018).

Brand name	Molecule name	Class	Indication	Sales in 2017	Manufacturer
Humira	adalimumab	monoclonal antibody	rheumtoid arthritis; Crohn's; ulcerative colitis	\$18.4 billion	AbbVie
Rituxan	Rituximab	monoclonal antibody	Non-Hodgkin's lymphoma; chronic lymphocytic leukemia; rheumatoid arthritis	\$9.2 billion	Roche (Genentech) and Biogen
Enbrel	etanercept	fusion of cytokine to Fc region of antibody	multiple varieties of arthritis	\$7.9 billion	Amgen and Pfizer
Herceptin	Roche (Genentech)	monoclonal antibody	breast cancer; gastric/gastroesophageal adenocarcinoma	\$7.4 billion	Roche (Genentech)
Remicade	Infliximab	monoclonal antibody	Crohn's; ulcerative colitis; spondylitis	\$7.2 billion	Johnson & Johnson and Merck & Co
Avastin	bevacizumab	monoclonal antibody	colorectal cancer; lung cancer; glioblastoma	\$7.1 billion	Roche (Genentech)
Lantus	insulin glargine	hormone	diabetes	\$5.7 billion	Sanofi

Prevnar 13	diphtheria CRM197 protein	vaccine	prevention of <i>Streptococcus pneumoniae</i> -induced pneumonia	\$5.6 billion	Pfizer
Rebif	interferon beta-1a	cytokine	multiple sclerosis	\$1.9 billion	Merck
NovoSeven	activated factor VIIa	blood factor	haemophilia	\$1.4 billion	Novo Nordisk

1.2 Expression hosts

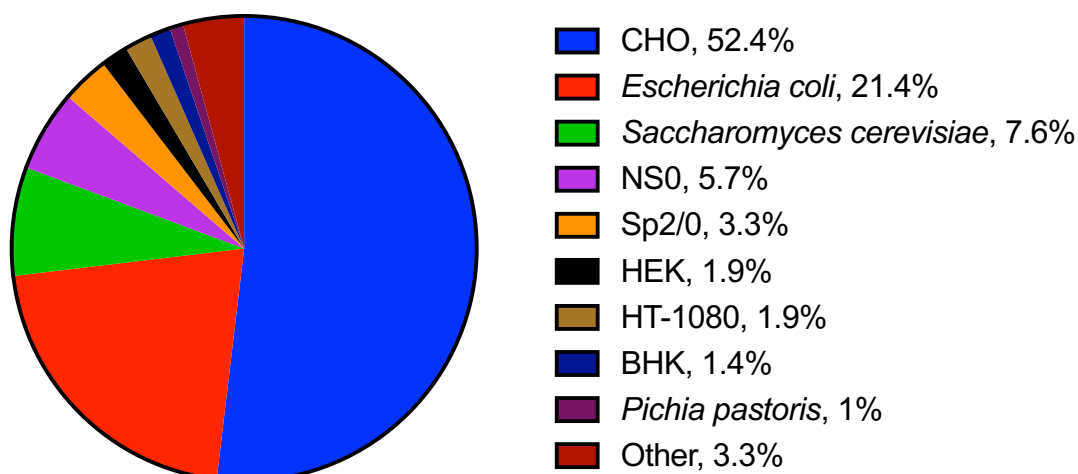


Figure 1.1. Production hosts used in the manufacture of biopharmaceuticals approved by the European Medicines Agency up to July 2018.

The majority of the ~200 approved biopharmaceuticals have been produced in *Escherichia coli*, yeast (primarily *Saccharomyces cerevisiae*), or mammalian cells (primarily Chinese hamster ovary or CHO cells) (Berlec & Strukelj, 2013), each with different strengths and weaknesses. A production host is selected with the complexity of the desired protein in mind - more extensive post-translational modifications (particularly glycosylation) require a eukaryotic host (typically mammalian), while proteins with fewer such modifications can be made more efficiently in *Escherichia coli* or yeast.

Of the 210 protein biopharmaceuticals that have been approved for sale by the European Medicines Agency up to July 2018, 69% were made in a mammalian host, 21% were made in *Escherichia coli*, and 9% were made in a fungal host (see figure 1.1 for a detailed breakdown) (European Medicines Agency, 2018; Walsh, 2005; Walsh, 2006; Walsh, 2010; Walsh, 2014). A full list of these drugs and their production hosts is given in appendix 9.1.

1.2.1 *Escherichia coli*

There are many advantages to using *Escherichia coli* as a production host, if its inability to perform mammalian-like post-translational modifications is not an issue for the protein being expressed. It allows rapid growth to high cell densities with inexpensive media, straightforward transformation using plasmid DNA, and a broad range of choice when it comes to expression strains and the regulatory genetic components used to optimise protein production.

The doubling time of *Escherichia coli* is exceptionally fast, dividing once every 20 minutes under optimal conditions (Sezonov et al., 2007). Indeed, the replication of *Escherichia coli* has been calculated as being near a theoretical maximum efficiency, wasting very little energy as heat (England, 2013). The benefit of this to a bioprocessing run is that a reaction vessel can be inoculated with one hundredth of its volume with a starter culture, and can reach stationary phase within a few hours. Additionally, *Escherichia coli* can reach high cell densities, with a theoretical density limit estimated to be around 200 grams of dry cell weight per litre of medium (Shiloach & Fass, 2005); high production titres of 19.2 grams of protein per litre of media have been seen with dry cell weights of 77 grams per litre, for example (Strandberg & Enfors, 1991).

Scaling up production with *Escherichia coli* is straightforward relative to mammalian cells, as they have a cell wall that makes them resistant to the physical forces introduced by the mixing and sparging necessary at higher volumes (Marks, 2003).

Transformation of *Escherichia coli* can be performed within minutes (Pope & Kent, 1996), and plasmid DNA co-expressing a resistance marker can be maintained indefinitely in the presence of a selection pressure (typically an antibiotic). This contrasts favourably with the weeks necessary to establish highly productive clones in *Pichia pastoris* and the months required for mammalian cells (Kunert et al., 2008), although the use of antibiotics is problematic with regard to product contamination and the emergence of antibiotic-resistant strains of bacteria (Mignon et al., 2015).

Many strains of *Escherichia coli* are available that enhance its capacity to, for example, form disulphide bonds or express toxic proteins (Baeshen et al., 2015).

There are also many well characterised regulatory genetic components that allow optimisation of protein expression, including a range of constitutive and tunable promoters (e.g. Alper et al., 2005; Marschall et al., 2017), with a broad range of possible inducing signals that includes the use of IPTG, mannose and red-light (Striedner et al., 2008; Kelly et al., 2016; Tabor et al., 2011).

Escherichia coli was used to produce insulin, the first biopharmaceutical on sale, and is still used for this purpose today. It is also used to make parathyroid hormone (brand name Natpar), sold as a treatment for hypothyroidism, amongst other drugs. *Escherichia coli* is the production host used to make 45 out of 210 (21%) of biopharmaceuticals approved by the EMA up to July 2018 (figure 1.1).

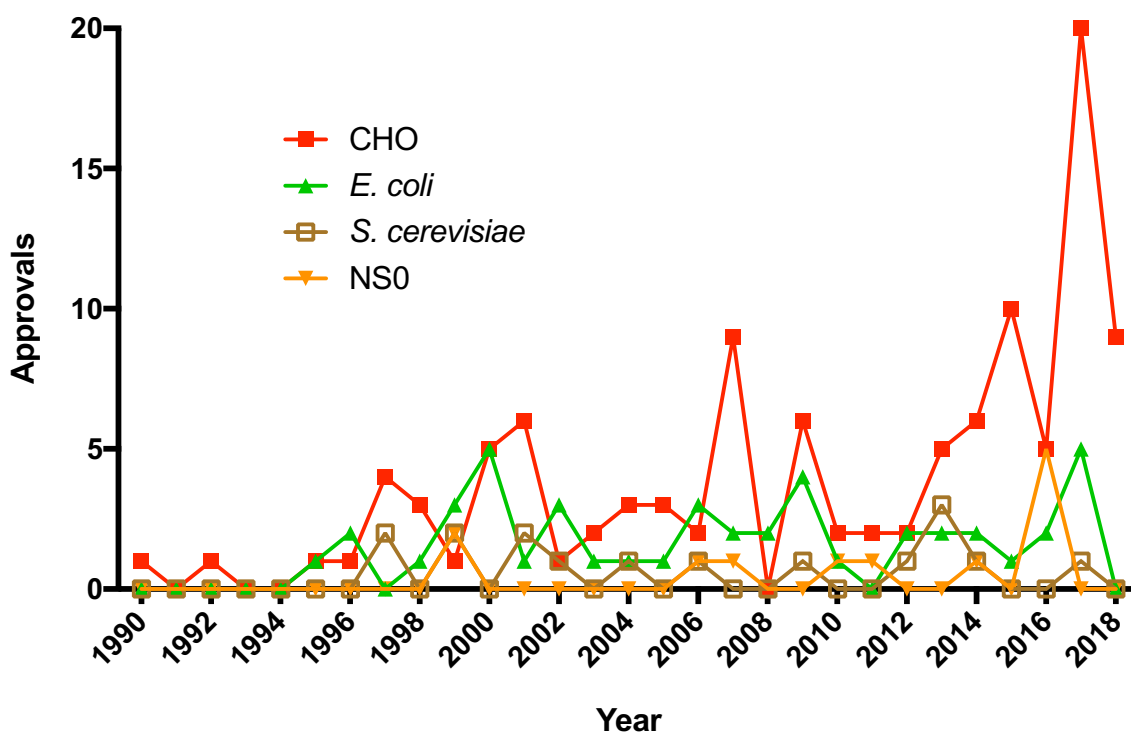


Figure 1.2. Number of biopharmaceutical approvals by the European Medicines Agency up to July 2018, by production host. Only the top four most commonly used hosts are displayed for clarity.

Despite the convenience and productivity of *Escherichia coli*, its use is limited to the production of proteins without significant post-translational modifications (PTM) and as such, it is likely that it will only be used to express a minority of biopharmaceuticals in the near future. In addition, *Escherichia coli* may not always be able to produce a soluble eukaryotic protein, owing to the differences between its native “folding machinery” and that of eukaryotes (Baneyx & Mujacic, 2004), and the endotoxin that *Escherichia coli* incorporates in its outer membrane is both toxic to patients, and difficult to remove during the purification of a desired protein (Mack et al., 2014). Some efforts have been made to generate a strain of *Escherichia coli* capable of producing glycosylated proteins, although the process remains very inefficient and currently only partially resembles a human glycosylation pattern (Jaffé et al., 2014).

1.2.2 Yeast

The two predominant yeast species used for biopharmaceutical production are *Saccharomyces cerevisiae* and *Pichia pastoris*, due to their relatively superior ability to produce complex eukaryotic proteins compared to *Escherichia coli*, as well as their favourable bioprocessing characteristics (e.g. rapid growth rates, high cell densities, secretion of overexpressed protein, etc.). Volumetric productivity for *Saccharomyces cerevisiae* can reach 5 to 10 mg/L/h, compared to 1 to 2 mg/L/h for mammalian cells (Gasser et al., 2007), reflecting the high cell densities that can be achieved during a production run, around 102 g/L dry cell weight for *Saccharomyces cerevisiae* (e.g. Wang et al., 2010), and 160 g/L dry cell weight for *Pichia pastoris* (e.g. Jahic et al., 2002). Protein production yields in *Pichia pastoris* have been reported as high as 14.8 g/L in a clarified broth (Werten et al., 1999), where the use of a secretion signal fused to the protein of interest simplified subsequent efforts at purification. This is in contrast to *Escherichia coli*, where the overexpressed protein is often localised to the periplasm, requiring cell lysis and consequent separation of cellular debris from the desired product (Gasser et al., 2007). Cultivation times are brief, as yeasts have short doubling times (~110 minutes in the case of *Pichia pastoris* (Larsen et al., 2013), and ~90 minutes for *Saccharomyces cerevisiae* (Sherman, 1991)). Yeasts are also known for their ability to consume inexpensive feedstocks. For example, there is media available for *Pichia pastoris* that costs ~\$2 per litre (Pais-Chanfrau & Trujillo-Toledo, 2016) and it can utilise methanol as its sole carbon source, which sells for ~\$0.42 per litre (Methanex, 2018), compared to media costs of ~\$52 per litre that can be used in large-scale mammalian cell culture (Xu et al., 2017).

18 biopharmaceuticals expressed in yeast systems were approved for sale by the EMA throughout its history, 16 of which were made in *Saccharomyces cerevisiae* and 2 in *Pichia pastoris* (see appendix 9.1). 7 of these were insulin or insulin-analogues.

Although yeast-based systems have a strong foothold as an expression system, their use in biopharmaceutical production is ultimately limited as a result of the undesirable glycosylation patterns they impart. *Saccharomyces cerevisiae* is known for producing hyper-mannosylated N-glycans, often exceeding 100 mannose residues (Dean, 1999), occurring to a lesser extent in *Pichia pastoris* (Krainer et al., 2013). Therapeutic proteins bearing these glycans have reduced serum half-life and are therefore less effective as drugs (Walsh, 2010). Multiple research groups and pharmaceutical companies are developing yeasts capable of human-like glycosylation patterns, and have shown progress towards this aim by knocking out key proteins and introducing others (reviewed by Piirainen et al., 2014), although these techniques have not yet been used to produce any currently marketed drugs.

1.2.3 Mammalian cells

Despite the significantly greater difficulty and expense of producing biopharmaceuticals in mammalian systems, the benefit of human-like glycosylation to the therapeutic efficacy and successful expression of these proteins means that mammalian cells are the predominant production hosts in use. Of 210 biopharmaceuticals approved for sale by the EMA, 145 (69%) were made in mammalian systems.

Gonadotropin (sold as Gonal-F) is an example where the glycosylation has been shown to assist proper folding, by facilitating disulphide bond formation (Feng et al., 1995). Glycosylation also enhances protein recognition - in the case of imiglucerase (sold as Cerezyme), its glycan allows it to be taken up by its targets cells more efficiently than otherwise (Furbish et al., 1978). Glycans are also important for the direct activity of therapeutic proteins; gonadotropin, for instance, can no longer transduce a signal without glycosylation at a key site (Ulloa-Aguirre et al., 1999). Additionally, proper sialylation of therapeutic proteins is important for the regulation of their serum half-lives (Kompella & Lee, 1991), which is performed natively by mammalian cell lines but not in yeast (Gemmill & Trimble, 1999). Glycosylation is also important for masking hydrophobic protein residues that would otherwise contribute to aggregation and precipitation, with interferon β -1a (made in CHO and sold as Avonex) and interferon β -1b (made in *Escherichia coli* and sold as Betaferon) illustrating this impact.

Although these drugs differ by only a single protein residue (Zago et al., 2009), Betaferon's lack of glycosylation means that it is highly hydrophobic and has a short serum half-life (Chiang et al., 1993; Khan et al., 1996), and as a result it is 8.7-fold less active than the interferon- β made in CHO cells (Antonetti et al., 2002). Glycosylation can also be necessary for the proper secretion of biopharmaceuticals from their production hosts (Walsh & Jefferis, 2006).

Chinese hamster ovary cells are by far the most commonly used biopharmaceutical production host, accounting for 52% of all such drugs approved by the EMA to date (see figure 1.1 and appendix 9.1).

Chinese hamsters were a source of important cell lines for medical research in the early 20th century, and its ovary cells were first isolated as a cell line in 1958 (Tjio & Puck). Many of the genes of CHO have been silenced, and so in effect it only has single copies of these genes, a consequence of which is that mutagenesis of CHO can produce useful auxotrophs (Simon et al., 1982). In particular, a cell line deficient in dihydrofolate reductase (DHFR) was found (Urlaub & Chasin, 1980). DHFR is responsible for producing tetrahydrofolic acid, which is needed for the production of glycine, hypoxanthine and thymidine. The DHFR-deficient CHO cell line can be co-transfected with plasmids encoding a protein of interest as well as DHFR, allowing cells that are capable of genomic integration of these vectors to be selected by withholding the DHFR-produced molecules. Additionally, amplification of the gene of interest can be carried out by cultivating the surviving cells in increasing concentrations of methotrexate (MTX), which blocks the activity of DHFR and so selects for cells with progressively higher levels of DHFR expression (Kingston et al., 1993). Amplification of DHFR activity correlates with greater expression of the protein of interest, as they will tend to be integrated together even when transfected separately (Chen & Chasin, 1998). It has been argued that the existence of this auxotrophic cell line was a key contributory factor to the industry-wide adoption of CHO as a mammalian biopharmaceutical production host (Jayapal et al., 2007). CHO is also considered a safe production host - many human pathogenic viruses do not survive in CHO (Wiebe et al., 1989), and its ability to grow well in serum-free media has allayed regulatory concerns about the transmission of prion diseases, such as Creutzfeldt-Jakob disease (Jayapal et al., 2007).

Although the human-like glycosylation imparted by mammalian cells is clearly compelling enough to compensate for their shortcomings, these cells do have many weaknesses relative to bioproduction in microbial host systems.

Cell line development in mammalian cells can take up to two orders of magnitude longer than in other systems. As described previously, *Escherichia coli* can be transformed with plasmid DNA in a matter of minutes, and can be used to express large quantities of protein within a matter of days (Pope & Kent, 1996). In a case study, Kunert et al. (2008) compared the expression of an antibody fragment in *Pichia pastoris* and CHO cells - 3 days was sufficient to generate a stably integrated *Pichia* clone, while the first round of selection for CHO took over two weeks, and had to be followed by further rounds of selection in the presence of MTX. Estimates of the duration required for the isolation of highly productive CHO cell lines vary, ranging from at least 5 months (Kunert et al., 2008) to at least a year (Wurm, 2004). Kunert et al. state that highly productive *Pichia* cell lines can be isolated within 2 weeks.

Yields in mammalian cells are typically 10- to 100-fold lower than in microbial systems (Jayapal et al., 2007), as a result of many contributory factors. Mammalian cells have low growth rates - CHO has a 20-hour doubling time, whereas *Saccharomyces cerevisiae* and *Escherichia coli* require 90 and 20 minutes, respectively (Sherman, 2002; Sezonov et al., 2007). CHO cell densities are also rather low, with a wet cell weight of approximately 3.5 grams per litre (Kunert et al., 2008), compared to 130 g/L for *Saccharomyces cerevisiae* (Gasser et al., 2006), 450 g/L in *Pichia pastoris* (Nylen & Chen, 2018), and 51 g/L in *Escherichia coli* (Glazyrina et al., 2010). Given that the volumetric productivity of a given system is typically proportional to the density of biomass generated (Cregg, 2007), CHO cells are at a disadvantage on this basis compared to other systems. The result of this overall lower productivity is that larger culture vessels and longer cultivation periods are required, with a concomitant increase in production costs. Other production issues faced by the use of mammalian cells are considered in more detail in the following sections.

1.2.3.1 Scale up

The relatively low density reached by CHO cells coupled with the high global demand for biopharmaceuticals means that they are often cultivated at very large scales, above 10,000 litres (e.g. Le et al., 2012). Achieving homogeneity of nutrients and properly aerating such large cultures is a difficult problem, as inadequately doing so will result in detrimental concentration gradients of nutrients, oxygen and CO₂, but excessive mixing and aeration can themselves be damaging to mammalian cells (Marks, 2003).

Nutrient gradients can lead to areas of starvation as well as areas with excessive nutrient levels within a bioreactor, and as consumption increases in one area relative to others, oxygen consumption increases and so contributes to the occurrence of oxygen gradients. In general, inadequate mixing will mean that cells are not exposed to the ideal concentrations of nutrient feedstocks. Bylund et al. (1998) indicate that these nutrient gradients can lead to significantly lower recombinant protein yield, relative to performance at smaller scales. The risk of oxygen gradients occurring (either as a result of inefficient aeration, nutrient gradients emerging from inadequate mixing, or a combination of the two) is due to the use of large bioreactors, as the ratio of surface area available for aeration to the overall volume of the bioreactor decreases rapidly as the volume increases. Oxygen limitation is detrimental to cell growth, productivity, and protein glycosylation (Serrato et al., 2004, Lin et al., 1993). CO₂ can also accumulate when mixing is inadequate, and can be detrimental to cell growth as a result of culture acidification or the addition of salts to combat pH decreases (Kimura et al., 1996).

To combat the emergence of the above gradients, mechanical mixing and the introduction of air (sparging) are required. However, both can potentially have their own negative impact on cell productivity, as mammalian cells lack the rigid cell wall of microbes. The impeller blades commonly used to mix cell cultures will dissipate energy in concentrated areas, such that even moderate levels of overall mixing can damage cells near the blades themselves. This is generally only an issue for adherent cells grown on microcarrier beads, as this can cause the cells to detach (Hu et al., 2011). The primary cause of cell damage (for both suspension-adapted cells and microcarrier cell culture) is from sparging, with a linear relationship between the rate at which gas is introduced and the cell death rate (Jöbses et al., 1990). Bubbles that burst at the liquid-air interface of a bioreactor rapidly dissipate energy to the surrounding liquid and damage nearby cells (Walls et al., 2017). These issues mean that mixing and aeration in large scale mammalian cell culture must be carefully balanced to avoid the emergence of concentration gradients or excessive and physically disruptive physical forces.

Additionally, cell line selection strategies can struggle to find the clones that maintain their productivity from the initial small scale to the later large scale, which is thought in part to result from the different environments that two scales present to cell lines (Porter et al., 2010).

1.2.3.2 Productivity loss through genetic instability, promoter silencing and stochastic gene expression

Mammalian cell cultures are known to lose a great deal of their productivity over extended cultivation periods, due to epigenetic silencing of the promoters powering gene expression, the loss of the genes themselves, and phenotypic diversity that can emerge in a genotypically homogeneous cell line. This loss of activity not only diminishes product yield but can also impact protein glycosylation, with the potential to compromise regulatory approval as this can have an impact on the therapeutic activity of the drug (Barnes et al., 2003). This loss of productivity has been seen in the manufacture of numerous therapeutic proteins, including interferon- γ (Cossons et al., 1991) and an IgG (Kim et al., 1998), dropping to 30-40% and 30-80% of their volumetric expression peaks after 120 and 56 days, respectively. Many other examples exist (see Barnes et al., 2003), and it occurs in an unpredictable manner both in the presence or absence of the selection pressure used to establish the cell lines (Kim et al., 2011).

The instability of expression in long-term cell culture is largely caused by gene excision and promoter silencing. Although the mechanisms by which genes are excised in clonal cell lines are not well understood, it is thought to be the primary factor in productivity loss (Kim et al., 2011). CHO is known to have an unstable karyotype (i.e. the number and arrangement of its chromosomes), and may lose gene copies through the complete or partial looping out of transgenes (Heller-Harrison et al., 2009), deletion over time of telomeres which aid chromosomal maintenance, and the loss of entire chromosomes (Jun et al., 2006). Transgenes can also be silenced by promoter methylation, occurring at cytosines within CpG dinucleotides (Osterlehner et al., 2011), which correlates with compacted chromatin that cannot be accessed by the transcriptional machinery (Cedar & Bergman, 2009). The human cytomegalovirus (CMV) promoter, widely used in biopharmaceutical production, is enriched with CpG dinucleotides and is heavily methylated in unstable cell lines (Kim et al., 2011). Once cells within a population are able to excise or silence its transgenes, any growth advantage that results from a reduced expression burden will allow them to gradually occupy a greater proportion of the overall population (Barnes et al., 2003). Pilbrough et al. (2009) state that the phenotypic diversity that emerges within isogenic cell lines as a result of stochastic gene expression (Raj & van Oudenaarden, 2008) can also contribute to apparent expression instability, as the clones identified as being particularly productive (due to stochastic

fluctuations in key characteristics) at early screening stages may return to a more typical, less productive state.

There are various approaches to avoid this loss of productivity. A promoter resistant to CpG methylation can be used, although these are usually weaker than the CMV promoter (Romanova & Noll, 2018). To the extent that MTX amplification of gene expression is what gives rise to instability, the targeted integration of transgenes into transcriptional hotspots has been suggested as a way of ensuring stable albeit reduced levels of expression, without the need for prolonged exposure to selection molecules such as MTX (Barnes et al., 2003). Yoshimoto & Kuroda (2014) demonstrated that clonal cell lines that are prone to stochastic variations in productivity can be screened early on, to eliminate this as a concern in the generation of cell lines.

1.2.3.3 Accumulation of the toxic metabolites lactate and ammonia

The accumulation of lactate and ammonia is widely considered to be detrimental to bioprocess productivity (Lim et al., 2010), resulting in media acidification and a consequent decrease in cell viability and productivity in the case of the former (e.g. Jeong et al., 2001), and a decrease in cell growth, productivity and glycosylation quality in the case of the latter (Andersen & Goochee, 1995; Yang & Butler, 2000). Lactate accumulation occurs as a result of aerobic glycolysis typical of rapidly proliferating cells in the early, exponential growth phase of a bioprocessing run (Zheng et al., 2012). This form of glycolysis occurs even in the presence of adequate oxygen, and the conversion of pyruvate to lactate produces only 4 molecules of ATP compared to the 36 of oxidative phosphorylation, where pyruvate is instead converted to acetyl-CoA (Jones & Bianchi, 2015). The breakdown of glutamine by glutaminolysis also produces lactate (Kreb & Bellamy, 1960). During exponential growth of mammalian cell cultures, they will convert up to 75% of the available glucose to lactate, with less than 10% converted to carbon dioxide via the citric acid cycle (Ahn & Antoniewicz, 2013).

It has been suggested that the preference of rapidly growing cells for the energetically inefficient aerobic glycolysis reflects a rewiring of metabolism in favour of NADPH production, which is produced as a result of lactate generation via glutaminolysis (Vander Heiden et al., 2009). NADPH is a cofactor required for fatty acid and nucleotide biosynthesis (Berg et al., 2006), the demands for which are elevated in rapidly dividing cells. Lactate consumption as cultures shift to the stationary growth phase has been highly correlated with greater final titres

in a study of 243 12,000 litre production runs, while continued accumulation reduced titres (Le et al., 2012).

The switch from lactate accumulation in the early stages of production to lactate consumption has been a topic of great interest, on the basis of its association with higher productivity, as well as the unpredictability of its occurrence (see figure 1.3 - Le et al., 2012). It has been shown to correlate with many different and complementary factors, including glucose and/or glutamine decrease or depletion (Altamirano et al., 2004; Altamirano et al., 2006; Martinez et al., 2013; Ghrobaniaghdam et al., 2014; Wahrheit et al., 2014; Zagari et al., 2013), extracellular lactate concentration beyond a certain threshold (Kyriakopoulos & Kontoravdi, 2014), and reduced glycolytic flux when approaching stationary phase (Mulukutla et al., 2012), amongst others. A review of these studies led to a suggestion that as cells approach the stationary growth phase, the consumption of lactate can restore NADH/NAD⁺ homeostasis while allowing decreased glycolytic flux and increased oxidative phosphorylation (Hartley et al., 2018).

Ammonia accumulates as a byproduct of glutamine metabolism, by glutaminases prior to arginine and pyrimidine biosynthesis (Hashizume et al., 2011) and also through glutamate dehydrogenation. Mammalian cell culture media is supplemented with glutamine that can also naturally decompose into ammonia.

Both lactate and ammonia are detrimental to cell culture performance. Ammonia can reduce growth rates (Kurano et al., 1990), cell productivity (Hansen & Emborg, 1994), and result in suboptimal glycosylation patterns (Borys et al., 1994; Andersen & Goochee, 1995). This is related in part to the net increase in acidity associated with its production (Ozturk et al., 1992) and the energetic burden imposed by the clearance of ammonium freely diffusing across the cell membrane (Chen & Harcum, 2007). The negative impact of lactate comes largely from its acidity. If the pH of a cell culture deviates from a narrow range, it will undergo increased rates of apoptosis (Furlong et al., 1997; Williams et al., 1999). As the pH in a bioreactor is tightly controlled by the addition of an alkali such as sodium carbonate, this can lead to a hyperosmotic environment, itself a cause of increased apoptosis (Kim & Lee, 2002), with the impact of severely reducing viable cell densities (Le et al., 2012) and leading to the premature termination of production runs (e.g. Lee et al., 2013).

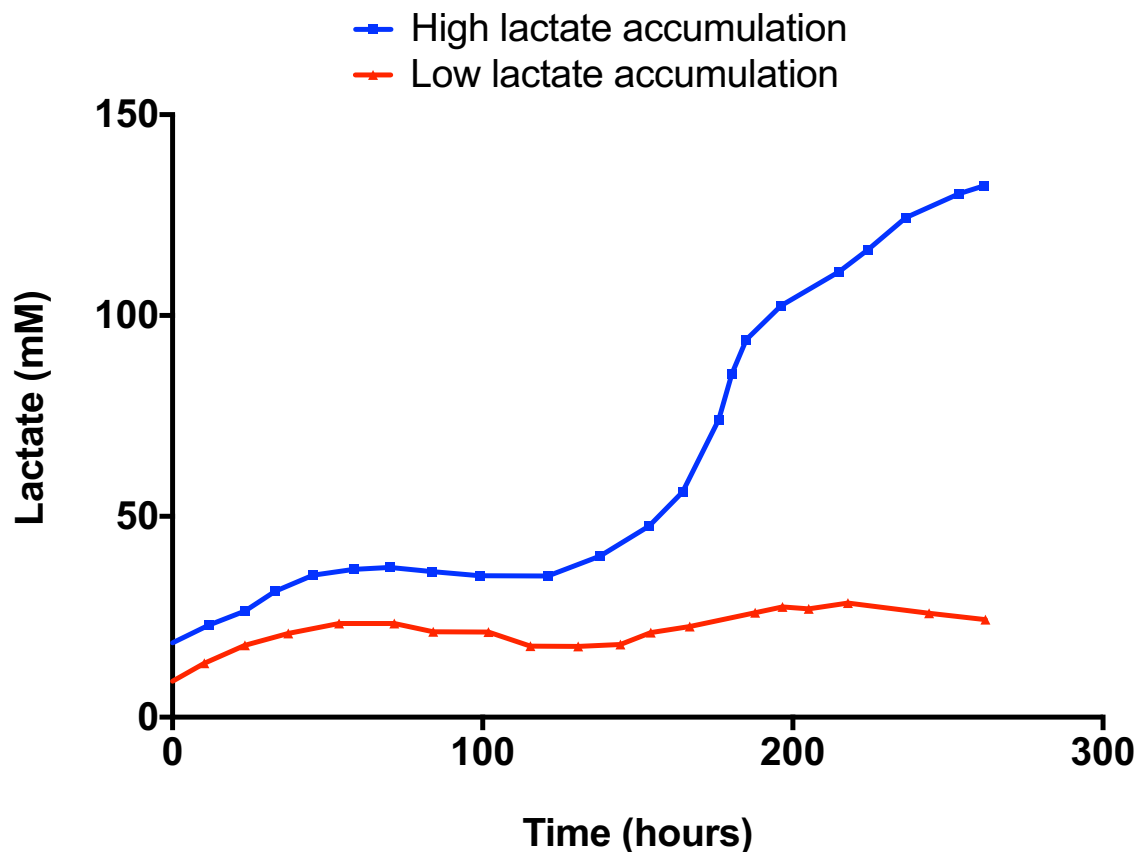


Figure 1.3. Variation in lactate accumulation for a single cell line across different production runs (Le et al., 2012). Two separate production runs for a single cell line at the 12,000 litre scale are displayed, exhibiting high levels of final lactate accumulation associated with lower productivity (shown in blue), and relatively low levels of final lactate accumulation associated with higher productivity (shown in red). Datapoints have been digitized from published figures using the WebPlotDigitizer (Rohatgi, 2016). Image reproduced with permission of the rights holder, Elsevier B.V., detailed in appendix 9.14.

1.2.3.3.1 Lactate metabolism in mammalian cells

Briefly, carbon sources can be metabolised to generate both energy for reactions (largely in the form of ATP) as well as to provide biosynthetic substrates. Glucose is the most commonly used carbon source for mammalian cell culture because it is readily taken up into cells by sodium-glucose linked transporters and glucose transporters, which overcome the otherwise low permeability of carbohydrates across cell membranes in general (Leong et al., 2017). As shown in figure 1.4, glucose is converted by multiple intermediate steps into pyruvate, at which point it can be converted into either acetyl-CoA (the key substrate for the citric acid cycle that

leads to efficient generation of ATP), or pyruvate can be converted into lactate by the enzyme lactate dehydrogenase. The L isomer is the predominant natural form of lactate in mammalian cells, and all references to lactate in this thesis are to this isomer. Although the conversion of pyruvate to lactate generates four ATP molecules, this is significantly fewer than the 36 if pyruvate is allowed to enter the citric acid cycle via acetyl-CoA (Jones & Bianchi, 2015). Glucose can also be metabolised through the pentose phosphate pathway, which generates intermediates used in nucleotide synthesis as well as NADPH, used in fatty acid synthesis.

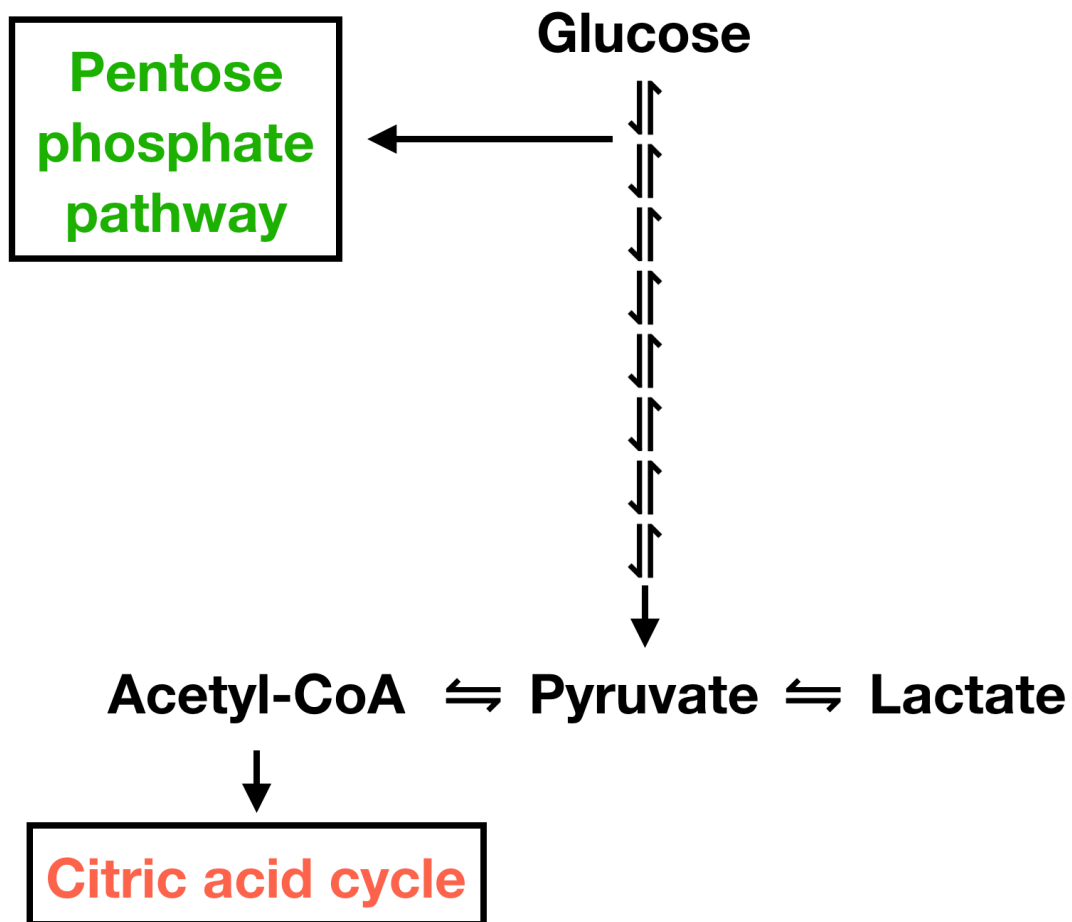


Figure 1.4. Overview of carbon metabolism in mammalian cells. Glucose is converted, via many reactions, into pyruvate. Pyruvate can then enter the citric acid cycle as acetyl-CoA to produce the substrates necessary for oxidative phosphorylation, or it can be converted into lactate by lactate dehydrogenase. Glucose can also flow into the pentose phosphate pathway, where it is used to generate intermediates for nucleotide and fatty acid synthesis. The arrows denote the directions of irreversible (\rightarrow) and reversible (\rightleftharpoons) conversions between metabolites. Intermediate metabolites are not shown for clarity, nor are the intracellular compartments where the reactions take place.

1.3 Combatting lactate accumulation

1.3.1 Growth media formulation

A straightforward approach to combatting lactate accumulation is through medium formulation. As the rapid conversion of glucose is a significant factor in lactate accumulation, multiple studies have looked at replacing or supplementing glucose with more slowly metabolised carbon sources. This includes the use of galactose (Altamirano et al., 2006), fructose (Le et al., 2013), mannose (Berrios et al., 2011) and maltose (Xu et al., 2016). Figure 1.5 shows lactate accumulation in the presence of 20 mM glucose, and if 20 mM galactose and 5 mM glucose was used instead, a shift to lactate consumption later in the culture process was seen, while maintaining high levels of cell growth otherwise not seen with galactose as the sole carbon source (Altamirano et al., 2006).

Xu et al. (2016) found that the addition of copper sulphate to media was able to encourage lactate consumption, possibly on the basis that as it is a cofactor of cytochrome c oxidase (which generates a proton gradient prior to ATP synthesis), it encourages the consumption of pyruvate via oxidative phosphorylation instead of lactate formation. Alongside other groups (e.g. Yuk et al., 2015), Xu et al. found that the use of copper supplementation led to increased proportions of aggregation-prone basic protein variants, which arise largely through the degradation of protein residues (Khawli et al., 2010).

A general issue with the replacement of carbon sources is that it may have an undesirable impact on the glycosylation pattern of the proteins (e.g. Xie et al., 1997; Liu et al., 2014; Chee Fung Wong et al., 2005; Xu et al., 2016), although this is not true in all cases (e.g. Berrios et al., 2011). For instance, Liu et al. (2014) found that glucose concentrations maintained below 15 mM (as can be typical of the previously described strategies) led to reduced site occupancy (the number of possible glycosylation sites that are occupied by glycans) and decreased galactosylation of antibodies, which can adversely affect their ability to induce antibody-dependent cell-mediated cytotoxicity (Houde et al., 2010), a mechanism of action for certain therapeutic glycoproteins such as rituximab/Rituxan.

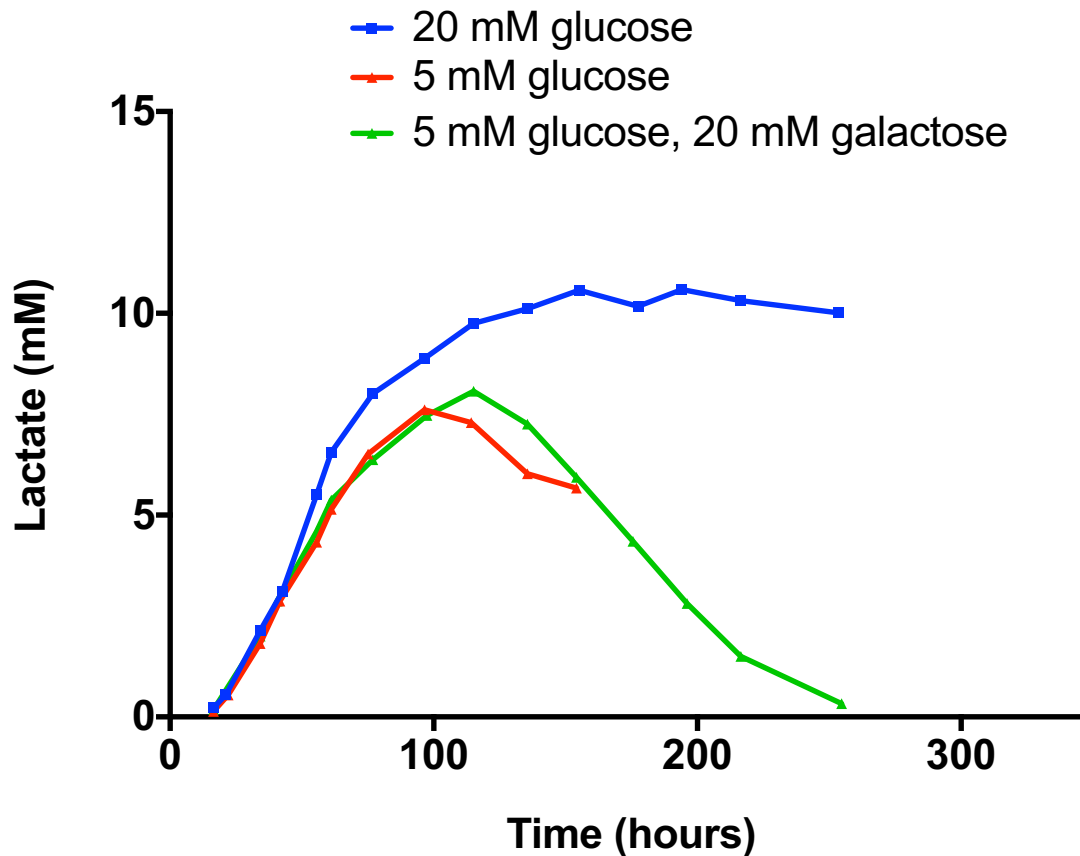


Figure 1.5. Alternative carbon sources such as galactose can encourage lactate consumption in the later stages of mammalian cell cultivation (Altamirano et al., 2006). Datapoints have been digitised from published figures using the WebPlotDigitizer (Rohatgi, 2016). Image reproduced with permission of the rights holder, Elsevier B.V., detailed in appendix 9.14.

1.3.2 Feeding strategies and metabolite monitoring

Lactate accumulation can be controlled with dynamic feeding strategies, if accurate metabolite data is available. For instance, Gagnon et al. (2011) determined that by only feeding glucose above a particular pH value, they could keep lactate concentrations much lower than if glucose was added continuously. A pH setpoint was used on the basis that any increase in acidity is likely due to the conversion of glucose to lactate, and so by stopping the addition of glucose at low pH, excess glucose and lactate would have to be consumed prior to the addition of more glucose. This had the result that lactate accumulation was prevented (see figure 1.6) and productivity was significantly increased, without having a significant impact on protein glycosylation.

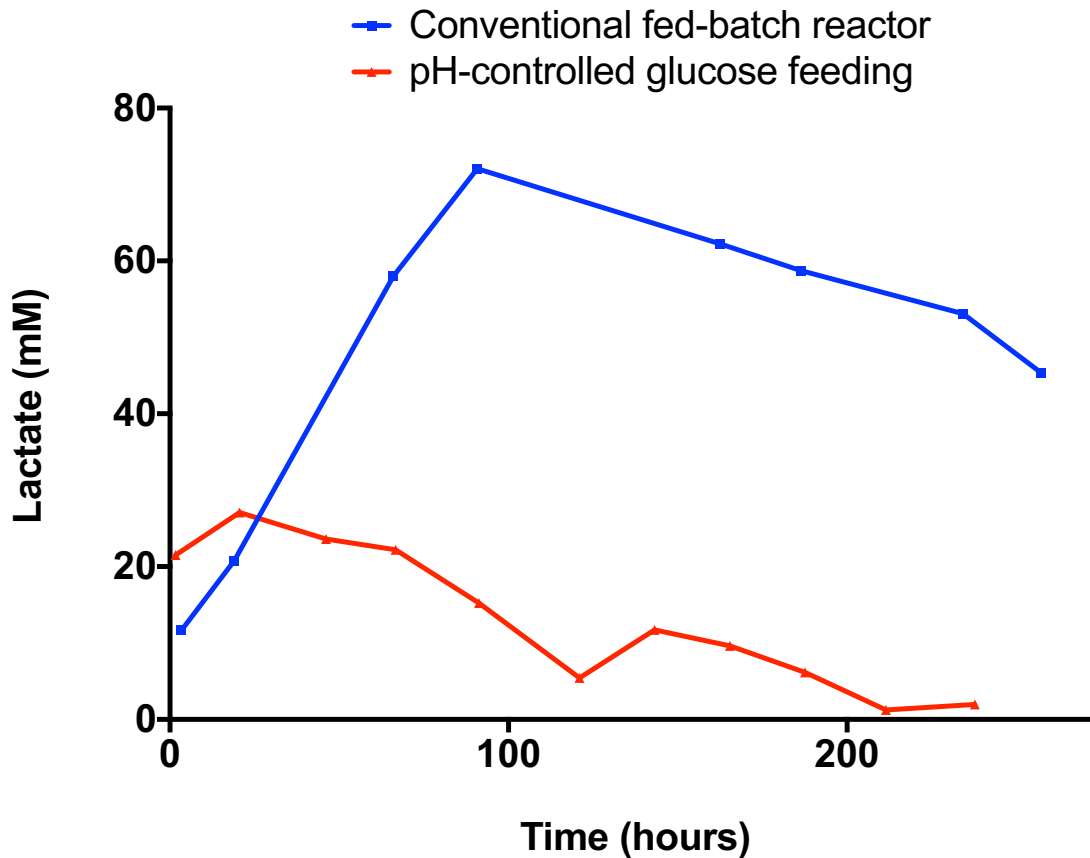


Figure 1.6. Different glucose feeding strategies can lead to different lactate accumulation outcomes. For example, Gagnon et al. (2011) show that the addition of glucose to a culture only when the pH rises above a particular threshold can maintain lactate at low levels, compared to a typical fed-batch production run where glucose concentrations will be higher. Datapoints have been digitised from published figures using the WebPlotDigitizer (Rohatgi, 2016). Image reproduced with permission of the rights holder, Elsevier B.V., detailed in appendix 9.14.

Matthews et al. (2016) were also able to control the accumulation of lactate through dynamic feeding, using Raman spectroscopy to maintain a low concentration of glucose. In so doing, they were able to prevent runaway lactate accumulation as well as boost cell viability, cell density and overall productivity.

Even temporary glucose starvation can have adverse effects on protein glycosylation, which is a risk factor in feeding strategies where the provision of glucose is limited (Toussaint et al., 2016), with Gagnon et al. (2011) acknowledging that although this was not the case for their tested protein, it could be true for others. Chee Fung Wong et al. (2005), for instance, showed

that a dynamic feeding strategy maintaining glucose at low levels led to suboptimal glycosylation patterns.

1.3.3 Host cell engineering

Host cell engineering has been defined by Fischer et al. (2015) as improvement of the “performance of CHO manufacturing cell lines by genetic engineering”. This can entail the upregulation, downregulation or knocking out of endogenous genes, or the introduction of heterologous ones. In general, host cell engineering addresses lactate accumulation in one of two ways - preventing the accumulation of lactate (by encouraging the metabolism of pyruvate to acetyl-CoA or downregulating lactate dehydrogenase, for example) or bolstering the ability of cells to cope with the harmful effects of lactate (typically by manipulating genes related to apoptosis). In an example of the former approach, the introduction of a yeast pyruvate carboxylase to the CHO genome increased the conversion rate of pyruvate to oxaloacetate, which can then enter the citric acid cycle. This promoted lactate consumption in the stationary growth phase (see figure 1.7), boosted viable cell density, and significantly improved overall productivity (Toussaint et al., 2016). Malate dehydrogenase II, which is responsible for converting malate to oxaloacetate, has been overexpressed by Chong et al. (2010) in order to increase flux through the citric acid cycle, resulting in reduced rates of lactate production and improved cell viability. Jeon et al. (2011) found similar benefits when engineering apoptosis resistance through the downregulation of lactate dehydrogenase (to limit the lactate produced from glucose) and the overexpression of Bcl-2, an anti-apoptotic protein that binds to and inhibits pro-apoptotic proteins (Hardwick & Soane, 2013). In another project, Cost et al (2010) knocked out the pro-apoptotic genes Bax and Bak, allowing the cells to better weather the impact of high lactate concentrations and leading to significantly greater densities of viable cells. Many other host cell engineering projects have been carried out and have been reviewed recently (Fischer et al., 2015, Young, 2013).

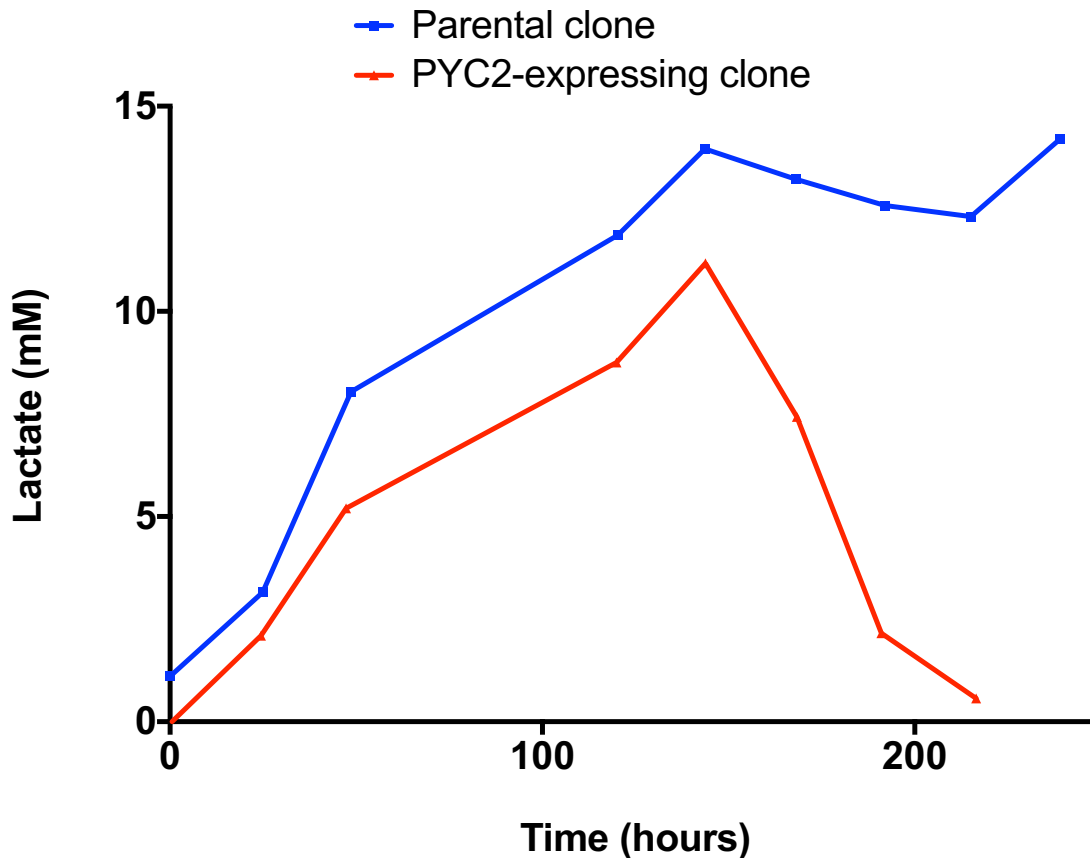


Figure 1.7. The use of host cell engineering to encourage cells to reliably shift to lactate consumption in the later stages of a production run. The example shown here is from Toussaint et al. (2016), who introduced a constitutively expressed pyruvate carboxylase to compete with lactate dehydrogenase for pyruvate, boosting pyruvate's conversion to oxaloacetate and subsequent entry into the citric acid cycle. This resulted in lactate consumption during the stationary phase of the production run. Datapoints have been digitised from published figures using the WebPlotDigitizer (Rohatgi, 2016). Image reproduced with permission of the rights holder, Elsevier B.V., detailed in appendix 9.14.

There are other examples where it has been useful to target multiple genes simultaneously. Zhou et al. (2011) used small inhibitory RNAs to downregulate lactate dehydrogenase A (which catalyses the interconversion of pyruvate and lactate) and pyruvate dehydrogenase kinases (which are responsible for phosphorylating and thereby suppressing the activity of pyruvate dehydrogenase, which itself catalyses the conversion of pyruvate to acetyl-CoA prior to entry into the citric acid cycle). In so doing, lactate accumulation was ~86% lower than the parental cell line, and the product titre was ~125% higher. Reducing lactate dehydrogenase A expression by itself does not always limit eventual levels of lactate accumulation (e.g. Kim & Lee, 2007), nor does knocking down pyruvate dehydrogenase kinases (Zhou et al., 2011).

Jeong et al. (2006) down-regulated lactate dehydrogenase A to limit the amount of lactate generated from pyruvate, and over-expressed glycerol-3-phosphate dehydrogenase to divert glycolytic intermediates to fatty acid synthesis, boosting the production titre of tissue-plasminogen activator and cell growth; these manipulations only had a small impact on productivity in isolation. Jeon et al. (2011) carried out work down-regulating lactate dehydrogenase A and up-regulating the expression of Bcl2 (an anti-apoptotic protein able to delay the onset of apoptotic death in CHO cells), which reduced the final levels of lactate accumulation and boosted viability. Again, productivity was most enhanced when both of these manipulations were performed simultaneously.

Host cell engineering is a useful complement to the previously described strategies, avoiding glycosylation issues associated with low glucose concentrations (e.g. Liu et al., 2014; Chee Fung Wong et al., 2005). However, the constitutive genetic manipulation employed in these projects may not be optimal in all cases - for example, Jeon et al. (2011) saw a reduced cell growth rate in their engineered strain relative to a negative control, indicating that the over-expression of Bcl-2 in particular may be detrimental in the growth phase. Additionally, identifying promising endogenous genes for targeting can be a laborious process, given the difficulties in cloning the appropriate constructs.

1.3.3.1 Constitutive and dynamic approaches to host cell engineering

Genetic engineering strategies such as those previously described can be enacted in response to a particular stimulus, in an approach known as dynamic metabolic engineering. In contrast, it is usually the case that host cell engineering approaches in mammalian cells simply constitutively express or downregulate whatever target of interest, in a fashion that may be suboptimal in some cases. In theory, dynamic metabolic engineering can allow cells to respond appropriately to their own needs, which will differ not only over time (e.g. lactate tends only to reach detrimental levels as cells approach the stationary growth phase) but which also differ as a result of heterogeneity within a bioreactor (e.g. nutrient and O₂ gradients).

Dynamic metabolic engineering is most commonly employed in bacterial and fungal production systems (Brockman & Prather, 2015). In *Escherichia coli*, Anesiadis et al. (2008) controlled the expression of key enzymes in response to cell density with improvements in yield matching the predictions of their computational model. Scalcinati et al. (2012) coupled a squalene synthase to a glucose-responsive promoter in *Saccharomyces cerevisiae*, such that

its expression was highest during exponential growth, contributing to an increase in titre. Fewer examples of dynamic metabolic engineering have been seen in mammalian cells, possibly because they typically employ signal transduction pathways of bacterial origin that require modification prior to use in mammalian systems. Weber et al. (2006) used an NADH responsive protein from *Streptomyces coelicolor* alongside its cognate promoter to report on the redox state of CHO cells, giving a signal that correlated with hypoxia, although the system was not used in this instance to attempt to improve productivity. A later effort from the same lab used a quorum-sensing system from *Aspergillus nidulans* to induce protein expression at higher cell densities (Weber et al., 2007). Kemmer et al. (2010) used a uric acid sensing protein from *Deinococcus radiodurans* to dynamically express a urate oxidase protein, limiting the accumulation of uric acid in mice (that would otherwise develop gout) to healthy levels. Le et al. (2013) used transcriptomic data to identify growth phase-dependent promoters, coupling one to the expression of a fructose transporter to encourage a shift from glucose to fructose consumption in the stationary growth phase. This resulted in greater lactate consumption, a higher growth rate, and a higher product titre.

1.3.3.1.1 Lactate-inducible host cell engineering

One way of combating lactate accumulation and its effects could involve lactate-dependent regulation of multiple genes. For example, the expression of anti-apoptotic genes and inhibition/knockout of pro-apoptotic genes have been shown as useful in mitigating the impact of lactate accumulation and boosting productivity (e.g. Wong et al., 2006; Lim et al, 2006; Cost et al., 2010). Lim et al. demonstrated this by downregulating the pro-apoptotic Bax and Bak proteins via RNAi. However, downregulating these two genes constitutively did not give rise to any substantial improvements in viable cell density or cell viability in the initial growth period, and this is also the case in other host cell engineering projects where genes relating to apoptosis are targeted (e.g. Yun et al., 2007, Wong et al., 2006). On the basis that these targeting these apoptosis-related genes does not confer any clear benefit in the early stages of a production run (and overexpression of certain anti-apoptotic proteins can actually reduce cell growth rates, e.g. Nivitchanyong et al., 2007), it may be a more efficient use of cellular resources to manipulate them in the later stages of production when apoptosis is occurring, and which is often as a result of lactate accumulation.

Additionally, it may be possible to induce a shift to lactate consumption in the later stages of production by altering the balance of lactate dehydrogenases present in response to levels of

lactate. The lactate dehydrogenase C (LDHC) isozyme is unique in that it is normally only expressed in the testes (Fagerberg et al., 2014), it preferentially converts lactate to pyruvate, in contrast to LDHA and LDHB (Koslowski et al., 2002), its upregulation in the late phase of fed-batch production can be accompanied by lactate consumption (Szperalski et al., 2011), and when constitutively expressed, it can lower lactate production and improve viable cell density (Fu et al., 2016). As described previously, LDHA has also been a target for downregulation, and is usually successful in limiting the rate of lactate accumulation (Kim & Lee, 2007, Jeong et al., 2001). However, for both LDHA and LDHC, constitutively manipulating these genes in isolation will still give rise to the same levels of lactate accumulation, albeit usually at a slower rate. It might be the case that simultaneously and constitutively overexpressing LDHC and downregulating LDHA could lead to lower levels of lactate accumulation or consistent lactate consumption; it may be optimal to dynamically regulate this with relation to lactate concentrations, so that normal LDHA expression is maintained at lower levels of lactate and repressed at higher levels, when LDHC could be expressed to encourage the conversion of lactate to pyruvate. This is supported by the hypothesis that the production of lactate can support rapid cell growth (Vander Heiden et al., 2009), which is preferable in the early stages of a production run.

1.3.3.2 Manipulation of endogenous gene expression as a host cell engineering strategy

One general approach to host cell engineering is the manipulation of endogenous gene expression, either by the transient introduction of endonucleases for gene deactivation or constitutive expression of siRNA/miRNA for gene downregulation. RNAi-mediated gene silencing via siRNA employs an siRNA sequence complementary to the mRNA of a target gene, allowing cleavage of the mRNA by the Argonaute protein to reduce transcript levels (Hammond et al., 2001). This can be readily engineered, as the design rules for siRNAs that will downregulate transcript levels in excess of 80% have been well characterised (e.g. Amarzguioui & Prydz, 2004). However, siRNA can suffer from significant off-target effects (Jackson et al., 2006), and this can produce strong phenotypes that are not related to the gene being targeted (Franceschini et al., 2014). Another mechanism of RNAi-mediated gene silencing is via miRNA, which works by binding to mRNA to which it has partial complementary, attracting (amongst other mechanisms) deadenylation factors that remove the poly(A) tail of mRNA and increase their susceptibility to degradation (Pasquinelli, 2012).

One successful example of miRNA engineering in CHO cells is the over-expression of a sequence complementary to the miRNA known as miR-23, thus boosting the expression of the various metabolic genes that miR-23 targets and increasing oxidative phosphorylation (Kelly et al., 2015). However, they found that their approach significantly upregulated 13 genes of wide-ranging function, and in addition to the fact that miRNAs will interact with 300 targets on average (Friedman et al., 2009), it is thus not clear how the miR-23 sponge achieves its increase in productivity. miRNA-based approaches are therefore difficult to engineer towards a specific outcome.

The DNA-binding proteins that have commonly been used in genome engineering approaches have major practical limitations, which are also indicative of why they are rarely used as transcriptional activators or repressors. Zinc finger (ZF) proteins are able to bind 3 specific bases of DNA (Pavletich & Pabo, 1993), with various ZFs of different specificities arrayed together in order to bind to longer DNA sequences. Although there are putative rules of DNA sequence recognition by particular ZFs, these are unpredictably influenced by surrounding ZF domains such that multiple rounds of selection for DNA binding specificity is required for targeting a given sequence (Greisman & Pabo, 1997), with modular assembly techniques resulting in high failure rates if no selection is applied (Ramirez et al., 2008). Characterised more recently are the transcription activator-like effector (TALE) proteins, which are DNA-binding proteins with up to 33 repeat modules, each of which is identical apart from a pair of variable residues within each module that confers DNA-binding specificity (Moscou & Bogdanove, 2009). Although they do have clear rules determining DNA sequence recognition, cloning a TALE protein can be arduous as a result of the multiple near-identical repeat sections, with one approach achieving an efficiency of 5-20% (Gogolok et al., 2016); another commonly employed cloning approach requires the use of a library of 68 repeat modules and two rounds of cloning for each construct (Weber et al., 2011).

1.3.3.2.1 Cas9-VPR as a convenient and versatile genome targeting tool

Cas9 is an RNA-guided endonuclease protein with an ability to conveniently cleave at a specific DNA sequence on the basis of a complementary, co-expressed guide RNA (Jinek et al., 2012). This ability to be directed to new DNA sequences without a lengthy process of cloning and screening has led to the widespread usage of Cas9 in various genome editing projects (Barrangou & Doudna, 2016).

Cas9 can also be fused to particular effector domains in order to engineer diverse genomic interactions. Kiani et al. 2015 showed that an endonucleolytically active Cas9 fused to a transcriptional activator (VPR, a fusion of VP64, p65 and Rta, shown to be more effective at transcriptional upregulation than any of its constituent parts in isolation - Chavez et al., 2015) can be used to upregulate transcription, downregulate transcription, and cleave DNA. If guide RNAs 20 nucleotides in length are expressed, they are able to bind and cleave a target. If a guide RNA of 16 nucleotides or fewer is used, this will no longer be able to cleave DNA, allowing the co-localised VPR to upregulate gene expression if targeted upstream of the transcription start site. However, the guide RNA can be positioned in such a way on the promoter to repress transcription, typically requiring that it is just downstream of a TATA box, if present. Thus, Cas9-VPR can potentially be used to upregulate, downregulate, and excise genes, theoretically allowing it to be used to manipulate the CHO genome in any of the ways previously described in this section.

Constructing plasmids expressing guide RNAs is rapid, requiring a single round of Golden Gate cloning, and allowing straightforward screening of positively transformed colonies - thus, testing the function of a guide RNA in combination with Cas9-VPR can be done quickly. By contrast, constructing a plasmid overexpressing an endogenous protein of interest requires the isolation of the relevant cDNA from a library or de novo synthesis, which can be laborious and expensive, possibly requiring thorough sequence verification and multiple rounds of cloning. Using ZFs or TALEs as activators or repressors of endogenous genes is similarly difficult to clone. This also hinders the ability of host cell engineering projects to target multiple genes, as to do so with multiple proteins requires a great deal of work, and may account for why it is seldom reported. By contrast, Golden Gate assembly methods exist for rapidly cloning multiple guide RNAs into a single plasmid (e.g. Sakuma et al., 2014).

The use of Cas9 to up- or downregulate genes regulated to lactate metabolism in Chinese hamster ovary cells had not been demonstrated at the start of this project, and presents a new approach for conveniently testing the manipulation of specific genes with regards to productivity characteristics, as well as being a platform for the integration of multiple previously characterised manipulations.

1.4 Biosensor design and testing for use in inducible transgene expression systems

1.4.1 What is a biosensor?

A biosensor can be defined as anything that is at least partially biological that is able to detect a molecule or condition of interest (Goers et al., 2013). They can be divided into systems that are genetically encoded, and those in which the detection is performed by an enzyme linked to a physical or chemical output (Dekker & Polizzi, 2017). The latter variety predominate in bioprocess monitoring (Dekker & Polizzi, 2017). Glucose and lactate, for example, can be detected through the use of immobilised glucose and lactate oxidases, in a system where the substrate of interest reacts with the enzyme, giving rise to electron exchange that can be detected at an electrode (Dzyadevych et al., 2008). The merits of such a system are that it can allow on-line measurement of a parameter continuously, frequently, and with high sensitivity (Dekker & Polizzi, 2017), which is of great relevance to bioprocessing. However, one drawback to this kind of system is that it cannot be linked directly to a cellular response at the transcriptional level, which can be a powerful mechanism for boosting process productivity (Brockman & Prather, 2015).

Genetically encoded biosensors comprise of transcriptional regulators (also known as transcriptional switches), post-transcriptional switches relying on RNA aptamers (Pfeiffer & Mayer, 2016), and Förster resonance energy transfer (FRET)-based sensors (e.g. San Martín et al., 2013). Additionally, these biosensors can be contained within the production host or can be within a bacterial chassis, where, in a bioprocessing context, culture media can be diverted for analysis by a whole cell (e.g. Goers et al., 2017). An advantage of genetically encoded biosensors is that they can have transcriptional outputs within the production host, allowing them not only to monitor a process of interest but to be used in host cell engineering projects.

1.4.2 One-component systems and their use as mammalian transgene expression systems

One-component systems (OCSs) are the predominant form of small molecule sensing in prokaryotes, and employ a simple signal transduction mechanism. They use a transcriptional regulator protein that is able to bind to a specific sequence of DNA (known as an operator),

but typically only in the absence of an inducing molecule. When the inducing molecule is present, the interaction of the transcriptional regulator with DNA is allosterically inhibited, causing it to unbind from its operator, allowing transcription to occur from the previously repressed promoter (Ulrich et al., 2005). The prototypical example of a one-component system is TetR, which binds to its operator tetO in the absence of its inducing molecule tetracycline, abolishing the affinity of TetR for tetO (Orth et al., 2000). TetR-based transgene expression systems are commonly used in mammalian cells, on the basis of its high inducibility and the orthogonality of the inducing ligand (Das et al., 2016), and one-component systems predominate as the source of transcriptional switch biosensors in mammalian cells (Kis et al., 2015).

This predominance is owed to the great number of possible inducing ligands (analysis by Ulrich et al., 2005, detected 17,000 such systems in the 145 prokaryotic genomes available at that time) and the simplicity of optimising and modifying a system with so few core components, opening them up for use in many applications. For instance, the inducibility of the system can be optimised by varying the number and placement of operators adjacent to a promoter (e.g. Gossen & Bujard, 1992; Merulla & van der Meer, 2016; Malphettes et al., 2005); the operator itself can be changed to modulate the affinity of the transcriptional regulator for it (e.g. Ike et al., 2015; Krell et al., 2007); ligand-dependent DNA binding can be reversed via protein engineering so that binding of the transcriptional regulator to its operator only occurs in the presence of the ligand (e.g. Gossen et al., 1995), which can be linked to the original regulatory system to improve inducibility (e.g. Mullick et al., 2006); the ligand binding pocket of a transcriptional regulator has in at least one case been modified to respond to a new ligand (de los Santos et al., 2015); they have also been used successfully as part of complex regulatory networks in mammalian cells, for example as a band-pass system where activity was only seen for an input within a certain range (Greber & Fussenegger, 2010) and as a variety of logic gates (Kramer et al., 2004).

To date, 29 different one-component systems have been used as transcriptional biosensors in mammalian cells (see table 1.2), and are somewhat diverse in terms of their protein family of origin. 59% are taken from the TetR family of transcriptional repressors, despite their accounting for approximately 10% of one-component systems (Ulrich et al., 2005). 24% have DNA binding domains from the Winged-helix superfamily, which includes VanR from the GntR family (Gitzinger et al., 2012), from which the lactate-dependent transcriptional repressor LldR

is also drawn. LldR was subsequently selected as a core component of the lactate-inducible transgene expression system designed in this thesis.

Table 1.2. Bacterial transcriptional regulators that have been employed in mammalian inducible transgene expression systems.

Name	Organism	Ligand	Mammalian cell line(s) used	Reference
AmtR	<i>Corynebacterium glutamicum</i>	Gln K protein	HEK293-T, CHO-K1	Stanton et al., 2014
ArgR	<i>Chlamydia pneumoniae</i>	L-arginine	CHO-K1, HT-1080, COS-7, HEK293-T, NIH/3T3	Hartenbach et al., 2007
Bacterial NADH-responsive DNA-binding protein (REX)	<i>Streptomyces coelicolor</i>	NADH	CHO-K1	Weber et al., 2006
BirA	<i>Escherichia coli</i>	Biotin	CHO-K1, HEK293-T, NIH/3T3, COS-7, HeLa, MCF7	Weber et al., 2009
BM3R1	<i>Bacillus megaterium</i>	Pentobarbital	HEK293-T, CHO-K1	Stanton et al., 2014
CbaR	<i>Comamonas testosteroni BR60</i>	benzoic acids	HEK293-T, HeLa, hMSC	Xie et al., 2014
CmeR	<i>Campylobacter jejuni</i>	bile acids including cholic acid	HEK293-T, BHK-21, COS-7, HeLa, HT-1080, CHO-K1	Rössger et al., 2014

CymR	<i>Pseudomonas putida</i>	Cumate	HEK293-T, HeLa, A549	Mullick et al., 2006
EthR	<i>Mycobacterium tuberculosis</i>	2-Phenylethylbutyrate	HEK293-T	Weber et al., 2008
GyrB	<i>Escherichia coli</i>	Coumermycin	HEK293-T, K562, HEK293-T	Zhao et al. 2013
HdnoR	<i>Arthrobacter nicotinovorans</i>	6-hydroxynicotine	CHO-K1, HEK293-T	Malphettes et al., 2005
HucR	<i>Deinococcus radiodurans</i>	uric acid	HeLa, HEK293-T, HT-1080, transgenic mice	Kemmer et al., 2010
IcaR	<i>Staphylococcus epidermidis</i>	Gentamicin	HEK293-T, CHO-K1	Stanton et al., 2014
lacI	<i>Escherichia coli</i>	IPTG	Rat2 fibroblasts, transgenic mice	Cronin et al., 2001
LmrA	<i>Bacillus subtilis</i>	Lincomycin	HEK293-T, CHO-K1	Stanton et al., 2014
McbR	<i>Corynebacterium glutamicum</i>	L-methionine	HEK293-T, CHO-K1	Stanton et al., 2014
MphR	<i>Escherichia coli</i>	Erythromycin	CHO-K1, NIH/3T3, HEK293-T, HT-1080, HUVEC cells, CHO-TWIN1	Weber et al., 2002
PhlF	<i>Pseudomonas protegens</i>	2,4-diacetylphloroglucinol	HEK293-T, CHO-K1	Stanton et al., 2014
Pip	<i>Streptomyces coelicolor</i>	pristinamycin I, virginiamycin	CHO-K1, BHK-21, HeLa, C2C12	Fussenegger et al., 2000

PmeR	<i>Pseudomonas syringae</i> pathovar <i>tomato</i> DC3000	parabens, including methylparaben, ethylparaben	HEK293-T, HeLa, HT-1080, hMSC-TERT, BHK-21, CHO-K1	Wang et al., 2015
QacR	<i>Staphylococcus epidermidis</i>	plant alkaloids	HEK293-T, CHO-K1	Stanton et al., 2014
RheA	<i>Streptomyces albus</i>	Heat	DT40	Weber et al., 2003a
ScbR	<i>Streptomyces coelicolor</i>	butyrolactones (e.g. SCB1)	CHO-K1, COS-7, HEK293-T, BHK-21, HepG2, HUVEC, mouse embryonic stem cells	Weber et al., 2003b
SpbR	<i>Streptomyces pristinaespiralis</i>	butyrolactones (e.g. SCB1)	See above	Weber et al., 2003b
TetR	<i>Escherichia coli</i>	tetracycline, doxycycline	HeLa	Gossen & Bujard, 1992
TraR	<i>Agrobacterium tumefaciens</i>	3-oxo-C8-HSL	HeLa, Huh7, HEK293-T, RD	Neddermann et al., 2003
TrpR	<i>Chlamydia trachomatis</i>	L-tryptophan	HEK293-T, HeLa, NIH/3T3	Bacchus et al., 2012
TtgR	<i>Pseudomonas putida</i>	flavonoids (e.g. phloretin)	CHO-K1, HEK293-T, BHK-21, COS-7, HT-1080, HaCat, NIH/3T3, human foreskin fibroblasts,	Gitzinger et al., 2009

			human foreskin keratinocytes	
VanR	<i>Caulobacter crescentus</i>	vanillic acid	CHO-K1, HEK293-T, HeLa, COS-7, BHK-21, HT- 1080, NIH/3T3	Gitzinger et al. 2012

1.4.3 Lactate-binding proteins suitable for use in a mammalian transgene expression system

There are a range of known prokaryotic lactate sensing proteins. IipR is a LysR-type transcriptional regulator that has lactate-dependent interactions with DNA (Brutinel & Gralnick, 2012), binding DNA in the presence of lactate such that the operator is kinked and becomes favourable to transcriptional activity, although notably IipR does not unbind from its operator (Maddocks & Oyston, 2008), a property which is a major hurdle for transferring any such system to a different organism, particularly one in a different domain. Bacterial proteins that have ligand-dependent binding to DNA will likely maintain this characteristic in a mammalian environment, allowing them to operate as part of a ligand-inducible transgene expression system; if the addition of a ligand only alters the conformation of the DNA to which a bacterial protein is bound, any upregulation that occurs may depend to some extent on additional proteins, the identity of which will not necessarily be known. Additionally, conformational changes to DNA that lead to transcriptional upregulation in bacteria will not necessarily have the same function in mammalian cells. LysR-type transcriptional regulators have not to date been used in mammalian systems for transgene expression, although it has been demonstrated in yeast after developing multiple protein mutants and screening over 80 responsive element designs (Skjoedt et al., 2016). LarR is a Crp-Fnr family protein that also has a lactate-dependent interaction with DNA that operates in a similar fashion as that of IipR (Desguin et al., 2015). No Crp-Fnr family protein has been used in a mammalian transgene expression system, to my knowledge. The lactate-dependent DNA binding protein LldR belongs to the FadR subfamily of the GntR family of transcription regulators (Georgi et al., 2008), which are mostly involved in the regulation of central metabolism (Jain, 2015). LldR functions by binding to its operator in the absence of lactate, and unbinding in its presence,

de-repressing an upstream promoter (Georgi et al., 2008; Gao et al., 2012). LldR is found in many bacterial species, but those from *Corynebacterium glutamicum* and *Pseudomonas aeruginosa* have been studied in detail (Georgi et al., 2008; Gao et al., 2012), with structural information available for both (Gao et al., 2008; Xin et al., 2016). There is also one example of a FadR family protein being used previously in a mammalian transgene expression system, VanR from *Caulobacter crescentus*, which is responsive to vanillic acid (Gitzinger et al., 2012). The LldR from *Corynebacterium glutamicum* was chosen for initial testing of a lactate-inducible transgene expression system, as information relating to its operator binding affinity was available (in contrast to LldR from *Pseudomonas aeruginosa*) and at the start of this project, it was the LldR for which structural information was available (Gao et al., 2008). Also, Georgi et al. (2008) present a large amount of evidence that LldR is de-repressing activity from a target gene in response to lactate in *Corynebacterium glutamicum*. They showed that the overexpression of LldR reduced the transcription and activity of its target LldD (a lactate dehydrogenase), while deletion of LldR had the opposite effect.

1.4.4 The LldR operators; designing an operator for use with a one-component system

LldR (from *Corynebacterium glutamicum*) has at least three characterised DNA binding sites (Gao et al., 2008). LldR has the greatest affinity for cgl2917-site1, which is given below:

TTGTGGTCTGACCATGT

Only the underlined portion was shown to be protected from DNaseI activity, suggesting that the initial “T” was not involved in binding, and so it was excluded from the operators used in the subsequent chapter. TGGT and ACCA are the two palindromic half-sites, both of which are essential for LldR binding. LldR binds to this sequence with a dissociation constant of 81 nM, which is comparable to some other transcriptional regulators that have been used in mammalian transgene expression systems (see table 1.3).

Table 1.3. Dissociation constants of bacterial transcriptional regulators that have previously been used in mammalian inducible transgene expression systems.

Name	Dissociation constant (nM)	Reference	Note
TetR	0.01	Lederer et al., 1995	
TtgR	15,400	Krell et al., 2007	Co-operativity of additional TtgR-operator binding events decreases the apparent dissociation constant
QacR	5.7	Schumacher et al., 2004	
EthR	146	Engohang-Ndong et al., 2004	
HdnoR	26	Sandu et al., 2003	

A spacer of 6 nucleotides between the end of one cgl2917-site1 and the beginning of another one was used for the assembly of multiple operators (where used), and resulted from a restriction enzyme scar made in the cloning process. It was assumed that a 6-nucleotide spacer would be sufficient to prevent any steric clashing between multiple LldRs bound to multiple operators. The use of multiple operators can increase the probability that LldR will be found adjacent to a promoter, and so increases the likelihood that it will be able to exert some action on that promoter, improving inducibility (e.g. Gossen & Bujard, 1992; Hartenbach et al., 2007). Initially, the use of 1, 2, 6 or 12 operator repeats was considered for *in vivo* testing of the lactate-inducible transgene expression system in chapter 3.

1.4.5 Inducible transgene expression system design

There are three configurations used in ligand-inducible transgene expression systems used in mammalian cells (based on ligand-dependent DNA binding proteins). These are trans-repression, or de-repression by steric hindrance, where operator binding sites are downstream of a constitutive promoter; trans-silencing, or de-repression by heterochromatin recruitment, where operators can be located upstream or downstream of a constitutive promoter and the

transcriptional regulator is fused to a heterochromatin-recruiting domain; and trans-activation, where operators are typically located upstream of a minimal promoter and the transcriptional regulator is fused to a trans-activating domain (see figures 1.8, 1.9 and 1.10). The aim of this thesis was to establish a lactate-inducible transgene expression system, without necessarily considering more complex regulatory architectures (such as the use of a feedforward system capable of amplifying a weak signal, detailed amongst others in a review by Mathur et al., 2017) at this early stage.

1.4.5.1 De-repression by steric hindrance

De-repression by steric hindrance predominates as a mechanism of signal transduction in prokaryotes, as transcriptional machinery is not restricted in its access to genomic DNA, unlike in eukaryotes (Struhl, 1999). However, the reliance on strong operator binding affinity and the lack of an effector domain to bolster activity on a promoter means that it may not always be an appropriate system for use in a eukaryotic transgene expression system. This mechanism of transcriptional regulation has been employed in mammalian cells by Cronin et al. (2001), Stanton et al. (2014), and Mullick et al. (2006), and given that prokaryotic transcriptional regulators typically work by unbinding from their operators in the presence of an inducing ligand, these typically form “on-type” switches where the presence of the ligand turns on expression (see figure 1.8).

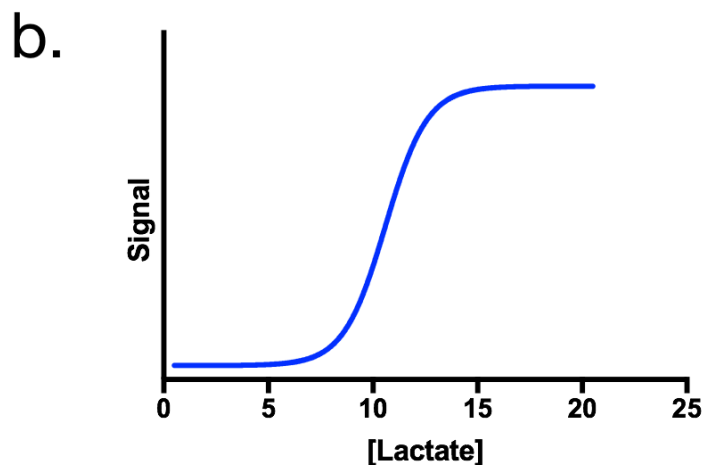
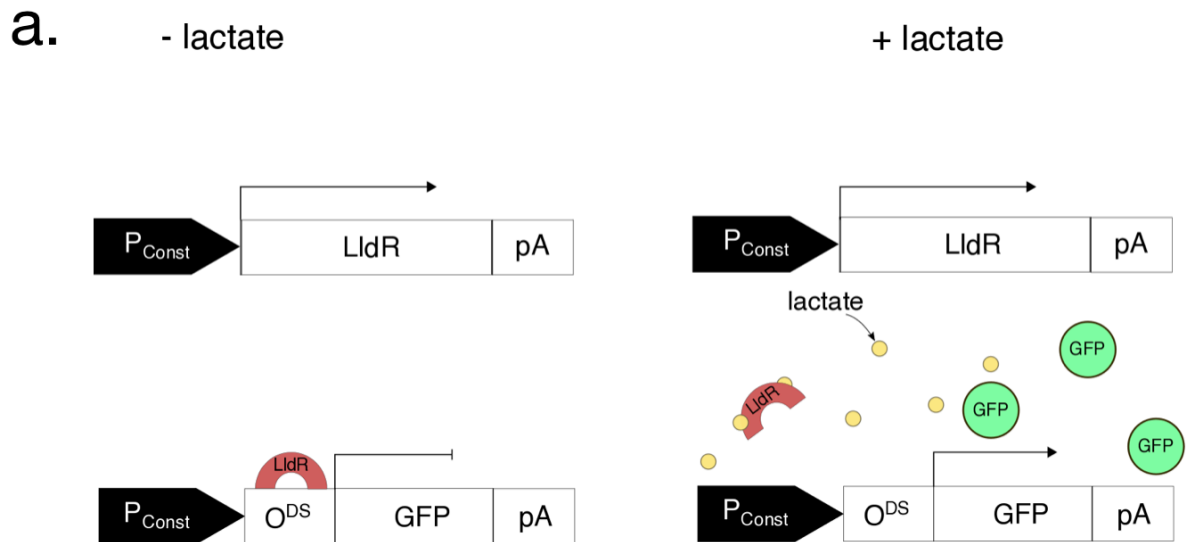


Figure 1.8. Transcription regulation by steric hindrance and de-repression. a. LldR binds to an operator downstream (O^{DS}) of a constitutive promoter (P_{Const}) in the absence of lactate, preventing the movement of an RNA polymerase past it to the transcription start site, inhibiting the expression of the reporter gene (in this case GFP). In the presence of lactate, LldR can no longer bind to its operator, relieving steric hindrance, de-repressing the expression of GFP. b. Ideal behaviour of such a system, with the x-axis showing increasing concentrations of lactate and the y-axis showing the signal produced.

1.4.5.2 De-repression by heterochromatin recruitment

De-repression by heterochromatin recruitment employs a transcriptional regulator fused to a mammalian repressor domain, typically the KRAB domain, which works by recruiting various heterochromatin-inducing factors to a targeted promoter (Deuschle et al., 1995; Groner et al., 2010), or the HDAC4 domain, which deacetylates histones, allowing tighter interactions between them and consequently silences expression from the targeted promoter (Bockamp et al., 2007; de Ruijter et al., 2003) (see figure 1.9). Although the dynamics of the transition from the silenced state once de-repressed can be slow (Bintu et al., 2016), it can exert relatively greater repression of a targeted promoter over a steric hindrance mechanism alone (e.g. Garriga-Canut et al., 2012), and operators can be placed both upstream and downstream of a promoter for a cumulative effect.

1.4.5.3 Transactivation

This configuration uses a transcriptional activator fused to the prokaryotic transcriptional regulator protein, typically the minimal herpes simplex virus VP16 domain by itself (e.g. Gossen & Bujard, 1992) or VP64 (e.g. Beerli et al., 1998), a tetrad repeat of VP16 with greater upregulation activity (Sawaki et al., 2013). VP16/VP64 works by forming a regulatory complex with the Oct1 transcription factor (Wysocka & Herr, 2003). In this configuration, the operator is normally placed upstream of a minimal promoter (one that has little to no basal activity and requires a trans-activator to be targeted to it). As prokaryotic transcriptional regulator proteins typically bind to their operators in the absence of a ligand, the transactivation configuration results in an “off-type” switch (see figure 1.10), although there are examples where the ligand-dependency of the DNA-binding has been reversed by protein engineering of the ligand-binding domain (e.g. Gossen et al., 1995, Mullick et al., 2006).

1.4.6 Protein fusion design

The core components of the transcriptional regulator protein are the ligand-binding domain (typically found at the C-terminus) and the DNA-binding domain (typically found at the N-terminus). In order to implement the trans-activating and trans-silencing configuration described above, it is necessary to fuse effector domains to the biosensor itself. Karlsson et al. (2011) suggest that these effector domains should be fused away from the DNA-binding

domain, which is located at the N-terminus in both the TetR and GntR families of transcriptional regulators (Cuthbertson & Nodwell, 2013; Suvorova et al., 2015). TetR family proteins may be able to weather N- or C-terminal fusions without hindering their DNA- or ligand-binding ability, given their use as bipartite transcriptional regulators where multiple such proteins are fused to each other (Folcher et al., 2013); however, the only current example of a GntR protein used in a mammalian transgene expression system, VanR, used a C-terminal effector domain fusion (Gitzinger et al., 2012), and when tested with an N-terminal fusion, inducibility was lost (Folcher et al., 2013). It was unknown at the start of this work whether this was the case for LldR, and so the optimal location for an effector domain fusion to LldR was tested in this chapter.

A nuclear localisation sequence (NLS) is also often fused to transcriptional regulator proteins in these inducible systems (e.g. Stanton et al., 2014, Deuschle et al., 1995, Cronin et al., 2001), allowing efficient movement of the biosensor to the nucleus. As the passive nuclear diffusion limit is below 90-110 kDa (Wang & Brattain, 2007), the fusion of an NLS to LldR (25.1 kDa), LldR-KRAB (35.7 kDa) or LldR-VP64 (32.4 kDa) is unlikely to be strictly necessary. The presence of the SV40 NLS may, however, be able to boost the proportion of a fusion protein that is found in the nucleus (Kalderon et al., 1984, Luo et al., 2004) and may also be able to increase the rate at which it is transported there (Hodel et al., 2006). Thus, its use was tested in this chapter.

1.4.7 Testing one component systems in mammalian systems

Transient transfection refers to the introduction of DNA into mammalian cells which is expressed for a short period of time, but is not integrated into the cell line and will gradually diminish. Stable transfections will integrate the DNA being tested into the genome of the host cell line, and will theoretically express the DNA indefinitely. Stable transfections will typically show greater levels of inducibility than transient ones; for example, both Mullick et al., (2006) and Weber et al., (2003) showed an induction ratio (that is, the fold-change between uninduced and induced states) that was an order of magnitude greater for stably integrated constructs, compared to the same ones transfected transiently. This may occur for multiple reasons - the interaction of the construct with chromatin will differ between transient, episomal expression and stable, genomic expression (Recillas-Targa, 2006); and transient transfection efficiencies, in terms of a plasmid or multiple plasmids being taken up by the cell and expressed, can vary widely between constructs with different plasmids used in a co-

transfection being taken up and expressed with different efficiencies (Zohar et al., 2001, Dalton & Barton, 2014), which can have a significant impact on the ability to acquire good promoter activity data (Taher, 2017). However, the time taken to perform transient transfections is much shorter, with experiments typically assayed 48 hours after transfection (e.g. Mullick et al., 2006), compared to the need for stable cell line construction, which involves lengthy selection procedures, ranging from 1 to 3 months (Longo et al., 2013), and will often require the use of additional genetic elements in order to aid the integration of the construct into the host genome (e.g. through use of the Flp Recombination Target with a compatible cell line - O’Gorman et al., 1991). Thus, it is often the case that preliminary work is performed with transient transfections to determine the optimal system configuration, prior to establishing stable cell lines with the constructs of interest.

The efficiencies of transient transfection plasmid uptake can be measured using flow cytometry, a technique where individual cells are passed through various sensors to measure multiple size and fluorescence parameters. A fluorescent protein can be linked to the transcriptional regulator protein of interest by a self-cleaving P2A peptide (Carey et al., 2009, Kim et al., 2011). When, for example, LldR-P2A-mCherry is translated, the P2A sequence will cleave itself and separate the regulator from the fluorescent protein. As a consequence, there should be a close correlation between the expression of mCherry and the expression of the LldR fusion construct, and mCherry-positive cells can be gated for further analysis, excluding all mCherry-negative cells. The ability to monitor transfection efficiency by single cell analysis means that flow cytometry is particularly well suited to the analysis of inducible promoters (Ducrest et al., 2002), by contrast with the more commonly used plate reader assay, which can only give population-level information about promoter expression.

The green fluorescent protein mAzamiGreen was the output used to measure the inducibility of the transgene expression system. mAzamiGreen is preferred over the more commonly used EGFP, as it has been shown to be slightly brighter in mammalian HeLa cell line (Karasawa et al., 2003).

The use of transient transfections into mammalian cells is how ligand-inducible transgene expression systems have been tested in previous published work, with optimisations of inducibility relating to operator number, position, etc., carried out in the same fashion before possible stable integration and testing of the optimised systems. To the best knowledge of the author, it is not the case that *in vitro* testing and optimisation of purified components is

performed prior to *in vivo* transfections, with it typically being sufficient for ligand-dependent DNA-binding to have been demonstrated previously for *in vivo* testing to proceed. As lactate-dependent DNA-binding of LldR has been demonstrated (Georgi et al., 2008), it was decided to forgo *in vitro* testing in favour of optimisation by *in vivo* transient transfections.

1.4.8 Promoters used

Two varieties of promoter are required for constructing the three inducible configurations given in figures 1.8, 1.9 and 1.10. For de-repression by steric hindrance and de-repression by heterochromatin recruitment, a constitutive promoter is required, where a high level of expression is seen in the absence of a targeted transcriptional regulator (possibly fused to a silencing effector domain). Both the SV40 and cytomegalovirus (CMV) promoters are popular in this regard (e.g. Kemmer et al., 2010; Deuschle et al., 1995), and the CMV promoter was chosen as the constitutive promoter in this project because of its relatively greater levels of expression. The transactivation configuration has typically used the minimal cytomegalovirus (minCMV) promoter, which exhibits a low basal activity and levels of activity comparable to a constitutive promoter when a trans-acting factor is targeted to it (Boshart et al., 1985). Although the minCMV promoter is known for its leakiness (Agha-Mohammadi et al., 2004), until recently there were no alternative minimal promoters for use in mammalian cells. A recent paper showed that YB_TATA, another minimal promoter, showed better basal expression levels (Ede et al., 2016); however, as it was released after work on this project commenced, it is only employed in chapter 5.

1.5 Project aims

Therapeutic glycoprotein production is typically performed in Chinese hamster ovary cells, which are prone to the deleterious accumulation of lactate. Current host cell engineering approaches can combat this accumulation, but do so in a constitutive and possibly suboptimal manner; in addition, assessing whether a particular endogenous gene will be a promising target for up- or downregulation with regards to mitigating lactate accumulation is a laborious process. This project aims to develop a lactate-inducible transgene expression system for use in Chinese hamster ovary cells, such that it can be used for the dynamic expression of genes able to mitigate the impact of lactate. Additionally, work was carried out to streamline future attempts at identifying putative target genes.

In pursuit of these aims, the following objectives were set:

- The testing and optimisation of a suitable lactate-sensing protein both *in vitro* and *in vivo*
- Testing a Cas9-VPR system capable of rapidly manipulating the CHO genome in order to identify new targets for lactate mitigation by host cell engineering

2 Materials and methods

Bacterial strains used are described in appendix 9.2. Additional plasmids and oligonucleotides are described in appendix 9.3, 9.7, 9.9, 9.11, 9.12 and 9.13.

2.1 Plasmids

Table 2.1. Plasmids used to express various LldR fusion constructs in *Escherichia coli*, prior to purification.

Plasmid name	LldR origin	N-/C-terminal fusion?	Notes	Plasmid origin
pCri-7b	N/a	N/a	Based on the IPTG-inducible pET-28a, proteins of interest cloned into this vector are fused to a C-terminal His6-tag.	Goulas et al., 2014
pCri-7b-LldR (<i>Corynebacterium glutamicum</i>)	<i>Corynebacterium glutamicum</i>	N/a		This study
pCri-7b-LldR (<i>Corynebacterium glutamicum</i>)-KRAB	<i>Corynebacterium glutamicum</i>	C-terminal KRAB		This study

pCri-7b-KRAB-LldR (<i>Corynebacterium glutamicum</i>)	<i>Corynebacterium glutamicum</i>	N-terminal KRAB		This study
pCri-7b-LldR (<i>Corynebacterium glutamicum</i>)-VP64	<i>Corynebacterium glutamicum</i>	C-terminal VP64		This study
pCri-7b-VP64-LldR (<i>Corynebacterium glutamicum</i>)	<i>Corynebacterium glutamicum</i>	N-terminal VP64		This study
pCri-7b-LldR (<i>Corynebacterium glutamicum</i>)-15-aa-VP64	<i>Corynebacterium glutamicum</i>	C-terminal 15aa-linker-VP64		This study
pCri-7b-LldR (<i>Pseudomonas aeruginosa</i>)	<i>Pseudomonas aeruginosa</i>	N/a		This study
pCri-7b-LldR (<i>Pseudomonas aeruginosa</i>)-VP64	<i>Pseudomonas aeruginosa</i>	C-terminal VP64		This study
pCri-7b-LldR (<i>Pseudomonas aeruginosa</i>)-15-aa-VP64	<i>Pseudomonas aeruginosa</i>	C-terminal 15aa-linker-VP64		This study

Table 2.2. Optimised single transient constructs. Plasmids with a P2A sequence refer to the use of a self-cleaving peptide fused between LldR and the red fluorescent mCherry protein. These constructs are used in chapter 5. *C. g.* refers to *Corynebacterium glutamicum*; *P.a.* refers to *Pseudomonas aeruginosa*.

Plasmid name	Encodes LldR or responsive element (RE)?	Promoter	LldR origin	Operator number	Operator location relative to promoter	Polyadenylation signal	Plasmid origin
MXS-CMV-L(<i>C. g.</i>)-P2A-mC-pA	LldR	CMV	<i>Corynebacterium glutamicum</i>	N/a	N/a	SV40	This study
	LldR is unfused to an effector domain.						
MXS-CMV-L(<i>C. g.</i>)-15-V-P2A-mC-pA	LldR	CMV	<i>Corynebacterium glutamicum</i>	N/a	N/a	SV40	This study
	LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.						
MXS-CMV-L(<i>P. a.</i>)-P2A-mC-pA	LldR	CMV	<i>Pseudomonas aeruginosa</i>	N/a	N/a	SV40	This study
	LldR is unfused to an effector domain.						
MXS-CMV-L(<i>P. a.</i>)-V-P2A-mC-pA	LldR	CMV	<i>Pseudomonas aeruginosa</i>	N/a	N/a	SV40	This study

	LldR has a C-terminal fusion of VP64.						
MXS-CMV-1x-mAz-pA	RE	CMV	N/a	1	Downstream	bGpA	This study
MXS-CMV-2x-mAz-pA	RE	CMV	N/a	2	Downstream	bGpA	This study
MXS-minCMV-1x-mAz-pA	RE	minCMV	N/a	1	Downstream	bGpA	This study
MXS-minCMV-2x-mAz-pA	RE	minCMV	N/a	2	Downstream	bGpA	This study
MXS-1x-minCMV-mAz-pA	RE	minCMV	N/a	1	Upstream	bGpA	This study
MXS-2x-minCMV-mAz-pA	RE	minCMV	N/a	1	Upstream	bGpA	This study
MXS-1x-minCMV-1x-mAz-pA	RE	minCMV	N/a	1 and 1	Up- and downstream	bGpA	This study
MXS-2x-minCMV-2x-mAz-pA	RE	minCMV	N/a	2 and 2	Up- and downstream	bGpA	This study
MXS-YB_TATA-1x-mAz-pA	RE	YB_TATA	N/a	1	Downstream	bGpA	This study

MXS-YB_TATA-2x-mAz-pA	RE	YB_TATA	N/a	2	Downstream	bGpA	This study
MXS-1x-YB_TATA-mAz-pA	RE	YB_TATA	N/a	1	Upstream	bGpA	This study
MXS-2x-YB_TATA-mAz-pA	RE	YB_TATA	N/a	2	Upstream	bGpA	This study

Table 2.3. Optimised double transient constructs, which express both a fusion variant of LldR and a variant of a response element. All plasmids include a P2A sequence referring to the use of a self-cleaving peptide fused between LldR and the red fluorescent mCherry protein. These constructs are used in section 5. *C. g.* and *C. glut.* refer to *Corynebacterium glutamicum*; *P.a.* and *P. aeru.* refer to *Pseudomonas aeruginosa*.

Plasmid name	Promoter for expression of:		LldR origin	Operator number	Operator location relative to promoter of response element	Polyadenylation signal of:		Plasmid origin
	LldR	Response element				LldR	Response element	
MXS-CMV-L(<i>C. g.</i>)-P2A-mC-pA-CMV-1x-mAz-pA	CMV	CMV	<i>C. glut.</i>	1	Downstream	SV40	bGpA	This study
LldR is unfused to an effector domain.								
MXS-CMV-L(<i>C. g.</i>)-P2A-mC-pA-CMV-2x-mAz-pA	CMV	CMV	<i>C. glut.</i>	2	Downstream	SV40	bGpA	This study
LldR is unfused to an effector domain.								
MXS-CMV-L(<i>P. a.</i>)-P2A-mC-pA-CMV-1x-mAz-pA	CMV	CMV	<i>P. aeru.</i>	1	Downstream	SV40	bGpA	This study
LldR is unfused to an effector domain.								
MXS-CMV-L(<i>P. a.</i>)-P2A-mC-pA-CMV-2x-mAz-pA	CMV	CMV	<i>P. aeru.</i>	2	Downstream	SV40	bGpA	This study

	LldR is unfused to an effector domain.							
MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-minCMV-1x-mAz-pA	CMV	minCMV	<i>C. glut.</i>	1	Downstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.							
MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-minCMV-2x-mAz-pA	CMV	minCMV	<i>C. glut.</i>	2	Downstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.							
MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-1x-minCMV-mAz-pA	CMV	minCMV	<i>C. glut.</i>	1	Upstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.							
MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-2x-minCMV-mAz-pA	CMV	minCMV	<i>C. glut.</i>	2	Upstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.							
MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-1x-minCMV-1x-mAz-pA	CMV	minCMV	<i>C. glut.</i>	1 and 1	Up- and downstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.							

MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-2x-minCMV-2x-mAz-pA	CMV	minCMV	<i>C. glut.</i>	2 and 2	Up- and downstream	SV40	bGpA	This study
LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.								
MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-YB_TATA-2x-mAz-pA	CMV	YB_TATA	<i>C. glut.</i>	2	Downstream	SV40	bGpA	This study
LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.								
MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-1x-YB_TATA-mAz-pA	CMV	YB_TATA	<i>C. glut.</i>	1	Upstream	SV40	bGpA	This study
LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.								
MXS-CMV-L(<i>C. glut.</i>)-15-V-P2A-mC-pA-2x-YB_TATA-mAz-pA	CMV	YB_TATA	<i>C. glut.</i>	2	Upstream	SV40	bGpA	This study
LldR has a C-terminal fusion of a 15 amino acid flexible linker and VP64.								
MXS-CMV-L(<i>P. aeru.</i>)-V-P2A-mC-pA-minCMV-1x-mAz-pA	CMV	minCMV	<i>P. aeru.</i>	1	Downstream	SV40	bGpA	This study
LldR has a C-terminal fusion of VP64.								
MXS-CMV-L(<i>P. aeru.</i>)-V-P2A-mC-pA-minCMV-2x-mAz-pA	CMV	minCMV	<i>P. aeru.</i>	2	Downstream	SV40	bGpA	This study

	LldR has a C-terminal fusion of VP64.							
MXS-CMV-L(<i>P. a.</i>)-V-P2A-mC-pA-1x-minCMV-mAz-pA	CMV	minCMV	<i>P. aeru.</i>	1	Upstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of VP64.							
MXS-CMV-L(<i>P. a.</i>)-V-P2A-mC-pA-2x-minCMV-mAz-pA	CMV	minCMV	<i>P. aeru.</i>	2	Upstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of VP64.							
MXS-CMV-L(<i>P. a.</i>)-V-P2A-mC-pA-1x-minCMV-1x-mAz-pA	CMV	minCMV	<i>P. aeru.</i>	1 and 1	Up- and downstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of VP64.							
MXS-CMV-L(<i>P. a.</i>)-V-P2A-mC-pA-2x-minCMV-2x-mAz-pA	CMV	minCMV	<i>P. aeru.</i>	2 and 2	Up- and downstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of VP64.							
MXS-CMV-L(<i>P. a.</i>)-V-P2A-mC-pA-YB_TATA-2x-mAz-pA	CMV	YB_TATA	<i>P. aeru.</i>	2	Downstream	SV40	bGpA	This study
	LldR has a C-terminal fusion of VP64.							

MXS-CMV-L(<i>P. a.</i>)-V-P2A-mC-pA-1x-YB_TATA-mAz-pA	CMV	YB_TATA	<i>P. aeru.</i>	1	Upstream	SV40	bGpA	This study
LldR has a C-terminal fusion of VP64.								
MXS-CMV-L(<i>P. a.</i>)-V-P2A-mC-pA-2x-YB_TATA-mAz-pA	CMV	YB_TATA	<i>P. aeru.</i>	2	Upstream	SV40	bGpA	This study
LldR has a C-terminal fusion of VP64.								

2.2 Molecular biology techniques

2.2.1 Restriction enzyme digestion

DNA was digested with high-fidelity restriction enzymes from New England Biolabs (Ipswich, USA) in CutSmart buffer, according to the manufacturer's instructions. This was followed by purification by agarose gel electrophoresis or a spin column kit (DNA Clean & Concentrator, Zymo Research, Irvine, USA).

2.2.2 Insert generation via annealed short oligonucleotides

This method was preferred for the addition of operators to existing vectors. Oligomers were designed in silico such that when annealed to each other, they would have overhanging bases complementary to a digested plasmid vector, and were ordered from Thermo Fisher (Inchinnan, UK). 5' phosphorylation of the separate oligomers was carried out with T4 Polynucleotide Kinase (NEB, Ipswich, USA) according to the manufacturer's instructions, after which each pair was annealed in Buffer EB (10 mM Tris-Cl, pH 8.5; QIAGEN, Hilden, Germany) using a thermocycler with an initial temperature step of 95 °C for 2 minutes, cooling to 25 °C at a ramping down rate of 1 °C per minute. The annealed oligomers were subsequently used as inserts in an otherwise typical restriction enzyme cloning procedure.

2.2.3 High-fidelity PCR with Phusion polymerase

When PCR products were to be used as inserts in cloning efforts, fidelity to a given template is vital and so the use of Phusion High-Fidelity DNA Polymerase (NEB, Ipswich, USA) was preferred. Primers pairs were designed using the NEB web application "Tm Calculator v1.8.1" such that they are between 18-25 bases in length, with a GC content between 40-60%, and with melting temperatures between 45 and 72 °C. Annealing temperatures subsequently used were also determined using this application. In certain cases it was necessary to incorporate additional restriction enzyme sites in these primers, but this is not relevant to melting temperature concerns. Reaction conditions were based on the manufacturer's instructions. The thermocycling conditions used were a 30 second initial denaturation step at 98 °C, followed by 35

cycles of a 10 second denaturation step at 98 °C, a 30 second annealing step at the previously determined temperature, and a 72 °C extension step of 30 seconds per kilobase of amplicon length. A final extension was carried out for 10 minutes at 72 °C. The samples were held at 10 °C until purification by agarose gel electrophoresis.

2.2.4 Gibson assembly

The web application Benchling (Benchling, 2019) was used to assist in the rapid assembly of multiple parts in a single round of cloning, facilitating the construction of primers to isolate these parts for use in Gibson isothermal assembly. The melting temperature of each primer in a pair was greater than 50 °C, with the difference between both melting temperatures kept within 5 °C, and the length of the homologous regions between each fragment kept between 20 and 50 bases. PCRs were performed with Phusion High-Fidelity DNA Polymerase (NEB, Ipswich, USA) according to the manufacturer's instructions, and were purified by gel electrophoresis (described in section 2.2.5). The parts to be assembled were combined in equimolar amounts (0.1 pmols each) and added to a Gibson master mix containing (at a 1x concentration) 5% PEG-800, 100 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 10 mM DTT, 0.8 mM dNTPs, 1 mM NADH, 0.06 units of T5 exonuclease (NEB, Ipswich, USA), 0.375 units of Phusion polymerase (NEB) and 60 units of Taq DNA ligase (NEB, Ipswich, USA), totalling 20 µL in volume. After incubation at 50 °C for 60 minutes, 10 µL of the reaction mix was transformed into chemically competent *Escherichia coli* Turbo cells (NEB, Ipswich, USA), and colonies were minipreped and sent for verification via Sanger sequencing.

2.2.5 Agarose gel electrophoresis, imaging and purification

Agarose gel electrophoresis was employed to separate DNA by weight. A 1% agarose to Tris-acetate-EDTA (TAE) weight/volume ratio was used for all gels, with SYBR Safe (Thermo Fisher, Inchinnan, UK) used as a DNA stain. 2.5 µL of HyperLadder 1kb (Bioline, Luckenwalde, Germany) was typically used as a molecular weight marker unless specified otherwise. Gels were subsequently imaged using a Gel Doc XR+ system (Bio-Rad, Hercules, USA). Relevant bands were excised and purified using a Zymoclean Gel DNA recovery kit (Zymo Research, Irvine, USA).

2.2.6 DNA ligation

400 units of T4 DNA ligase (NEB, Ipswich, USA) were used in 10 μ L reactions containing 50 ng of plasmid and a 3:1 molar ratio of insert to plasmid. The reaction was carried out according to manufacturer's instructions.

2.2.7 Preparation of chemically competent *Escherichia coli* cells; transformation of *Escherichia coli*

Both Turbo (NEB, Ipswich, USA) and Rosetta2 (Novagen, Madison, USA) – two strains of *Escherichia coli* – were prepared to be chemically competent in the same manner, after Inoue et al., 1990. In brief, this procedure concentrates and freezes *Escherichia coli* cells at a particular growth phase in a solution containing 1,4-piperazinediethansulphonic acid (PIPES), manganese chloride, calcium chloride, potassium chloride and dimethyl sulphoxide (DMSO), rendering the cells chemically competent. Transformation of these cells was performed by thawing aliquots on ice for 15 minutes, incubating them with plasmid DNA for 30 minutes, and heating to 42 °C for 45 seconds in a water bath. After incubating on ice for 2 minutes, 500 μ L SOC was added to the aliquot and the cells were placed in a 37 °C shaking incubator for 1 hour. The cells were then plated on agar plates containing the appropriate antibiotic, and moved to a 37 °C incubator to grow overnight, with transformed bacterial colonies appearing the next day.

2.2.8 Colony PCR with Taq polymerase

Colony PCR was used to determine the presence and size of an insert from bacterial colonies resulting from a transformation reaction. Primers pairs were designed to give an amplicon indicative of a successful transformation using the NEB web application "Tm Calculator v1.8.1", with base lengths between 15 and 30 bases, and with matched melting temperatures. Bacterial colonies were stabbed with a small pipette tip and mixed in 100 μ L volumes of water, from which 1 μ L was used as a template. REDTaq Readymix (Sigma-Aldrich, St. Louis, USA) was used as per the manufacturer's instructions. The standard cycling conditions used were an initial 10 minute cell lysis step at 94 °C, followed by 30 cycles of 1 minute template denaturation at 94 °C, 1

minutes of primer annealing dependent on the previously determined primer melting temperatures, a 72 °C extension step for 1 minute per kilobase of amplicon length, followed by a final extension period of 10 minutes at 72 °C. Agarose gel electrophoresis was used to determine the amplicon sizes. A template without the desired insert was typically used as a negative control.

2.2.9 Plasmid DNA isolation and purification

Plasmid DNA was purified from the *Escherichia coli* cloning strain Turbo (NEB, Ipswich, USA). Colonies were grown in 3.5 mL of 2xYT medium overnight at 37 °C, and were extracted using a Plasmid Mini Kit (QIAGEN, Hilden, Germany), following the manufacturer's instructions.

2.2.10 Golden Gate approach to design, construction and screening of guide RNA expression plasmids for use with Cas9-VPR

Guide RNAs were designed against specific sequences using the Benchling web application, with 5' sequence overhangs automatically added that were complementary to a Bsal-digested guide RNA entry vector. These oligonucleotides were ordered from Thermo Fisher (Inchinnan, UK) and phosphorylated with T4 Polynucleotide Kinase (NEB, Ipswich, USA) as in section 2.2.2. These annealed and phosphorylated oligonucleotides were ligated into Bsal-digested entry vector in a one-pot Golden Gate reaction, containing 5 units of amount Bsal (NEB, Ipswich, USA), 1 µL T7 DNA ligase (NEB, Ipswich, USA), 1x T4 DNA ligase buffer (NEB, Ipswich, USA), 75 ng of the entry vector, and ~600 ng of oligonucleotides (a 2 to 1 molar ratio of oligonucleotides to entry vector). This mix was incubated in a thermocycler at 37 °C for 1 hour, then at 55 °C for 5 minutes, after which point it was transformed into *Escherichia coli* Turbo cells (NEB, Ipswich, USA), as described in section 2.2.7. To streamline the selection of positively transformed cells, superfolder GFP (sfGFP) containing the two Bsal sites was present in the guide RNA entry vector, such that a successful cloning reaction would excise the sfGFP and allow the ligation of the guide RNA construct of interest. Thus, it was possible to distinguish between non-transformed cells (which expressed a highly visible level of sfGFP, apparent to the naked eye and especially on transillumination equipment - see panel G of figure 2.1)

and transformed cells, which did not express sfGFP and so appeared as normal colonies on selective media, resulting from the presence of a correctly integrated guide RNA construct. An overview of this process is given in figure 2.1.

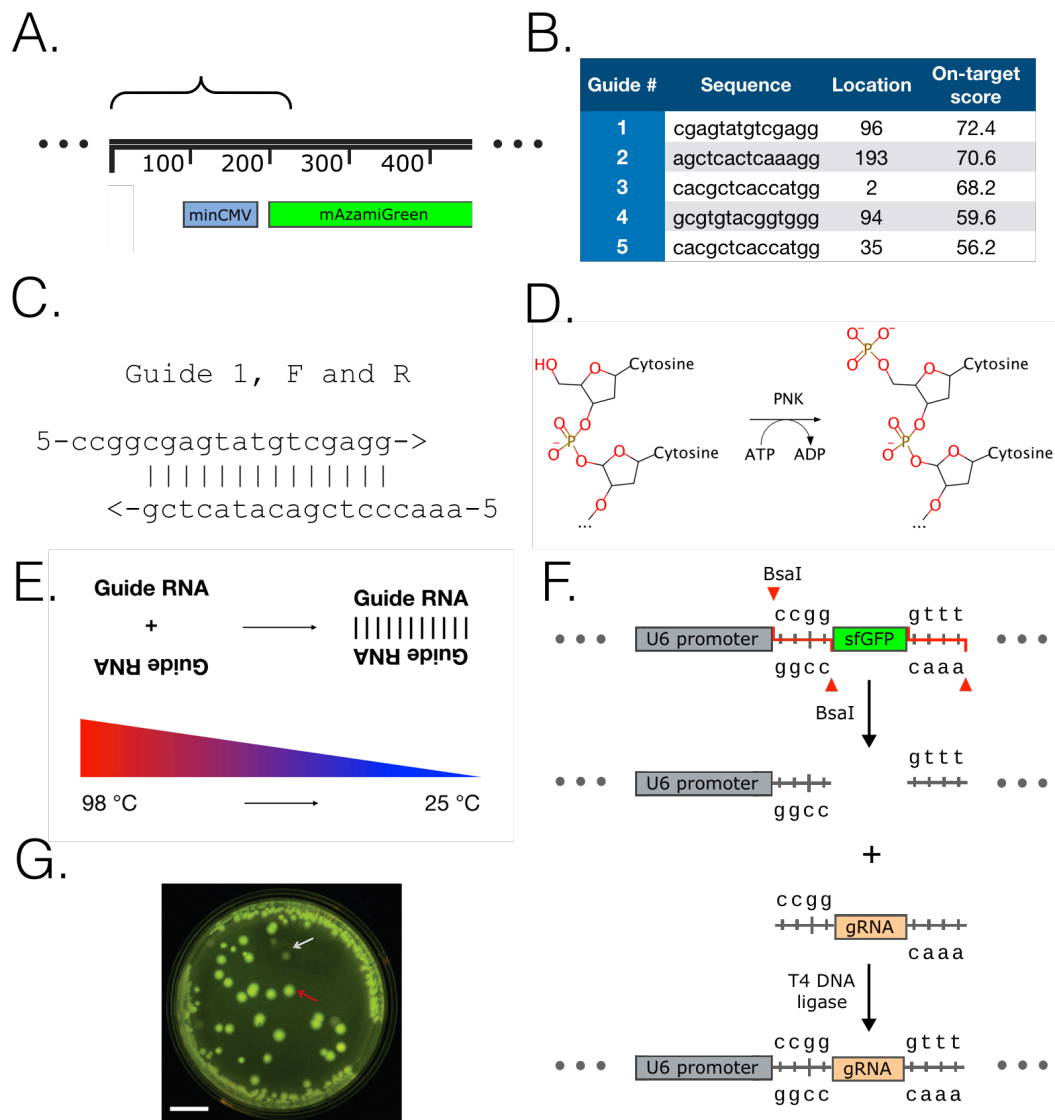


Figure 2.1. Depiction of the overall process required for the design, construction and screening of guide RNA-expressing constructs. A. The region upstream of the transcription start site is scanned for possible guide RNA targets using Benchling. B. A number of possible guide RNAs are returned and can be sorted by their on-target score. C. Two short oligonucleotides can be ordered to construct the guide RNA expression plasmid, with added overhangs complementary to the BsaI-digested entry vector. D. Oligonucleotides have phosphate groups added to their 5' hydroxyl groups to aid cloning efficiency. E. The two complementary guide oligonucleotides are incubated initially at a high temperature, cooling to room temperature over a 1 hour period to allow them to anneal without the generation of any off-target products. F. The

guide RNA entry vector encodes an sfGFP expression unit which is active in *Escherichia coli* and contains two BsaI recognition sites. Upon incubation with BsaI, the sfGFP expression unit is excised and the annealed and phosphorylated guide RNA oligonucleotide pair can be ligated in by T4 DNA ligase. G. After transformation and plating of the newly constructed guide RNA expression plasmid, colonies can be screened according to their fluorescence. An absence of fluorescence indicates no sfGFP expression and therefore is usually a positive transformant.

2.2.11 Assessment of RNA quality by bleach gel

To establish the integrity of the RNA isolated from CHO-S samples, bleach gel electrophoresis was performed (Aranda et al., 2012). This was performed as described previously here (under section 2.2.5) but with the addition of 5% w/v sodium hypochlorite to a molten agarose solution, to occupy 1% v/v of the final volume. The addition of bleach denatures the secondary structure of RNA and destroys any contaminating RNases.

2.2.12 cDNA preparation

Total RNA samples were first treated with 2 units of DNase I (NEB, Ipswich, USA) in 100 μ L of 1x DNase I reaction buffer for 10 minutes at 37 °C to degrade any contaminating DNA. 50 μ L of 5 M NaCl and 500 μ L of 60% isopropanol were added to the treated samples which are then loaded onto a silica membrane column (EZ-10 RNA Column and Collection Tube, Bio Basic, Toronto, Canada) and centrifuged at 13,000 x *g* for 1 minute. 700 μ L of wash buffer (15 mM Tris-HCl, 85% ethanol, pH 7.4) was spun through the column at 13,000 x *g* for 1 minute, and the RNA was eluted in 50 μ L of RNase-free water (Ambion, Inchinnan, UK).

Reverse transcription of the RNA samples was performed with the QuantiTect Reverse Transcription Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. For all RNA samples on which reverse transcription was performed, a "no reverse transcriptase" condition was also included, to be used as a control in subsequent qRT-PCR experiments to determine the presence of any DNA contamination.

2.2.13 Quantitative Reverse Transcription PCR (qRT-PCR)

2X SYBR Green JumpStart Taq Ready Mix (Sigma-Aldrich, St. Louis, USA) was used to set up the qRT-PCRs. cDNA was diluted 1/10 into water, 3 μ L of which was added to 1 μ L of 5 mM forward and reverse primers (2 μ L of primers total), in a total reaction volume of 20 μ L. The qRT-PCR was performed in an Eppendorf Mastercycler ep realplex (Eppendorf, Hamburg, Germany) as follows: initial denaturation at 95 °C for 5 minutes, 40 cycles of amplification with a 95 °C step for 30 seconds, a 55 °C step for 30 seconds, and a 72 °C step for 30 seconds. A melting curve was performed at the end of the run by heating from 55 °C to 95 °C, increasing by 0.2 °C every 2 seconds.

Primers were designed using the Primer3Plus web application (Untergasser et al., 2012), such that they would amplify across an exon-exon junction. The efficiency of each primer pair was determined by performing amplifications on a series of tenfold dilutions of cDNA templates, with the threshold cycles (CT) plotted against the log of the starting template quantity. The slope of the regression line of this plot and its *r* value was used to determine the efficiency of the primers. Primer efficiencies are given in appendix 9.12.

Relative levels of gene expression were analysed using the Pfaffl method (Pfaffl, 2001), following the below formula:

$$Ratio = \frac{(E_{target})^{\Delta CT, ref (untreated-test)}}{(E_{ref})^{\Delta CT, ref (untreated-test)}}$$

E_{target} and E_{ref} are the amplification efficiencies of the target and reference genes, respectively (in this case, actin-beta or ACTB). “ ΔCT , target (untreated - test)” is the CT value of the target gene in the untreated sample minus the CT of the target gene in the test sample; “ ΔCT , reference (untreated - test)” is the CT of the reference gene in the untreated sample minus the CT of the reference gene in the test sample.

2.3 Mammalian cell handling

2.3.1 Mammalian cell culture

The suspension cell line CHO-S (catalogue number A1155701, Thermo Fisher, Inchinnan, UK) was revived from low passage number stocks stored in liquid nitrogen, and were maintained in CD-CHO medium supplemented with 8 mM L-glutamine and 1X sodium hypoxanthine/thymidine (all from ThermoFisher). Cells were subcultured every 3-4 days at 1×10^5 cells/mL in 20 mL of complete medium, in a 125 mL non-baffled Erlenmeyer flask with a vented cap (VWR, Leuven, Belgium), rotating at 125 revolutions per minute on a shaking platform with an orbit of 16 mm, within a humidified incubator set at 37 °C and 8% CO₂. The cells were used in experiments after three passages.

Adherent CHO-K1 cells were purchased from Thermo Fisher (catalogue number R75807, Inchinnan, UK), and were maintained in 75 cm² cell culture flasks (VWR, Leuven, Belgium) in Ham's F12, 10% FBS, 2 mM L-glutamine and 100 µg/mL Zeocin (all from Thermo Fisher, Inchinnan, UK), in a humidified incubator set at 37 °C and 8% CO₂. Cells were passaged when ~80% confluent using TrypLE Express Enzyme (1X) with phenol red (Thermo Fisher, Inchinnan, UK) according to the manufacturer's instructions, and were diluted into new 75 cm² cell culture flasks with 15 parts fresh medium to 1 part cells. The cells were used in experiments after three passages.

All manipulations of the CHO-S and CHO-K1 cells were performed in a laminar flow hood.

2.3.2 Chemical transfection of mammalian cells

CHO-S cells were seeded at 1×10^6 cells/mL in complete medium the day prior to the transfection experiment. TransIT-PRO and PRO Boost transfection reagents (both Mirus, Madison, USA) were complexed with plasmid DNA and supplemented medium according to the manufacturer's instructions. In co-transfection experiments, equal amounts of each plasmid were used. 100 µL of transfected cells were plated onto a tissue-culture treated 96-well plate (Corning, Tewksbury, USA) at a density of 1×10^6 cells/mL, with 93.75 ng of each construct per well (unless stated otherwise). The plates were sealed with Breathe-Easy sealing membranes (Sigma-Aldrich, St. Louis, USA),

and were incubated on an LSE Digital microplate shaker (Corning, Tewksbury, USA), shaking at 800 revolutions per minute with an orbit of 3 mm, inside an incubator maintaining a humidified atmosphere of 37 °C and 8% CO₂. All experiments were performed in triplicate.

CHO-K1 cells were split once at a confluency of ~80% to a ratio of 1 volume of cells to 24 volumes of medium into tissue-culture treated 6-well plates (Corning, Tewksbury, USA). Immediately after passaging, transfection complexes were formed as above and added to the plates according to the manufacturer's instructions. If subsequently used in fluorescence microscopy experiments, round glass coverslips (VWR, Leuven, Belgium) were placed in the centre of each well prior to plating the cells.

2.3.3 Fluorescence microscopy

48 hours after the addition of transfection complexes to adherent CHO-K1 cells, the medium was removed and the cells were washed 3 times with PBS, fixed for 10 minutes at 37 °C in a 4% formaldehyde solution (Sigma-Aldrich, St. Louis, USA), stained with 1 µg/ml Hoechst 33342 (Thermo Fisher, Inchinnan, UK), mounted on a microscope slide (VWR) overnight with ProLong Diamond Antifade Mountant (Thermo Fisher, Inchinnan, UK) at room temperature. Images were collected with an Eclipse Ti Inverted Fluorescence Microscope (Nikon, Tokyo, Japan), with an excitation wavelength of 375 nm and an emission wavelength of 430 nm.

2.3.4 Flow cytometry

Flow cytometry data were collected 48 hours after transfection of CHO-S cells. Measurements were taken with an Attune NxT Flow Cytometer (Thermo Fisher, Inchinnan, UK), with mAzamiGreen measured with a 488-nm laser and a 525/20 filter, and mCherry measured with a 561-nm laser and a 610/20 filter. A non-transfected control was used in each experiment. Approximately 15,000 flow cytometry events were taken per biological replicate.

2.3.5 RNA isolation from CHO-S cells

Extraction of RNA from CHO-S cells was based on the RNASwift protocol (Nwokeoji et al., 2016). 10^7 CHO-S cells were resuspended in 200 μL of room temperature lysis reagent (4% SDS, 0.5 M NaCl, pH 7.5) and were incubated at room temperature for 3 minutes. After lysis, 100 μL of 5 M NaCl was added and the lysate was centrifuged for 4 minutes at 13,000 x g . The supernatant was transferred to a new tube, to which 500 μL of 60% isopropanol was added. This mixture was loaded onto a silica membrane column (EZ-10 RNA Column and Collection Tube, Bio Basic, Toronto, Canada) and centrifuged at 13,000 x g . The flow-through was discarded and 700 μL of wash buffer (15 mM Tris-HCl, 85% ethanol, pH 7.4) was spun through the column at 13,000 x g for 1 minute. The RNA was eluted in 100 μL of RNase-free water (Ambion, Inchinnan, UK). The concentration of RNA was measured using a Nanodrop spectrophotometer (Thermo Fisher, Inchinnan, UK).

2.4 Techniques relating to *in vitro* assays

2.4.1 Protein expression and purification

pCri7-b/LldR (and variant) expression vectors were transformed into Rosetta2 (Novagen, Madison, USA), an *Escherichia coli* BL21 DE3 strain co-expressing pRARE2 (providing the otherwise rare tRNAs used by the CHO codon optimised coding sequence). Colonies were used to inoculate overnight cultures which were then used to inoculate 250 mL of LB containing 50 $\mu\text{g/ml}$ kanamycin and 25 $\mu\text{g/ml}$ chloramphenicol, incubated on a shaking platform at 37 °C until an OD_{600} of ~ 0.6 was reached. At this point expression was induced using 0.5 mM IPTG, and the temperature was shifted to 25 °C for 20 hours. Cells were harvested by centrifugation at 4,000 x g for 20 minutes at 4 °C, the supernatant discarded and the pellet stored at -80 °C overnight. Once thawed, it was resuspended in lysis buffer (50 mM Na_3PO_4 , 300 mM NaCl, 10 mM imidazole; pH 8.0 with NaOH), 1 mg/mL lysozyme and 1 mM phenylmethanesulfonyl fluoride, and left on ice for 30 minutes. Sonication was then performed using a 130 watt VCX 130 PB sonicator (Sonics, Newtown, USA) for 2 minutes with 10 second pulses at 50% amplitude, with a 10 second cooling period between each pulse. The lysate was cleared by centrifugation at 10000 x g at 4 °C for 30 minutes and loaded onto a 5 mL HisTrap HP column (GE Healthcare Life Sciences, Chicago, USA) equilibrated with lysis buffer. The column was eluted with a linear

gradient of imidazole (up to 500 mM), using an AKTA Pure (GE Healthcare Life Sciences, Chicago, USA) (see appendix 9.5 for details). SDS-PAGE analysis was used to determine which elution fractions contained the protein of interest. Three rounds of dialysis were sequentially performed on the selected fractions, where the fractions were loaded into Spectra/Por 4 Dialysis Tubing (Spectrum Labs, Irving, USA) that had been soaked in water for 30 minutes and placed into 100 volumes of buffer B (20 mM Tris-HCl, pH 7.5, 0.2 M NaCl, 10% glycerol), which was then gently agitated by magnetic stirring for 3 - 4 hours in a 4 °C environment. Protein concentrations were quantified using a DC Protein Assay (Bio-Rad, Hercules, USA), following the manufacturer's instructions.

2.4.2 Gel shift assays

Following Gao et al. (2008), LldR construct variants in buffer B (20 mM Tris-HCl, pH 7.5, 0.2 M NaCl, 10% glycerol) were mixed with DNA fragments in the same buffer. To test the impact of lactate on the DNA binding properties of LldR, lactate was incubated with the protein for 15 minutes prior to mixing with the DNA fragment. The protein-DNA mixture was incubated for 1 hour at room temperature and separated on 15% native polyacrylamide gels in Tris-borate-EDTA buffer (TBE: 89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3), running at a constant current of 8 mA per gel at 4 °C. The gels were stained with GelRed (Biotium, Fremont, USA) for 10 minutes before imaging with a Gel Doc XR+ system (Bio-Rad, Hercules, USA).

3 Design and *in vivo* testing of a lactate-inducible transgene expression system in CHO cells

3.1 Chapter aims

- To design a heterologous lactate-inducible transgene expression system for use in CHO, via a trans-repressing, trans-silencing, or trans-activating configuration
- To conduct initial *in vivo* testing of these systems in CHO cells
- Investigate nuclear localisation of the lactate-dependent DNA binding protein LIdR in CHO cells

3.2 Chapter summary

Components described in the introduction (LIdR, its operator, a constitutive or minimal promoter, an effector domain fused to LIdR, and a GFP output) were combined with the intention of creating a lactate-inducible transgene expression system. The cloning and experimental strategy to construct and test such a system are described. Fluorescence microscopy experiments demonstrated that the fusion of a nuclear localization signal was unnecessary, and its use was consequently discontinued. None of the configurations tested in this chapter produced a change in signal in response to the addition of lactate, necessitating the use of *in vitro* testing of individual components, detailed in the following chapter.

3.3 Results

3.3.1 Summary of experiments performed in this chapter

There are numerous variables that can influence the inducibility of a transgene, including the necessity and location of the NLS; the location of the KRAB/VP64 effector domain fusion relative to LIdR; the number of operators; and the location of the operators relative to a promoter. This gives rise to 216 possible experimental combinations, which would not be practical to carry out. Given the number of these variables, it was decided to limit the number of conditions that were tested initially, prior to more granular optimisation once a functional

system had been established. The assumptions used to narrow the experiments performed are given below:

It was assumed that NLS, KRAB and VP64 would not themselves be influenced by positioning effects, as KRAB is notably flexible at both of its termini (ref, Saito et al., n.d.; PDB: 1V65) and has previously been used in mammalian transgene expression systems with fusions at both N- and C-termini (Xie et al., 2014); although VP64 is normally fused at its N-terminus, as it is a tetrad of VP16, this suggests that fusions to either terminus should not have a detrimental impact on the function of VP64; and NLS is often used with both N- and C-terminal fusions, without any apparent issue (e.g. Chavez et al., 2015). It is assumed that only LldR would be affected by fusion positions. As an example of an outcome from this assumption, with the LldR fusion constructs LldR-KRAB-NLS and LldR-NLS-KRAB, it would only be considered necessary to test one of these conditions.

It was assumed that it may be necessary to include an NLS, so this features in many of the designs.

It was assumed that single operators would not give the optimal level of induction compared to multiple operators (e.g. Gitzinger et al., 2012), and so they were not used in the initial round of testing.

Operators were not tested downstream of a minimal promoter in any instance, on the assumption that steric hindrance would exceed any effect on transcriptional upregulation.

No more than 2 operator repeats would be placed downstream of a constitutive promoter, on the basis that anything longer could severely diminish the activity of the promoter.

Thus, these limiting assumptions were able to reduce the number of conditions to be tested initially from 216 to 38, and these conditions are listed in table 3.1 alongside their configuration types and observed fold-induction.

Table 3.1. All protein-operator-promoter construct combinations that were initially tested, alongside the configuration type, fold-induction and p-value calculated through the use of a Student's t-test (one-tailed, two-sampled, equal variance) to determine whether the uninduced and induced populations were significantly different. L refers to LldR (and is from *Corynebacterium glutamicum* in all cases), N refers to the SV40 nuclear localisation signal, K refers to the KRAB transrepressor domain, V refers to the VP64 transactivator domain, CMV refers to the cytomegalovirus promoter, and minCMV refers to the minimal cytomegalovirus promoter. Further details are given in table 9.4, including the average geometric mean of GFP and standard deviations.

Experiment number	Protein construct	Operator-promoter construct	Fold-induction (-lactate/+lactate)	p-value
<i>De-repression by steric hindrance configurations</i>				
1	L	CMV-2x	0.96	0.37
2	L	2x-CMV-2x	0.96	0.15
3	LN	CMV-2x	1.02	0.41
4	LN	2x-CMV-2x	0.98	0.09
5	NL	CMV-2x	0.96	0.37
6	NL	2x-CMV-2x	1.06	0.24
<i>De-repression by heterochromatin recruitment configurations</i>				
7	LK	CMV-2x	1.04	0.39
8	LK	2x-CMV	1.25	0.17
9	LK	6x-CMV	1.12	0.30
10	LK	12x-CMV	1.06	0.16
11	LK	2x-CMV-2x	1.06	0.32
12	KL	CMV-2x	0.70	0.16
13	KL	2x-CMV	1.42	0.11
14	KL	6x-CMV	0.98	0.39
15	KL	12x-CMV	0.76	0.10

16	KL	2x-CMV-2x	0.97	0.33
17	LKN	CMV-2x	0.94	0.08
18	LKN	2x-CMV	0.82	0.06
19	LKN	6x-CMV	0.98	0.31
20	LKN	12x-CMV	1.05	0.27
21	LKN	2x-CMV-2x	0.94	0.12
22	NKL	CMV-2x	0.81	0.08
23	NKL	2x-CMV	0.91	0.12
24	NKL	6x-CMV	0.82	0.07
25	NKL	12x-CMV	1.00	0.45
26	NKL	2x-CMV-2x	0.81	0.06
<i>Transactivation configurations</i>				
27	LV	2x-minCMV	1.01	0.46
28	LV	6x-minCMV	0.86	0.08
29	LV	12x-minCMV	0.88	0.33
30	VL	2x-minCMV	1.07	0.18
31	VL	6x-minCMV	1.03	0.44
32	VL	12x-minCMV	1.10	0.26
33	LVN	2x-minCMV	0.57	0.07
34	LVN	6x-minCMV	1.03	0.31
35	LVN	12x-minCMV	0.91	0.27
36	VNL	2x-minCMV	0.85	0.11
37	VNL	6x-minCMV	1.03	0.30
38	VNL	12x-minCMV	1.10	0.26

3.3.2 Use of MXS chaining

As table 3.1 shows, it is necessary to construct many different LldR fusion variants and response constructs. Various cloning approaches are available for this, such as Gibson Isothermal Assembly (Gibson, 2011) and Golden Gate Assembly (Engler et al., 2009). Although these approaches are highly useful, certain characteristics of the current project pose significant barriers to their implementation here. The protein fusion components (e.g. LldR, NLS, KRAB and VP64) are all to be ordered in multiple ways, and must be in frame, which can be difficult to achieve with the rigid pre-determined assembly order of the Golden Gate method. The response constructs can contain multiple operator repeats, which are unlikely to be amenable to a Gibson cloning approach due to its use of PCR. Although there may be possible workarounds to these issues using the aforementioned approaches, the MXS chaining approach (Sladitschek & Neveu, 2015) was selected. Analogous to the BioBrick assembly method (Shetty et al., 2008), all components are located initially in a cloning vector flanked at the 5' end with MluI and XhoI and Sall at the 3' end. If it is desired that block A and block B are to be fused together in-frame as block A-B, block A is digested with MluI and Sall, block B is digested with MluI and XhoI, and the two are ligated together. If the ordering desired is B-A, then B is digested with MluI and Sall and A with MluI and XhoI instead (see figure 3.1). In both cases, a translatable scar of valine-glutamic acid is made between the two parts, and it can be re-used subsequently as original pattern of restriction sites is recreated in the new construct.

3.3.3 Transient transfections into CHO-S, flow cytometry

After cells were transiently co-transfected with given constructs in the absence and presence of lactate, they were incubated for 48 hours before being assayed by flow cytometry. The flow cytometry data was processed in the following way, to limit subsequent analyses to viable and transfected cells: cells which do not resemble the forward and side scatter profiles of a negative control are excluded as non-viable cells (Homann et al., 2017) (see figure 3.2a); then, single cells are selected for by looking at the area and height of forward scatter, excluding the cells which have a greater area than would be expected for their height (see figure 3.2b); then, the RFP+ cells (those expressing mCherry as part of a P2A fusion to a transcriptional regulator construct) are selected (see figure 3.2c and 3.2d). From these cells, the geometric mean of the GFP signal is taken. These geometric means are averaged for the three biological replicates of each condition, for both uninduced and induced conditions. A Student's T-test (one-tailed, two-sampled, equal variance) is performed to determine if there is any significant difference between the results.

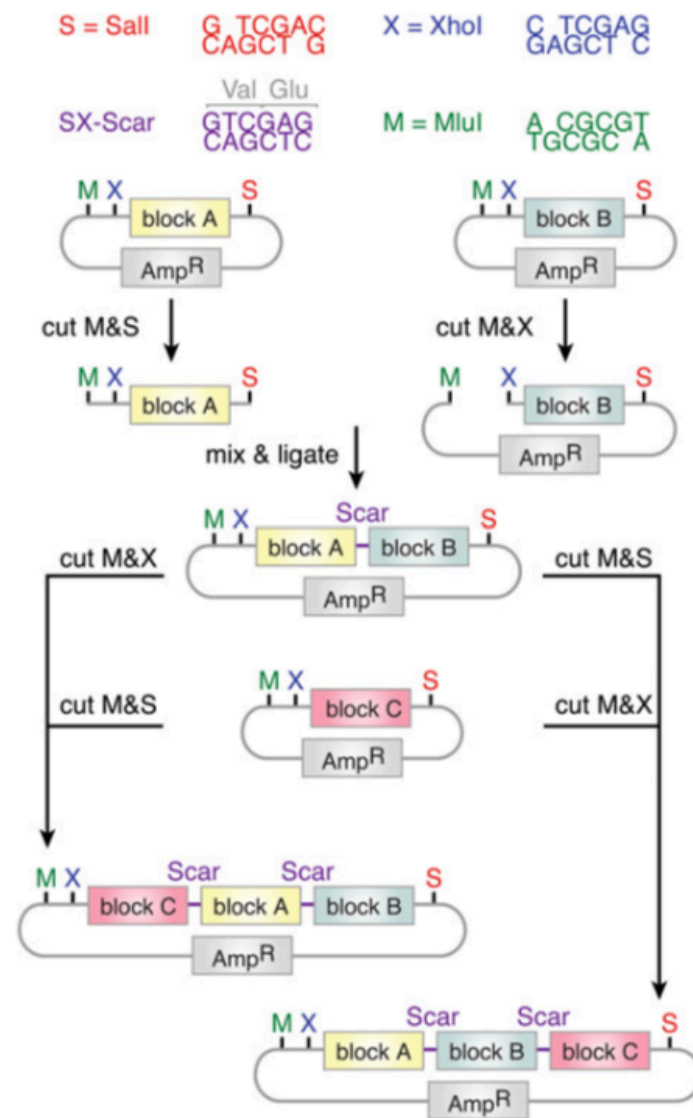


Figure 3.1. The principle of the MXS chaining cloning method (Sladitschek & Neveu, 2015). The restriction enzymes MluI (M), XhoI (X) and SalI (S) are used in this approach. Components or blocks are flanked at the 5' end with MluI and XhoI, and at the 3' end with SalI. The example in the image above shows how block A and block B can be assembled as block A-B by digesting the former with MluI and SalI and the latter with MluI and XhoI, with the two subsequently ligated together. The scar generated from the ligation of the SalI and XhoI overhangs no longer forms a restriction endonuclease recognition sequence. As a demonstration of how MXS chaining can be used to direct the assembly order of genetic components, if block A-B is digested with MluI and XhoI, and block C is digested with MluI and SalI, the ligation of the two digested components will produce block C-A-B; digestion of block A-B with MluI and SalI and block C with MluI and XhoI will produce block A-B-C. Image reproduced with permission of the rights holders, Sladitschek & Neveu, detailed in appendix 9.14.

The initial transient co-transfections did not show any significant results. The results are shown in Figure 3.3, with the remaining results shown in appendix 9.3, table 9.4.

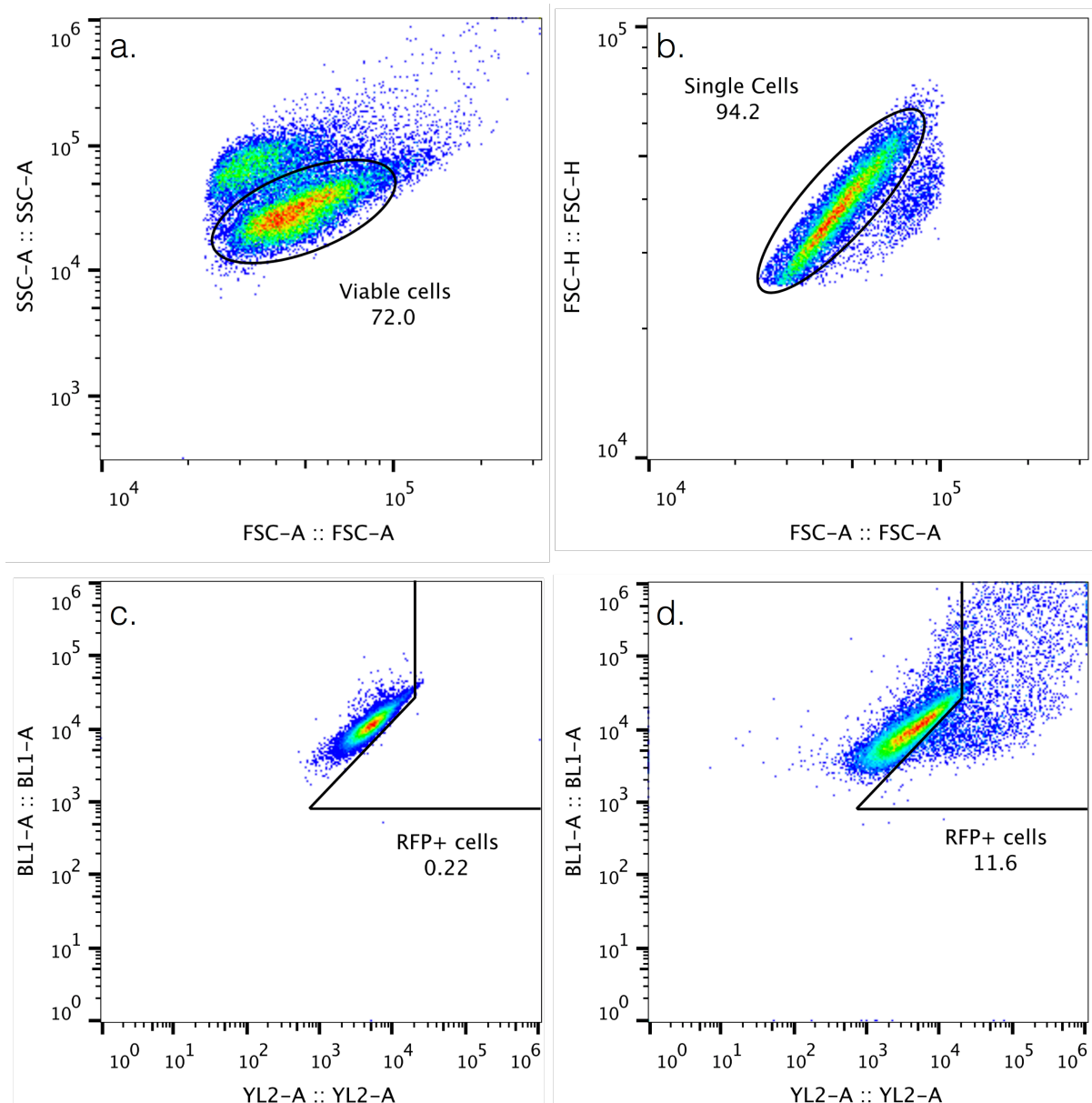


Figure 3.2. Flow cytometry gating strategy. The numbers inside each dot plot show the percentage of events within the adjacent gate. a. selection of viable cells, based on forward and side scatter. b. selection of singlet events based on the height and area of the forward scatter measurement. c. gating of RFP+ cells based on a sample that was incubated with transfection reagents but no DNA. d. shows the application of this gate to a sample transiently transfected with CMV-LV-P2A-mCherry.

Constructs sorted by mean fold-change in geometric mean of GFP signal

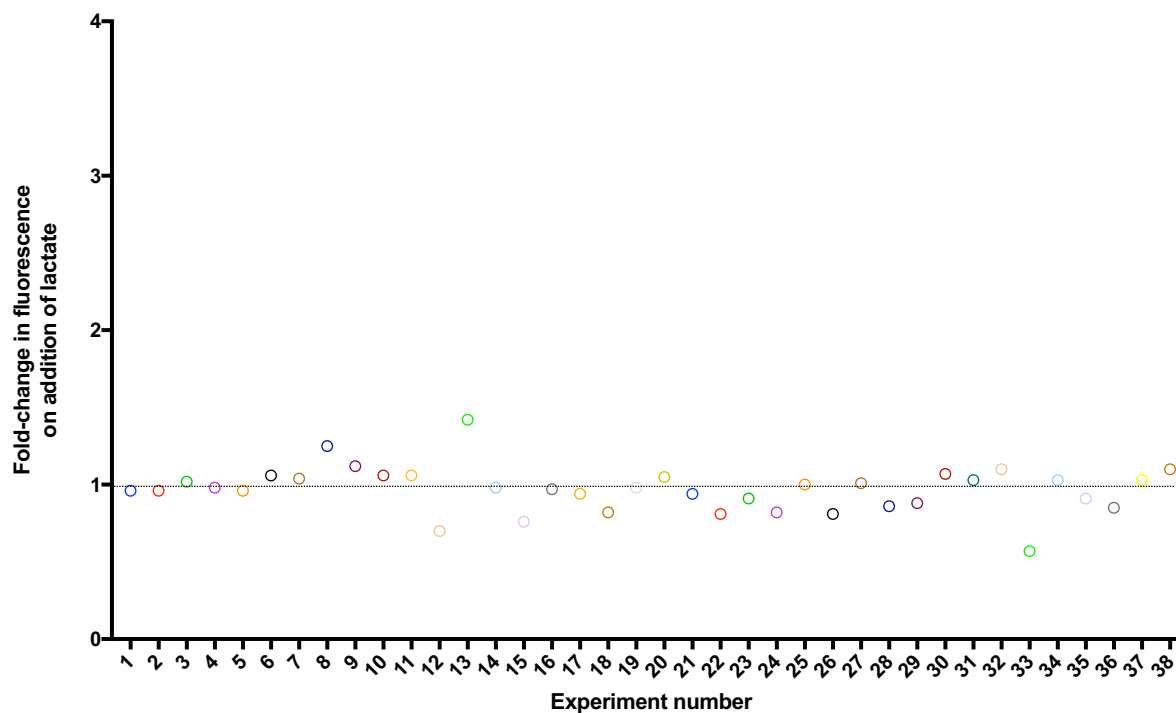


Figure 3.3. Conditions sorted by experiment number. No configurations tested here returned a significant result. The horizontal dotted line shows a fold-change of 1, where the uninduced geometric mean of the GFP signal is equal to the induced signal. Hollow circles indicate a p value greater than 0.05. The identity of the protein construct and operator-promoter construct used in each experiment number is given in table 3.1 and table 9.4.

3.3.4 Transfection of constructs for visualisation of LldR nuclear localisation

To determine whether the LldR fusion constructs are found in the nucleus (where they should exert their function on a co-transfected response module), adherent CHO-K1 cells were transfected with either CMV-LldR-mAzamiGreen or CMV-LldR-NLS-mAzamiGreen and grown on glass coverslips. This adherent CHO-K1 cell line was chosen over the suspension CHO-S cell line primarily used in this thesis, owing to the convenience of using this cell line for microscopy purposes. 48-hours post-transfection, each condition was treated with the Hoechst 33342 nuclear stain and imaged under a fluorescence microscope. This showed that LldR was able to localise to the nucleus, even in the absence of a nuclear localisation signal (see figures 3.4 and 3.5). A number of untransfected cells are seen in both conditions, as the transfection reagents used (TransIT-Pro) are primarily for use in suspension CHO cells. Scale

bars are not given in either figure as the microscope was incorrectly configured at the time when the photographs were taken.

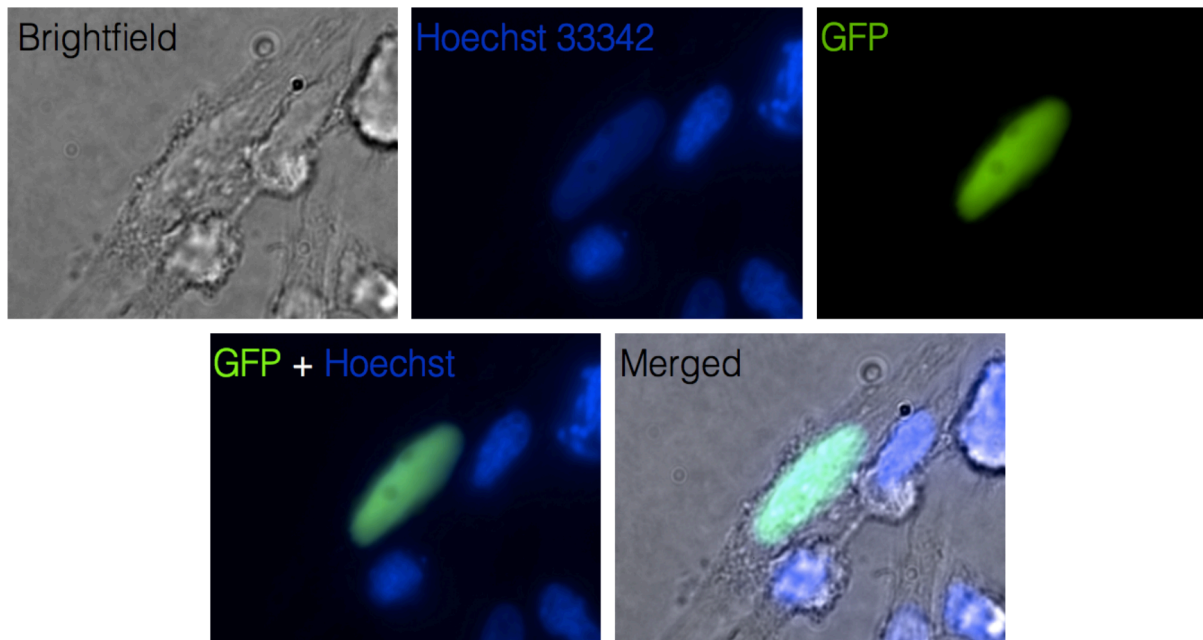


Figure 3.4. Subcellular localisation of LldR-NLS (fused to GFP/mAzamiGreen) in adherent CHO cells. The brightfield image shows the entire cell. The Hoechst 33342 dye is a DNA stain and is used to indicate the location of the nucleus. The GFP filter indicates the location of the LldR-NLS fusion protein, with the GFP + Hoechst composite showing that for one cell they completely overlap. This shows that the LldR-NLS fusion is localised to the nucleus 48-hours after transfection.

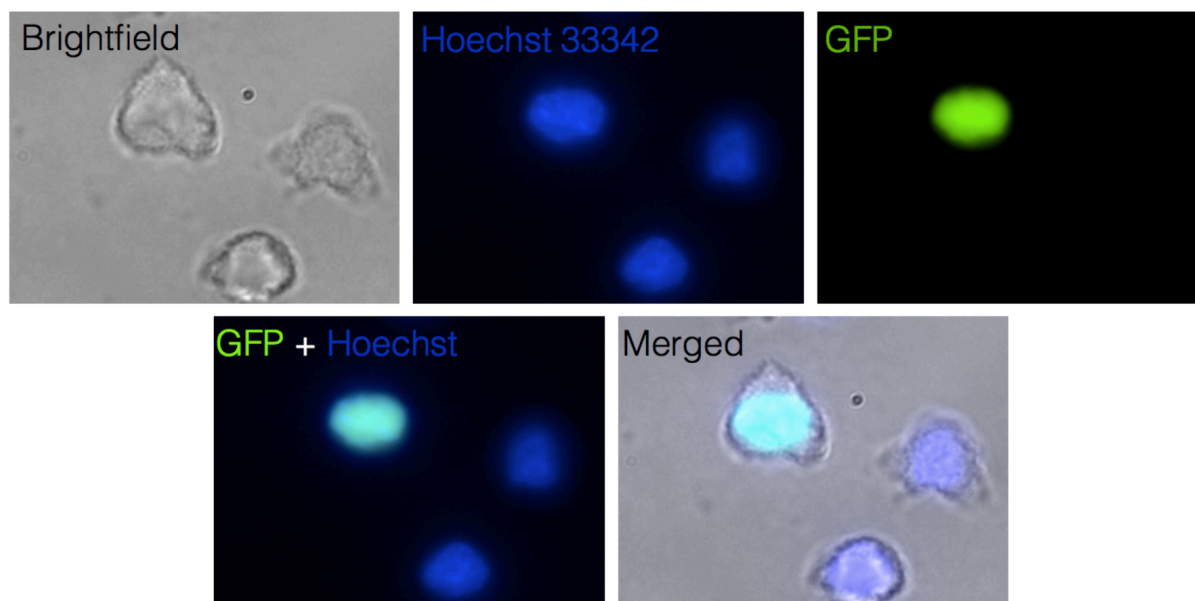


Figure 3.5. Subcellular localisation of LldR (fused to GFP/mAzamiGreen) in adherent CHO cells. The brightfield image shows the entire cell. The Hoechst 33342 dye binds to the minor groove of DNA, allowing it to stain the nucleus. The GFP filter indicates the location of the LldR fusion, with the GFP + Hoechst composite indicating that they fully overlap in the nucleus, and that LldR is able to passively translocate to the nucleus 48-hours post-transfection without the need for a nuclear localisation signal.

3.4 Summary of results; discussion

Of the constructs tested so far, none were found to be functional. At this stage, it was difficult to identify the cause of this lack of induction. Expression of LldR itself in CHO cells is unlikely to be an issue, given the number of proteins from the same or similar families that have been expressed in mammalian cells previously - additionally, the successful localisation of the LldR(-NLS)-mAzamiGreen construct to the nucleus of CHO-K1 indicates proper expression and folding of LldR. The fluorescence microscopy images show that LldR is able to be trafficked to the nucleus independently of fusion to an NLS, so it would have been available to act upon an inducible promoter construct there. The fact that no significant results showing induction were found means that no conclusions can be reached about why the system is not working as intended - it is impossible to know if the operator/promoter constructs, protein-effector fusions, or both were an issue.

4 *In vitro* optimisation of lactate biosensor components

4.1 Chapter aims

- Design and construct plasmids allowing LldR and fusion variants to be expressed in and purified from *Escherichia coli*.
- Express and purify LldR and fusion variants from *Escherichia coli*.
- Conduct EMSAs to:
 - Determine the impact of the operator sequence on protein-DNA binding affinity
 - Determine the impact of fusion variants on DNA binding ability
 - Determine the impact of operator spacing
 - Determine the impact of fusion variants on DNA unbinding in the presence of lactate

4.2 Chapter summary

In order to characterise the components constituting the lactate-inducible transgene expression system, LldR was expressed and purified from *Escherichia coli* (with and without fused effector domains) and its binding to various operator sequences was tested using the electrophoretic mobility shift assay. Expression of LldR fused to a KRAB domain was attempted but due to its low stability, it could not be tested and so was excluded from a subsequent round of *in vivo* testing, described in chapter 5. Improvements in the binding of LldR to its operator were established by slight modifications to the sequence flanking the LldR-binding portion, as well as increasing the distance between two adjacent operators. It was found that fusion of VP64 to the N-terminus of LldR (*Corynebacterium glutamicum*) abolished all DNA-binding activity, while the fusion of VP64 to the C-terminus of LldR was significantly detrimental but could be partially mitigated by the use of a flexible fusion linker. LldR from *Pseudomonas aeruginosa* was also tested with a C-terminal VP64 fusion, which only had a marginal impact on its ability to bind to an operator. The best performing components in these *in vitro* tests were taken forward to a subsequent round of *in vivo* testing.

4.3 Background

4.3.1 General outline of protein expression in *Escherichia coli*

The previous chapter showed that the initial designs of the transcriptional lactate biosensors were not functional *in vivo*, likely requiring the optimisation of the response element and/or the fusion of any effector domains. The typical method for characterising a protein-DNA interaction is to isolate these two components and measure the extent to which they bind to each other using an electrophoretic mobility shift assay (EMSA). To the author's knowledge, this approach has not been used previously for the optimization of a ligand-inducible transgene expression system. The operator DNA can rapidly and cheaply be synthesised by a third-party, while the protein must be expressed and purified.

The process of purifying a protein has multiple stages. The DNA coding sequence must either be synthesised or isolated from genomic DNA, after which point it is cloned into an expression vector that suits the intended expression host, mechanism of induction, and purification protocol to be followed. Once the protein of interest has been expressed successfully, it is purified and subsequently exchanged into a buffer suitable for experimentation and long-term storage.

4.3.1.1 Isolation and cloning of LldR and fusion variant constructs

Two LldR proteins were tested in this chapter, either in isolation or fused to one of two effector domains, KRAB or VP64. These components were either synthesised with codon usage optimised for expression in CHO cells (LldR from *Corynebacterium glutamicum*, KRAB, and a flexible (GGGS)₃ linker); isolated from existing vectors optimised for mammalian expression (VP64 was amplified from the hCas9-VPR vector); or they were isolated from genomic DNA, in the case of LldR from *Pseudomonas aeruginosa*. All vectors used in this chapter are contained in Table 2.1.

These components were cloned via Gibson assembly into the pCri vector system (Goulas et al., 2014), based primarily on the pET expression vectors. This vector system includes many plasmids that are suitable for use in *Escherichia coli* with the benefit that moving a coding sequence from one plasmid to another is simplified by the use of a common set of restriction enzyme sites. The different pCri plasmids streamline the testing of multiple expression tags (such as maltose binding protein, thioredoxin, and glutathione-S-transferase), if protein yields

are otherwise insufficient, as well as allowing the positioning of hexahistidine purification tags at either (or both) termini of the fusion construct.

4.3.1.2 Codon usage; transformation of LldR constructs into Rosetta2 cells

Codon usage is the most important factor in determining the efficiency of protein expression in prokaryotes, compared to e.g. the sequence of the ribosome binding site or the identity of the stop codon (Lithwick & Margalit, 2003). In particular, the presence of AGA/AGG and other rare codons can reduce the level of protein expressed in *Escherichia coli* (Kane, 1995). The provision of these tRNAs on a separate plasmid can significantly boost production of a heterologous protein in *Escherichia coli* (e.g. Brinkmann et al., 1989).

The Rosetta2 BL21 DE3 *Escherichia coli* expression host used here has such a plasmid and is preferable for expression of the various LldR fusion constructs tested in this chapter, which contain many of the rare AGA/AGG codons (see Table 4.1).

Table 4.1. Frequency of rare codons found in LldR biosensor components.

Component name	Number of rare AGA/AGG codons	Number of consecutive rare codons	Proportion of codons contained with <10% usage for given amino acid
LldR (<i>Corynebacterium glutamicum</i>), codon usage optimised for CHO expression	6	3	8%
LldR (<i>Pseudomonas aeruginosa</i>)	2	0	6%
VP64	1	1	8%
KRAB	2	0	4%
(GSSSS) ₃	0	0	13%

4.4 Results

4.4.1 Induction of protein expression; cell lysis; protein solubility testing

After transformation of the LldR expression vectors into *Escherichia coli* strain Rosetta2 BL21 DE3, the expression conditions were based on those previously described for the *Corynebacterium glutamicum* LldR (Gao et al., 2008). The use of a long and relatively cool expression period here (25 °C for 20 hours) is known to improve the solubility of a protein of interest (e.g. Vera et al., 2007).

After induction and expression, cells must be lysed in order to prior to purification of the protein being overexpressed. The aim of cell lysis is to disrupt the membranes of as many cells as possible, without damaging the protein itself. Although many physical and chemical strategies to achieve this exist (e.g. multiple freeze-thaw cycles of a sample or use of solubilising detergents), sonication was found to provide sufficient yields of protein and so no additional approaches were tested.

Two fractions can be separated from the cell lysate - a soluble fraction in which proteins are possibly in a native conformation, allowing subsequent purification, and an insoluble fraction resulting from misfolding and aggregation of proteins, where purification is still possible but with the requirement that the protein of interest is refolded into an active conformation. It is highly preferable that a protein of interest is found in the soluble fraction of a cell lysate, as refolding protocols are lengthy, have variable levels of efficiency, require a significant additional level of testing and optimisation, and may not be possible within a given timeframe or at all (Yamaguchi & Miyazaki, 2014). Solubility of a protein is determined by conducting a sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) experiment on soluble and insoluble fractions, both prior to and after induction. SDS-PAGE analysis allows the separation and visualisation of molecules according to their molecular weight. The testing of both pre- and post-induction samples indicates whether a protein has been successfully expressed, in addition to the use of a protein size ladder to determine the size of the bands in each lane.

Most of the LldR fusion variants expressed successfully and were found at sufficient levels in the soluble fraction, allowing subsequent purification. Figure 4.1 shows the solubility testing of LldR (*Corynebacterium glutamicum*), by itself and with KRAB fused at either of its termini. The fusion of KRAB to LldR (at either termini) was the only instance in this chapter of a protein fusion variant that localised primarily to the insoluble fraction, and was consequently subjected

to attempts at optimising its expression. Solubility testing by SDS-PAGE analysis of all other proteins expressed in this chapter is found in appendix 9.4.

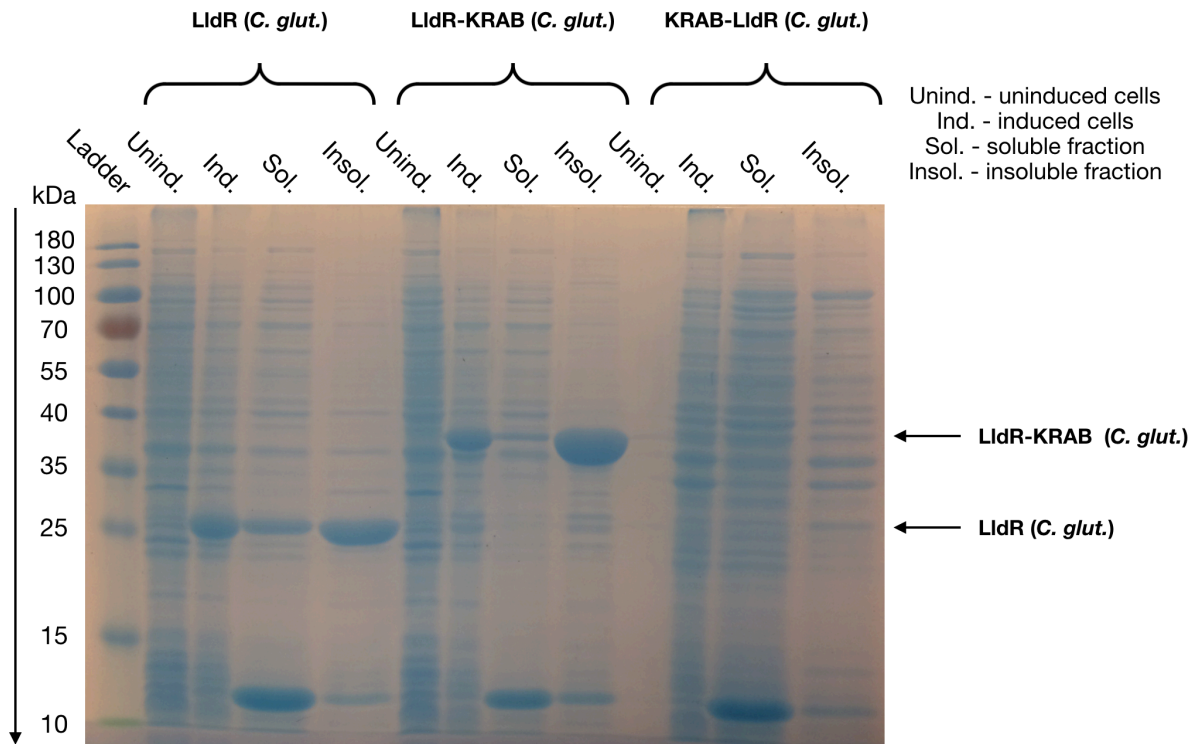


Figure 4.1. SDS-PAGE analysis of the solubility of LldR (*Corynebacterium glutamicum*) by itself and with the KRAB domain fused to either its N- or C-terminus. Cell samples were taken before and after induction with 0.5 M IPTG, and post-expression cell lysates samples were separated into soluble and insoluble fractions by high-speed centrifugation. Expression is shown by the presence of a large additional band in the induced sample, relative to the uninduced sample, at the expected size (indicated by the arrows to the right). Expression was seen for LldR and LldR-KRAB, but not for the KRAB-LldR condition. LldR was found at significant levels in its soluble fraction, sufficient to proceed to purification. Although successfully expressed, the LldR-KRAB protein appeared to be primarily in the insoluble fraction, with a relatively small portion in the soluble fraction. An attempt to optimise the expression of both LldR-KRAB and KRAB-LldR was carried out as a result. The lane showing the uninduced fraction of KRAB-LldR was incorrectly normalised and so is not apparent in this figure. The downward arrow shows the direction in which the proteins migrated.

4.4.1.1 Induction and lysis optimisation of constructs containing LldR

As shown in figure 4.1, both LldR-KRAB and KRAB-LldR initially failed to produce significant amounts of soluble protein. Two complementary approaches were employed in tandem in an attempt to boost successful folding of these proteins. As overexpression in *Escherichia coli* at low temperature can improve solubility and stability (Schein, 1989), the temperature of the expression was decreased from 25 °C to 16 °C, while the expression duration was increased to 24 hours to compensate for the reduced protein synthesis rate. Additionally, three lower concentrations of the inducing molecule IPTG were tested (0.25 mM, 0.01 mM and 0.005 mM), as this has been shown to be effective in limiting protein aggregation in some cases (e.g. Sadeghi et al., 2011). A small improvement in the solubility of LldR-KRAB was only achieved through the use of 0.25 mM IPTG, with KRAB-LldR seemingly not expressed at visible levels (see figure 4.2). On this basis it was decided to proceed with a higher scale expression of LldR-KRAB at 16 °C over 24 hours, using IPTG at a concentration of 0.25 mM. No further efforts were made to express KRAB-LldR.

The KRAB domain is notoriously difficult to express. Peng et al. (2000) described it as remaining “profoundly insoluble” over multiple expression conditions and as part of many larger constructs, and only succeeded in obtaining a soluble and active KRAB protein when using a protocol where the protein was refolded on the Ni-NTA column. Structural analyses have shown the flexibility and tendency to aggregate of KRAB-containing proteins (Mannini et al., 2006), which is likely to be the source of inclusion body formation and therefore the difficulty in expressing this protein in *Escherichia coli*.

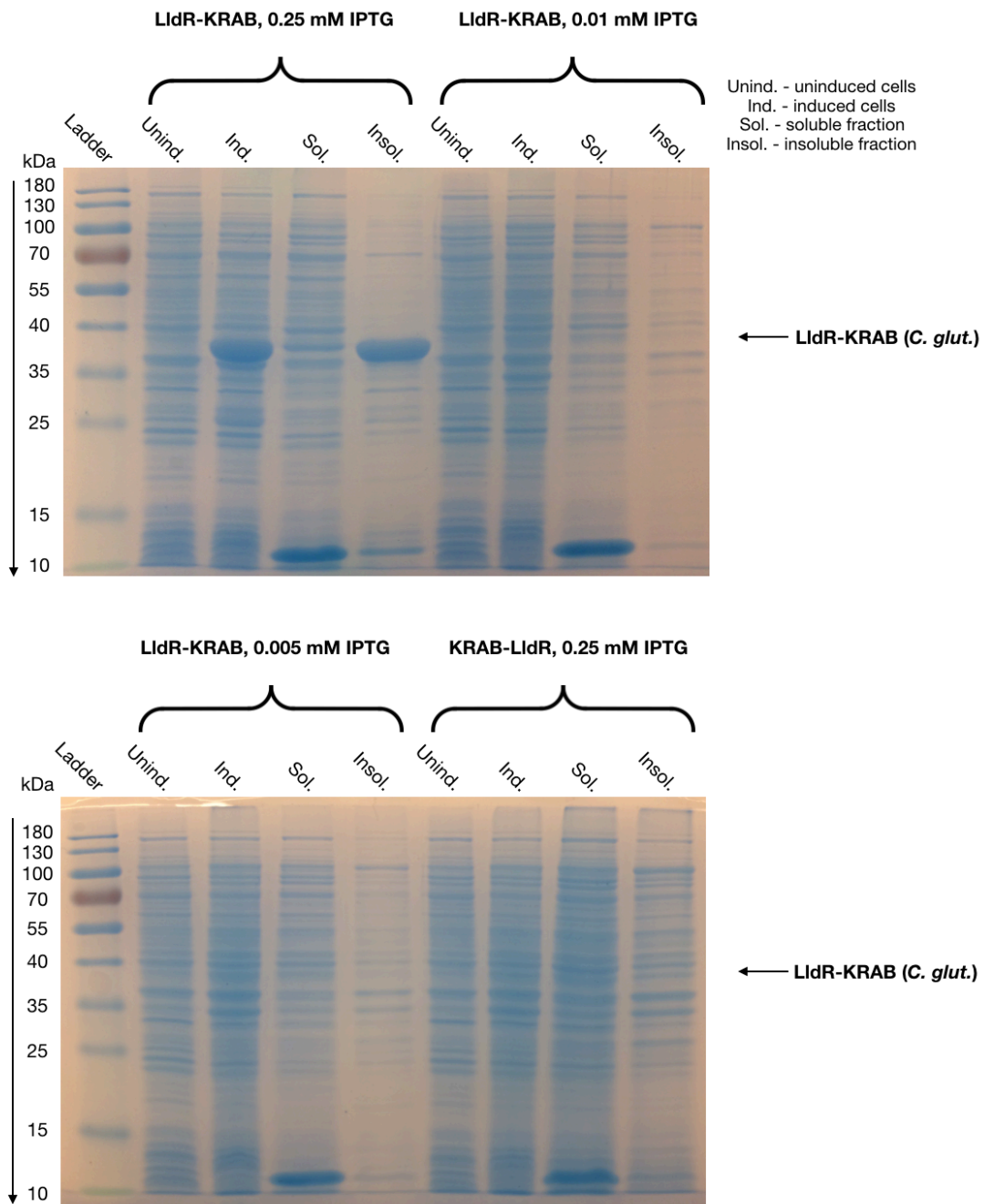


Figure 4.2. Testing the use of lower concentrations of IPTG to boost solubility of LldR (*Corynebacterium glutamicum*) fused to KRAB with SDS-PAGE. Only the use of 0.25 mM IPTG to induce expression was partially successful in shifting the amount of LldR-KRAB found in the soluble fraction. At lower concentrations of IPTG, expression of LldR-KRAB was not apparent. KRAB-LldR did not seem to be expressed at levels suitable for purification in any of the conditions tested. Data for KRAB-LldR expressed with 0.01 mM and 0.005 mM IPTG is not shown due to the poor quality of the gel used. The downward arrow shows the direction in which the proteins migrated.

4.4.2 Protein purification with nickel resin affinity via fast protein liquid chromatography

The recombinant LldR proteins expressed here all have a hexahistidine (6xHis) tag fused to their C-termini in order facilitate their purification from a cell lysate. This 6xHis tag has a high affinity for nickel-nitriloacetic acid (Ni-NTA) which can be immobilised in a matrix within a purification column. Other proteins within the lysate will generally not bind to the Ni-NTA and will pass through the column; the tagged protein will itself be eluted as the imidazole concentration of the buffer washed through the column is increased, as this also binds to the Ni-NTA and will gradually displace the tagged protein. Figure 4.3 shows how this increase in imidazole concentration can lead to the displacement of a 6xHis-tagged LldR (*Corynebacterium glutamicum*) protein, as evidenced by the UV absorbance signal taken of the eluent. This data is also provided in appendix 9.5 for the other proteins purified in this chapter.

During the elution process, multiple samples are collected and are analysed with SDS-PAGE in order to determine in which fraction the protein of interest is contained. Figure 4.4 shows the SDS-PAGE analysis of the purification of LldR (*Corynebacterium glutamicum*), with the protein of interest appearing to be the predominant component in elutions 5 and 6. Dialysis was subsequently performed on these fractions, the purpose of which is to exchange the purification/elution buffer (50 mM Na₃PO₄, 300 mM NaCl, 100 to 300 mM imidazole depending on elution concentration) for a previously described reaction buffer (20 mM Tris-HCl, pH 7.5, 0.2 M NaCl, 10% glycerol - Gao et al., 2008). All other SDS-PAGE analyses of purifications are contained in appendix 9.6.

All proteins were successfully purified, including LldR-KRAB (*Corynebacterium glutamicum*).

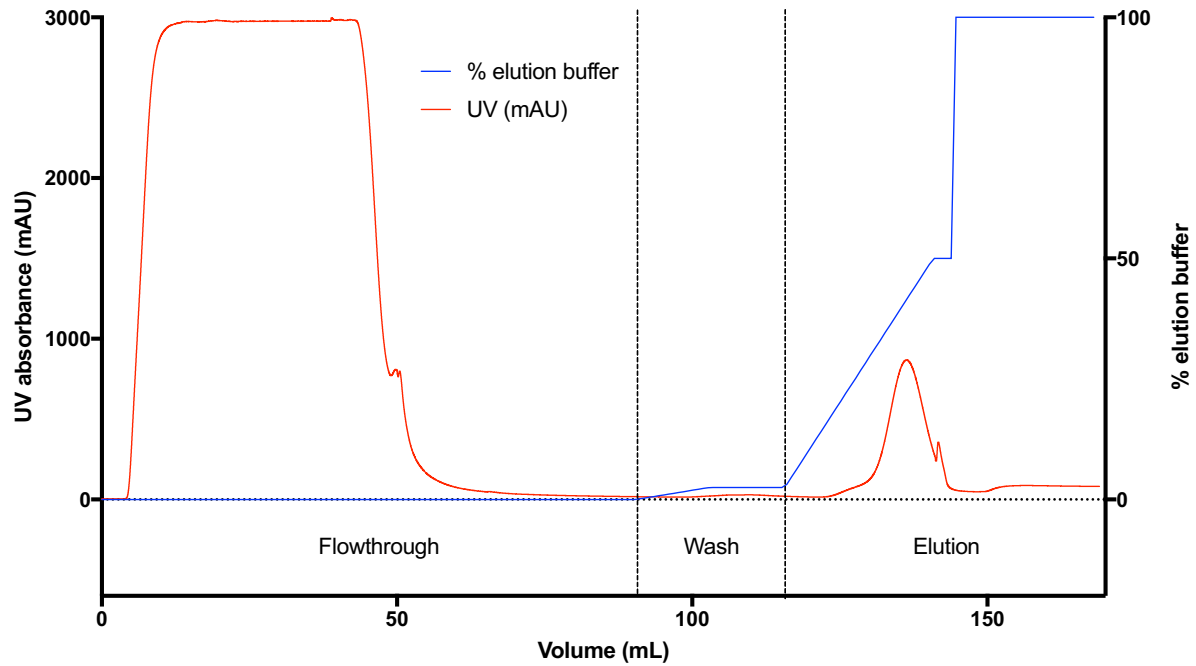


Figure 4.3. Chromatogram produced during the purification of LldR (*Corynebacterium glutamicum*). The x-axis shows the volume of buffer progressively passed through the nickel-NTA column. The left y-axis shows the ultraviolet (UV) absorbance, given in milli-arbitrary units (mAU), which is indicative of protein content. The right y-axis shows the percentage of the elution buffer (containing 500 mM imidazole) present in the buffer being passed through the nickel-NTA column. The flowthrough, wash and elution stages of the purification run are demarcated in the figure by the vertical dotted lines. The large initial UV absorbance peak represents the unbound portion of the cleared cell lysate. This is followed by a wash stage with a buffer containing a low amount of imidazole. Elution of the protein of interest is achieved with a higher percentage of the elution buffer.

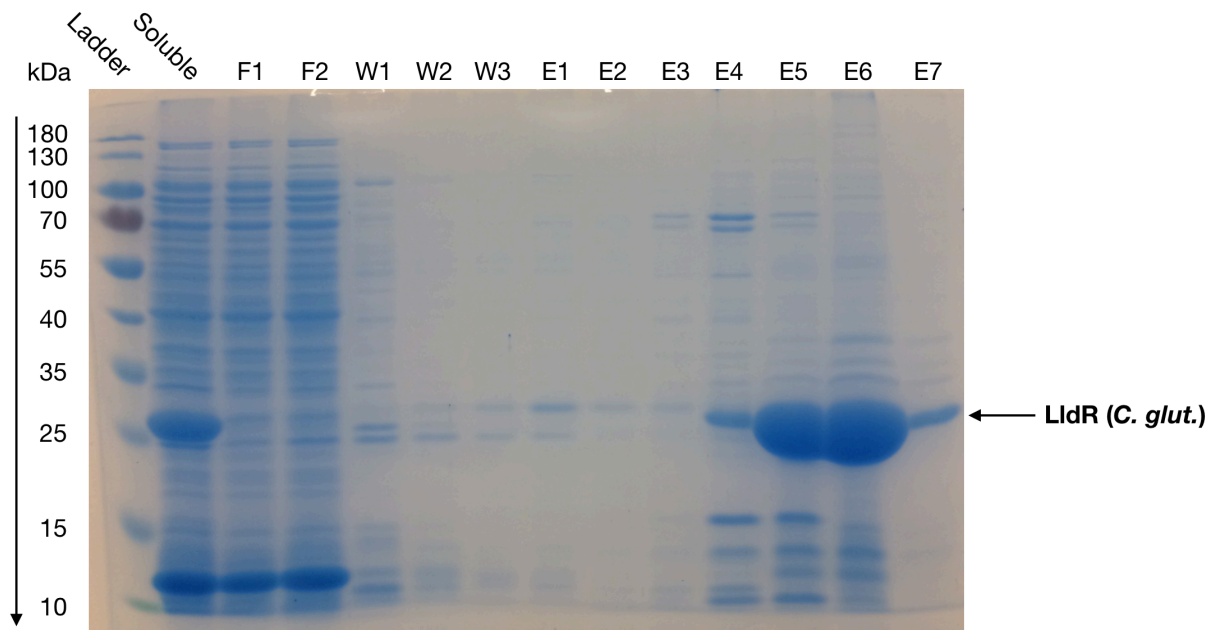


Figure 4.4. SDS-PAGE analysis of the fractions collected during the purification of LldR (*Corynebacterium glutamicum*). The soluble fraction shows the presence of LldR at approximately 26 kDa (indicated by the leftmost arrow), which is not seen in either of the flowthrough fractions (F1 and F2). LldR is released at high levels into the fifth and sixth elution fractions (E5 and E6). These two fractions are pooled and dialysed into a reaction buffer. F refers to flowthrough, W refers to a wash step, and E refers to an elution step.

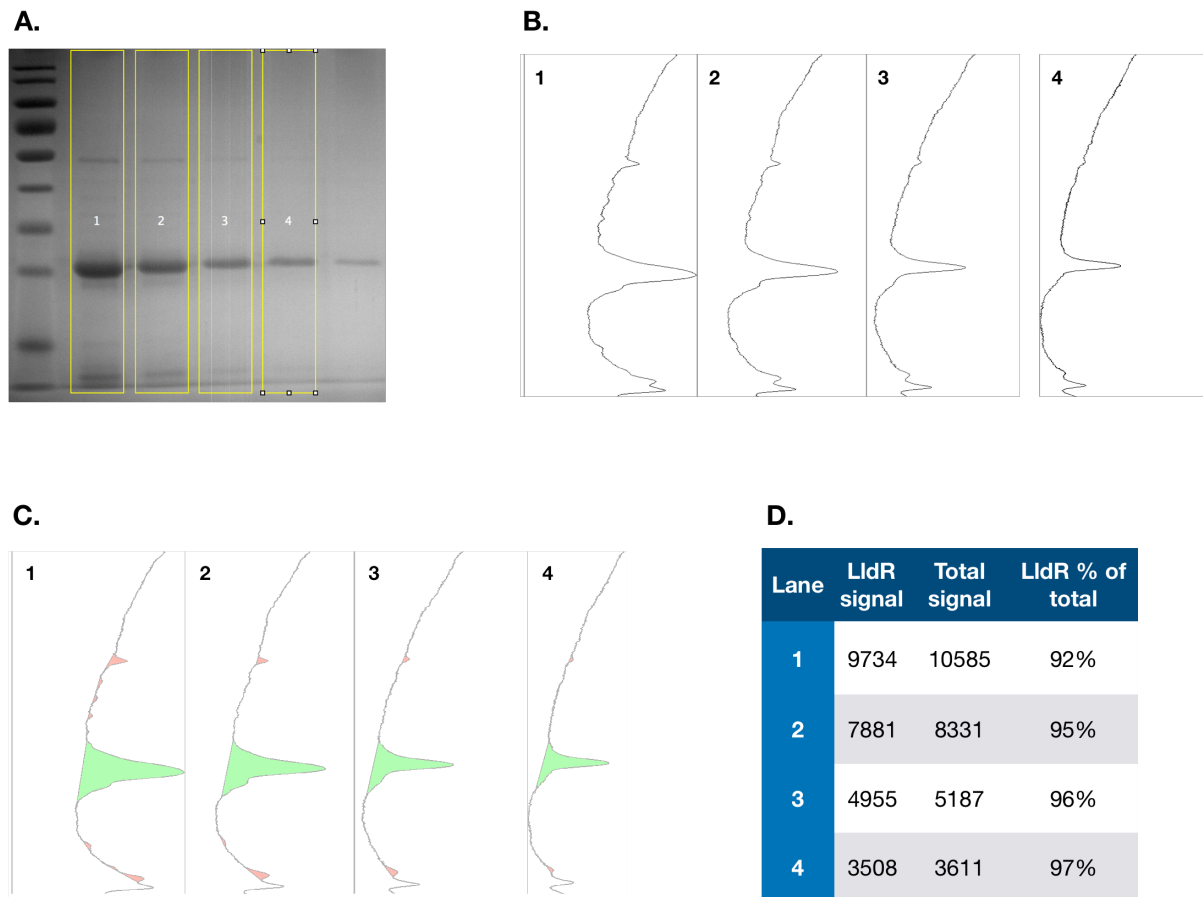


Figure 4.5. Using densitometry to determine the purity of LldR (*Corynebacterium glutamicum*). A. The sample was visualised using SDS-PAGE and analysed using ImageJ (Schneider et al., 2012). A box was drawn around 4 lanes corresponding to sequentially diluted LldR samples, in order to increase the accuracy of purity determination. B. These boxes then returned a value giving the intensity of the signal (arbitrary units) along the y-axis of the boxes. C. Lines were drawn across each peak to exclude the background signal, with the green shaded area showing the putatively LldR-containing region, and the red shaded areas showing the contaminating bands. D. Based on the signals given from LldR and the contaminating bands, the purity of LldR can be calculated.

4.4.3 Densitometry for determination of protein purity and extent of protein-DNA binding

Densitometry is an approach for quantifying the signal given by a band from an electrophoresis experiment. By assessing the visual signals observed within a lane, densitometry can be performed on SDS-PAGE images to quantify the signal given not only by the protein of interest but also that of contaminating bands, giving a measure of purity. Additionally, when DNA is

visualised in an EMSA, the signal from the free species can be compared to the bound species to determine the extent to which protein-DNA binding was seen.

Table 4.2. Estimated purity of LldR fusion variants isolated from *Escherichia coli*.

Construct name	Purity
LldR (<i>Corynebacterium glutamicum</i>)	96%
LldR-KRAB (<i>Corynebacterium glutamicum</i>)	54%
LldR-VP64 (<i>Corynebacterium glutamicum</i>)	78%
VP64-LldR (<i>Corynebacterium glutamicum</i>)	56%
LldR-15aa-VP64 (<i>Corynebacterium glutamicum</i>)	64%
LldR (<i>Pseudomonas aeruginosa</i>)	94%
LldR-VP64 (<i>Pseudomonas aeruginosa</i>)	87%
LldR-15aa-VP64 (<i>Pseudomonas aeruginosa</i>)	88%

Figure 4.5 shows the process of determining the signal from an SDS-PAGE assay using ImageJ (Schneider et al., 2012), where a concentrated sample of LldR (*Corynebacterium glutamicum*) has been run in a two-fold dilution series. Figure 4.5 A shows the initial selection of areas within a lane where protein bands are found. Upon using the “plot the lanes” function within ImageJ, the signals seen for each area are seen along the y-axis, with strong peaks typically corresponding to the presence of protein (figure 4.5 B). Lines were drawn across these peaks as seen in figure 4.5 C, in order to distinguish the background signal from the protein signal. The areas of these peaks were calculated to determine the signal from LldR and the total seen from all bands, allowing the apparent purity of the major LldR band to be quantified (figure 4.5 D). Although not necessary in this particular case, multiple dilutions of the sample were run in order to ensure that the signal from the LldR band was not saturated. The same process is applied to EMSA analysis, although DNA was visualised in this case. Performing these experiments confirmed the high purity of the LldR variants tested here, and made it possible to normalise the concentrations used in subsequent EMSA experiments.

4.4.4 Description of the EMSA experiments

The electrophoretic mobility shift assay (EMSA) is a rapid, sensitive and inexpensive method for investigating protein-DNA binding (Hellman & Fried, 2009). It works on the basis that DNA complexed to protein has a lower electrophoretic mobility compared to unbound DNA, meaning that both species of DNA will migrate differently when passed through a polyacrylamide or agarose gel. Having previously been used in multiple papers describing the interactions of LldR (from *Corynebacterium glutamicum* and *Pseudomonas aeruginosa* with its operator DNA (Georgi et al., 2008, Gao et al, 2008; Gao et al., 2012), it was decided to use EMSAs to investigate the impact of varying operator and protein fusion characteristics on binding behaviour.

To perform each EMSA, the protein was typically incubated for 1 hour at room temperature with double-stranded operator DNA in a binding buffer (described by Gao et al., 2008). Afterwards, the mixture was separated at 4 °C using native polyacrylamide gel electrophoresis. The amount of LldR protein used was based on the estimated purity information of the sample (an example of which is found in figure 4.5) and the colorimetric DC protein assay (Bio-Rad).

4.4.4.1 Testing the affinity of LldR (*Corynebacterium glutamicum*) to two single operator repeats

The strength of protein-DNA binding is highly sensitive to changes in the operator sequence. For example, Gao et al. (2008) showed that cgl2917 and cgl1934, two natural operators of LldR (*Corynebacterium glutamicum*), had dissociation constants of 81 and 1700 nM respectively, despite differing by only one nucleotide. Any operator used in an inducible transgene expression system must therefore be carefully designed, as the tighter the binding between a biosensor and its operator, the more likely it will be able to exert some effect on an adjacent promoter.

The operator (named 1xOperator here for convenience) used in the previous chapter differed slightly from the cgl2917 operator found in the genome of *Corynebacterium glutamicum* and used in the ITC experiments by Gao et al. (2008). Shown in bold in figure 4.6, the differing bases come only at the ends of the sequences and were not within a DNase-protected region in experiments by Gao et al., indicating that they were unlikely to be involved in the protein-

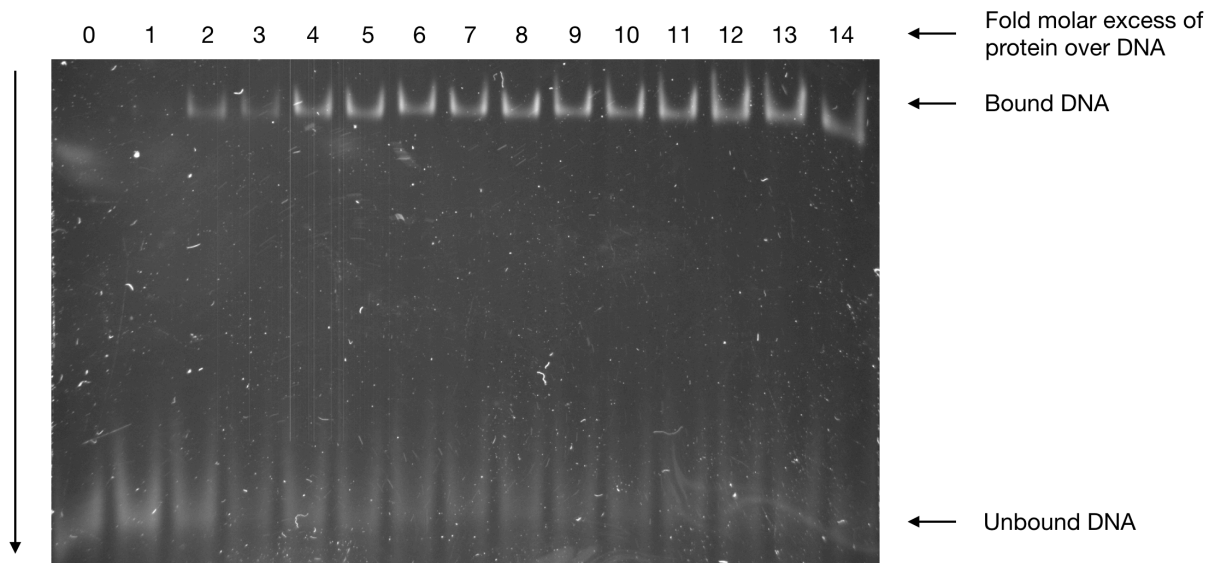


Figure 4.7. EMSA showing *cgl2917* incubated with increasing concentrations of LldR (*Corynebacterium glutamicum*). 50% of the total DNA appeared to be bound as the fold molar excess of protein reached 3 to 4.

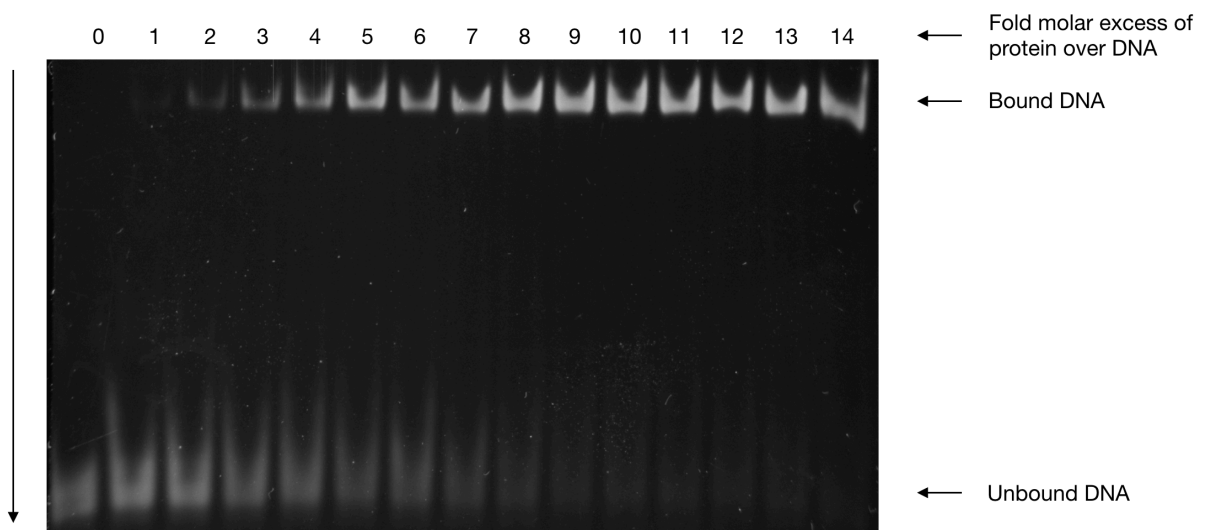


Figure 4.8. EMSA showing 1xOperator incubated with increasing concentrations of LldR (*Corynebacterium glutamicum*). 50% of the total DNA appeared to be bound as the fold molar excess of protein reached 5 to 6.

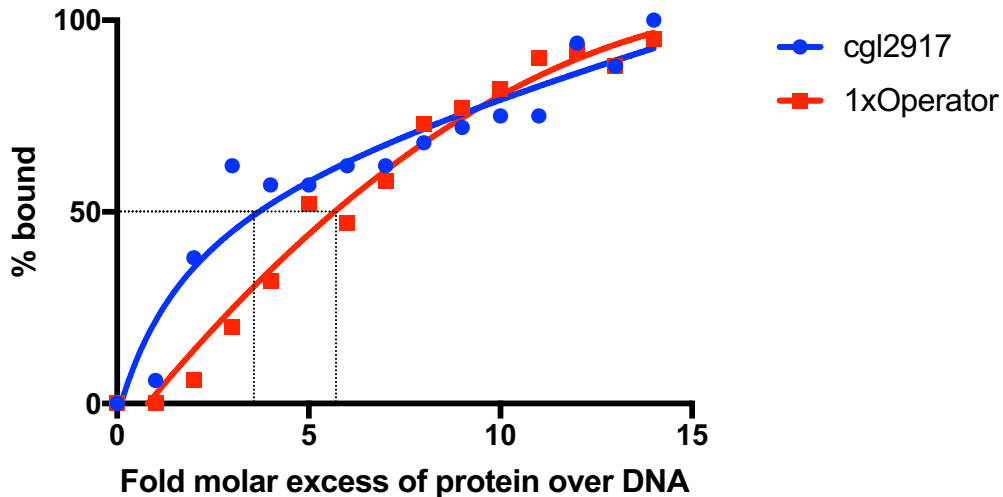


Figure 4.9. Analysis of the binding of LldR (*Corynebacterium glutamicum*) to two similar operators, 1xOperator and cgl2917. Fitting a curve to the data points allows an estimation of the fold molar excess (FME) of LldR needed for 50% of the operator DNA to be bound. For cgl2917, this point is found at an FME of 3.6; for 1xOperator, this point is found at an FME of 5.8. Data is taken from densitometry processing of gel electrophoresis experiments.

4.4.4.2 Effect of inter-operator spacing on binding affinity

To maximise the probability that LldR will be adjacent to a promoter (on which it can exert some effect on transcription), multiple operator repeats are often used together (e.g. Gossen & Bujard, 1992; Hartenbach et al., 2007). These operator repeats must be sufficiently spaced from each other, or steric hindrance can occur with a concomitant reduction in apparent binding affinity. This is seen with one of the natural promoter targets of LldR in *Corynebacterium glutamicum*, where a single cgl2917 operator has a dissociation constant of 81 nM, but the addition of an insufficiently spaced downstream operator increases the dissociation constant to 1100 nM (which corresponds to a weaker binding interaction overall). As a potential reason for the non-inducing behaviour of the previously tested system, the effect of spacing between two operator repeats on LldR-operator binding affinity was investigated.

Figure 4.10 shows the spacing between two operators. 2xOperatorA has 10 bases between the two operator repeats, and is the same as that used in the previous chapter. 2xOperatorB has a total of 22 bases between the two operator repeats. Figure 4.11 shows the results of an EMSA where 2xOperatorA was incubated with LldR (*Corynebacterium glutamicum*), and

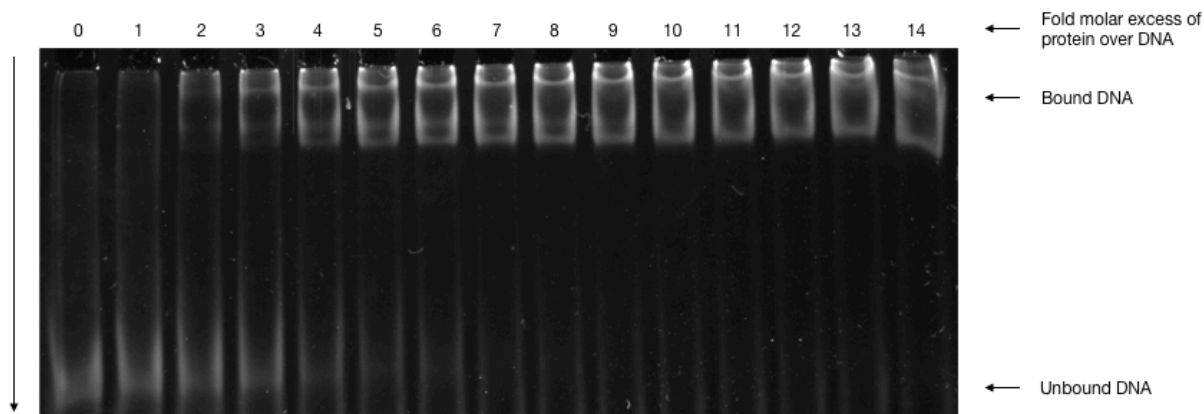


Figure 4.12. EMSA showing 2xOperatorB incubated with increasing concentrations of LldR (*Corynebacterium glutamicum*). 50% of the total DNA appeared to be bound as the fold molar excess of protein reached 2 to 3.

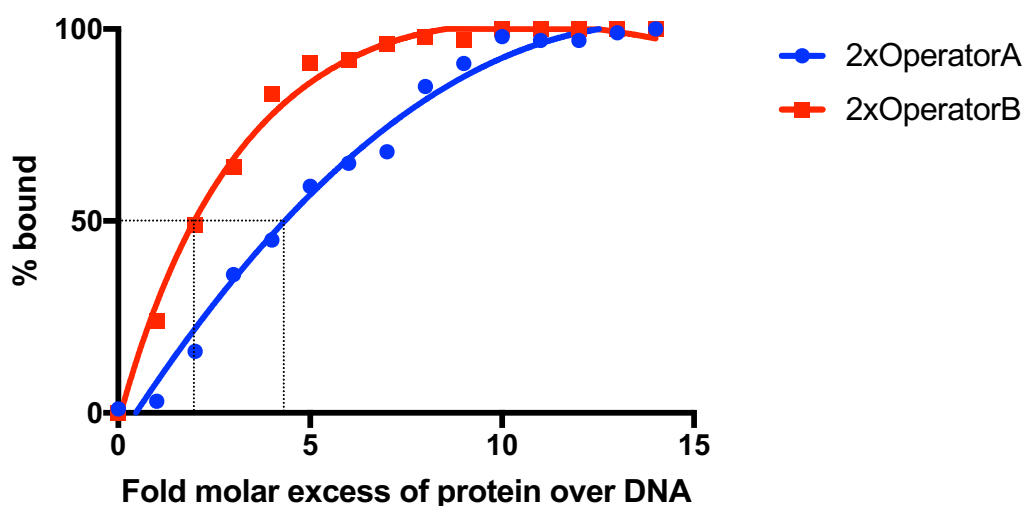


Figure 4.13. Analysis of the binding of LldR (*Corynebacterium glutamicum*) to 2xOperatorA and 2xOperatorB. Fitting a curve to the data points allows an estimation of the fold molar excess (FME) of LldR needed for 50% of the operator DNA to be bound. For 2xOperatorA, this point is found at an FME of 4.4; for 2xOperatorB, this point is found at an FME of 2. Data is taken from densitometry processing of gel electrophoresis experiments.

4.4.4.3 Determination of the impact of effector domain orientation on DNA binding affinity

An effector domain fused to a transcriptional biosensor is necessary for the implementation of trans-activating and trans-silencing configurations. It has been suggested that any effector

domain should be fused away from the DNA-binding domain of the transcriptional biosensor (Karlsson et al., 2011), and there is an example where a fusion at the N-terminus of a GntR-family protein (of which LldR is a member) was no longer able to induce a previously functional transgene expression system (Folcher et al., 2013). The impact of effector domain fusions to LldR (from *Corynebacterium glutamicum* and *Pseudomonas aeruginosa*) at both termini was tested. Whether the commonly used flexible linker (GGGGS)₃ could mitigate the impact of an effector domain fusion was also tested. This linker is referred to as “15aa” in subsequent figures. A linker is often used between two fused protein domains in order to prevent them from misfolding or sterically hindering the activity of the other (Chen et al., 2013).

4.4.4.3.1 Variants of the *Corynebacterium glutamicum* LldR protein

Figure 4.7 shows the EMSA image of LldR (*Corynebacterium glutamicum*) incubated with the cgl2917 operator. Figures showing the incubation of cgl2917 with LldR-VP64, VP64-LldR, LldR-15aa-VP64 and LldR-KRAB are located in appendix 9.8. Figure 4.14 is a summary of this EMSA data, showing that an effector fusion of VP64 to LldR hinders its ability to bind to cgl2917 (the FME for 50% occupancy of LldR for cgl2917 is 3.6; the FME for 50% occupancy of LldR-VP64 for cgl2917 is 16.6, approximately 4.6 times weaker than without the fusion). This C-terminal fusion is still preferable to an N-terminal fusion, which is no longer capable of binding to cgl2917 at the FME tested here.

The use of a 15aa linker can restore some of LldR's ability to bind to cgl2917 (the FME of LldR-15aa-VP64 for 50% occupancy is 13.0, a percentage difference of ~25%), and so this linker was used with LldR (*Corynebacterium glutamicum*) in the subsequent testing of an *in vivo* lactate-inducible transgene expression system in chapter 5.

LldR-KRAB was unstable after purification, and precipitated prior to its use in subsequent EMSA experiments. On the basis that it could not easily be purified and tested, no attempt was made to repurify this construct, nor to advance it to *in vivo* testing in an optimised inducible system.

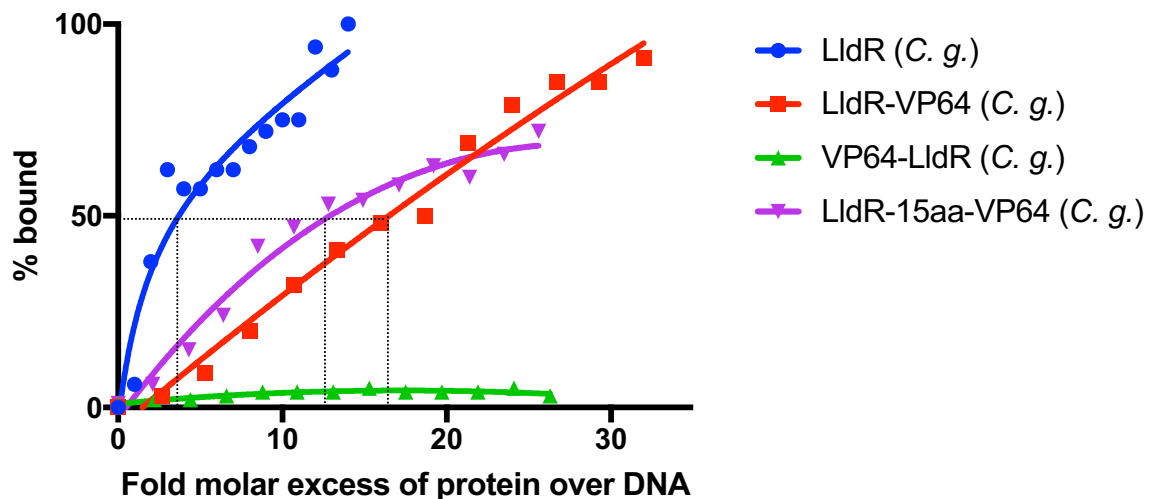


Figure 4.14. Analysis of the binding of LldR (*Corynebacterium*) and fusion variants to cgl2917. The FME required for LldR to bind 50% of the available operator DNA was 3.6, which serves as a baseline to compare the impact of effector fusion on the strength of LldR-operator binding. The addition of the C-terminal VP64 the FME for 50% occupancy to 16.6, with an N-terminal fusion of VP64 abolishing the protein-DNA interaction entirely. The use of a flexible linker (15aa) restores some of the DNA-binding capacity of LldR, evident in the FME for 50% occupancy of 13.0. Data is taken from densitometry processing of gel electrophoresis experiments.

4.4.4.3.2 *Pseudomonas aeruginosa* variants

The LldR from *Pseudomonas aeruginosa* was chosen as an alternative to that of *Corynebacterium glutamicum*, although it was not preferred initially owing to a relative lack of structural and biochemical information. It also binds to a very similar operator, differing only by a single base between the palindromic half-sites (identified as TGGTCTTACCA by Gao et al., 2012, compared to the TGGTCTGACCA identified by Georgi et al., 2008). Although DNA-binding character of proteins can be very sensitive to changes in its target operator, the similarity of the operator of the *Pseudomonas aeruginosa* LldR to cgl2917 means that it was relatively straightforward to test alongside LldR from *Corynebacterium glutamicum*. Given the previous difficulty with expressing and purifying KRAB, there was no attempt to use it with as a fusion domain with LldR from *Pseudomonas aeruginosa*.

All EMSA images of LldR (*Pseudomonas aeruginosa*) and fusion variants with cgl2917 are contained in appendix 9.8. A summary of this EMSA data is given in figure 4.15. Interestingly,

it appears that LldR (*Pseudomonas aeruginosa*) binds with greater strength to cgl2917 than LldR (*Corynebacterium glutamicum*) itself, with the FME of 3.2 required for 50% occupancy of cgl2917. The fusion of a C-terminal VP64 had a limited impact on this FME, shifting it to 3.9; additionally, the use of a flexible linker had limited impact on the FME, increasing it to 4.2. On this basis, it was decided to subsequently test LldR and LldR-VP64 from *Pseudomonas aeruginosa* (as well as from *Corynebacterium glutamicum*) in an *in vivo* system.

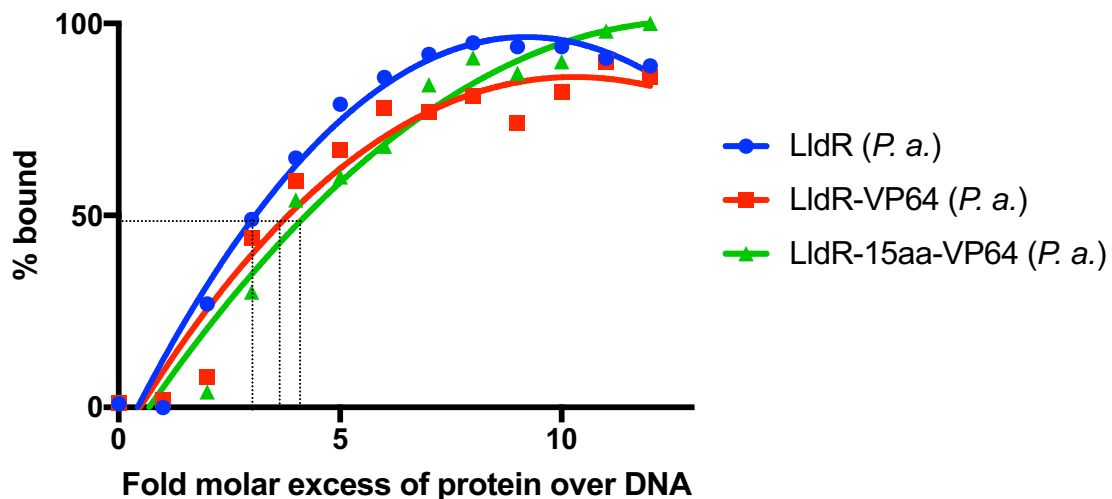


Figure 4.15. Analysis of the binding of LldR (*Pseudomonas aeruginosa*) and fusion variants to cgl2917. The FME required for LldR to bind 50% of the available operator DNA was 3.2, which indicates slightly stronger binding to cgl2917 than seen with LldR (*Corynebacterium glutamicum*). The addition of a C-terminal VP64 slightly shifted the FME for 50% occupancy to 3.9; the use of a 15aa flexible linker between LldR and VP64 increased the necessary FME even further to 4.2, suggesting that it is preferable not to use such a linker in this instance. Data is taken from densitometry processing of gel electrophoresis experiments.

4.4.4.4 Effector domain fusion impact on lactate unbinding capability.

The ability of LldR proteins to unbind from their operator in the presence of lactate is fundamental to their usage in an inducible transgene expression system, both in the natural context and for heterologous use in Chinese hamster ovary cells. Thus, it is important to determine whether the presence of any effector domains could have an impact on the extent to which this unbinding occurs. Additionally, the current data on the amount of lactate necessary to induce this unbinding is vague, stated as occurring at between 20 and 40 mM (Georgi et al., 2008), and so it was intended to investigate this more thoroughly. However, the

extent to which lactate-induced unbinding of LldR (both from *Corynebacterium glutamicum* and *Pseudomonas aeruginosa*) from cgl2917 was seen here did not match the previously published data.

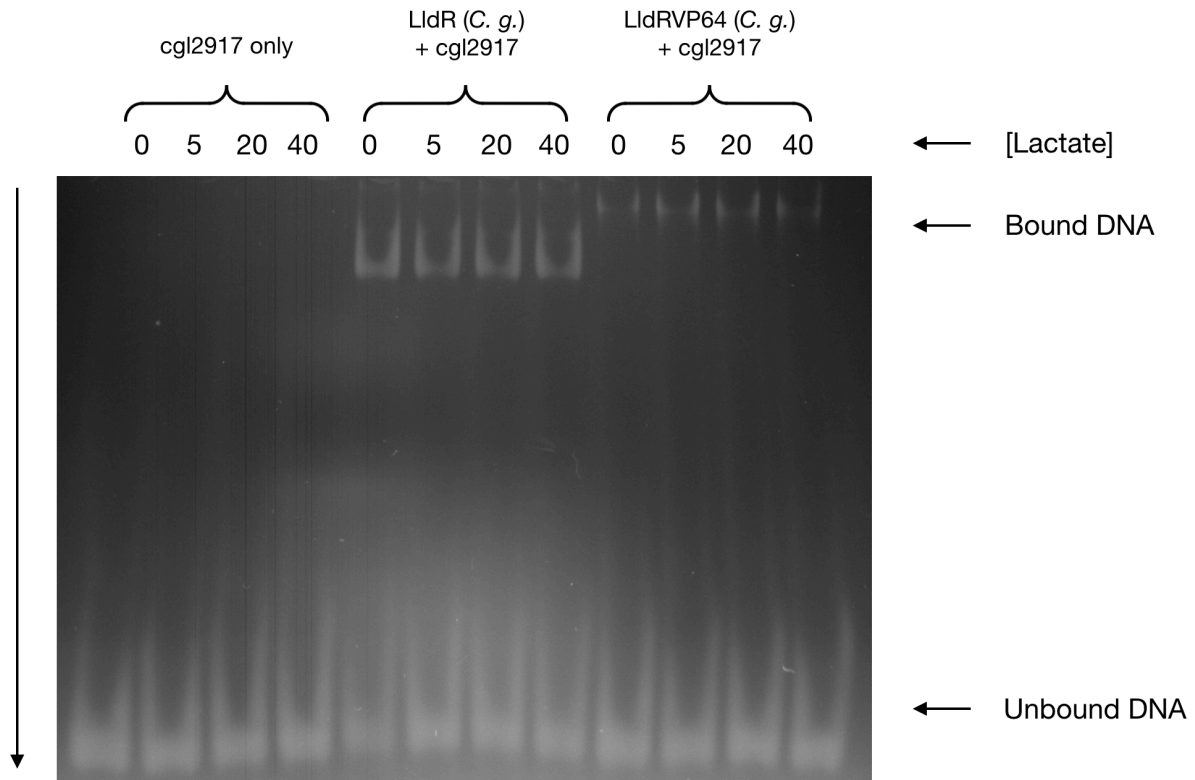


Figure 4.16. EMSA showing the incubation of cgl2917 with a 4-fold molar excess of LldR and 10-fold molar excess LldR-VP64 (both *Corynebacterium glutamicum*) and either 0, 5, 20 or 40 mM of lactate. No unbinding in the presence of lactate was apparent.

An initial attempt was made to show lactate-induced unbinding of LldR and LldR-VP64 (both *Corynebacterium glutamicum*) from cgl2917 in the presence of 5, 20 and 40 mM lactate (figure 4.16). No obvious increase in the free species of cgl2917 was seen. This experiment was then repeated with higher levels of lactate, ranging from 0, 20, 40 and 80 mM of lactate (figure 4.17). Additionally, the EMSA as carried out by Georgi et al. (2008) was followed more closely, using a 20-fold molar excess of LldR over cgl2917 as in their experiment (previously a 4-fold molar excess for LldR and a 10-fold molar excess for LldR-VP64 were used, as it was clear that a significant amount of binding would occur at this ratio). There is some evidence that unbinding in the presence of lactate did occur, with an excess of 20 mM of lactate leading to a greater proportion cgl2917 found as free DNA compared to the no lactate conditions for both LldR and LldR-VP64. Nonetheless, this partial unbinding did not match the behaviour of LldR

in the Georgi et al. paper, where almost complete unbinding in the presence of lactate was obvious (see figure 4.18).

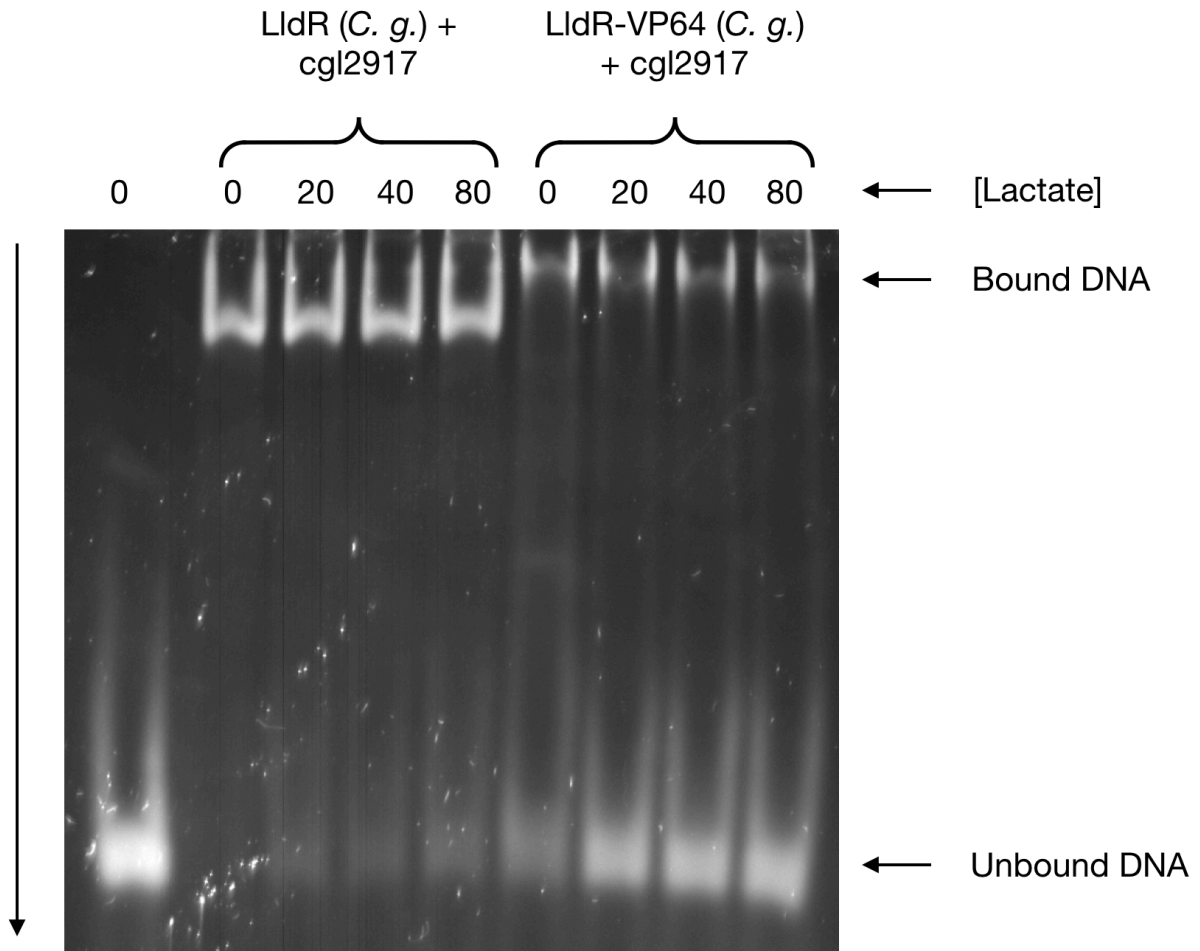


Figure 4.17. EMSA showing the incubation of cgl2917 with a 20-fold molar excess of LldR and LldR-VP64 (both from *Corynebacterium glutamicum*) and either 0, 20, 40 or 80 mM of lactate. The proportion of cgl2917 running as unbound did increase in the presence of lactate in this experiment, although still not to the extent seen in the original paper by Georgi et al., 2008.

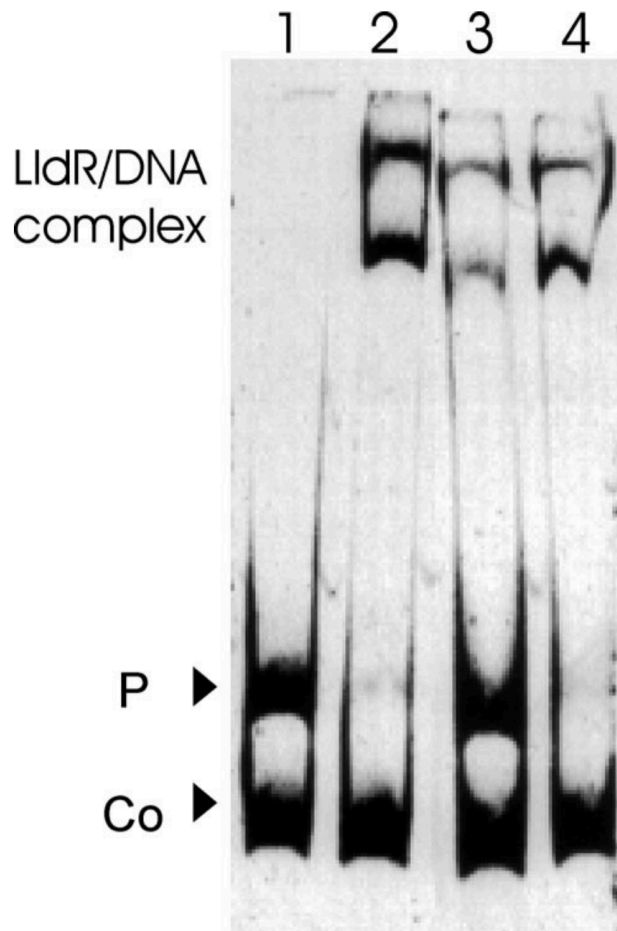


Figure 4.18. EMSA from Georgi et al. (2008) showing unbinding of LldR (*Corynebacterium glutamicum*) from a fragment containing the operator sequence in the presence of lactate. In this case, the DNA has been visualised. Lane 1 shows the operator incubated without any protein. Lane 2 shows the operator in the presence of LldR, with the operator appearing to bind to LldR. Lane 3 shows the operator in the presence of LldR and 40 mM L-lactate, with the majority of the LldR no longer able to form a complex with the operator. Lane 4 shows the operator in the presence of LldR and 40 mM D-lactate, with no unbinding apparent. The “P” indicates the running location of the unbound operator DNA. The “Co” indicates the running location of a control DNA that is not bound by LldR. LldR was present at a 20-fold molar excess in lanes 2 to 4. Annotations are as in the original figure. Figure is reproduced here with permission from the copyright holders, the American Society for Microbiology, detailed in appendix 9.14.

In a further attempt to replicate the unbinding behaviour of LldR observed by Georgi et al., much higher levels of lactate were tested (0, 20 and 200 mM), as well as testing LldR-VP64 in the same buffer as used in their experiments (50 mM Tris-HCl, 10% glycerol, 50 mM KCl, 10 mM MgCl₂, 0.5 mM EDTA; pH 7.5). The reaction buffer used otherwise was 20 mM Tris-

HCl, 10% glycerol, 0.2 M NaCl, 1 mM dithiothreitol, at pH 7.5, as used by Gao et al., 2008. The use of the very high 200 mM lactate did boost the extent of unbinding slightly (see figure 4.19), although it was still not complete. The use of the same buffer as in the Georgi experiment also did not replicate their result.

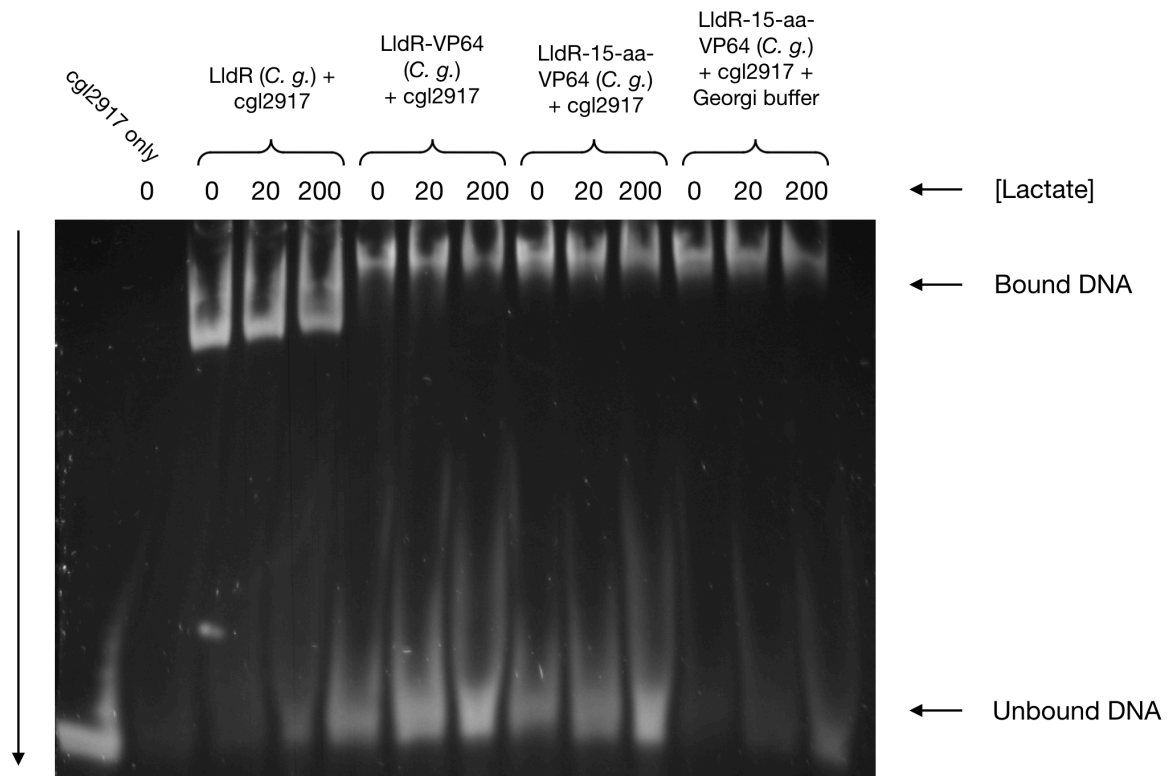


Figure 4.19. EMSA of LldR, LldR-VP64 and LldR-15aa-VP64 (all from *Corynebacterium glutamicum*) incubated with 0, 20 and 200 mM of lactate. In addition, the same buffer as that used by Georgi et al. (2008) was employed here. Although a partial shift of cgl2917 towards an unbound state was seen in the presence of lactate, this was still not to the same extent as seen in the Georgi paper, in any of the tested conditions.

This unexpectedly weak unbinding from cgl2917 in the presence of lactate was also true of the fusion variants based on LldR from *Pseudomonas aeruginosa*. Figure 4.20 shows these variants incubated with cgl2917 in the presence of 0, 20 and 200 mM lactate, with only small shifts of the bound cgl2917 to the unbound form being seen. Figure 4.21 shows the expected behaviour from Gao et al. (2012), where almost complete unbinding is seen in the presence of 40 mM lactate.

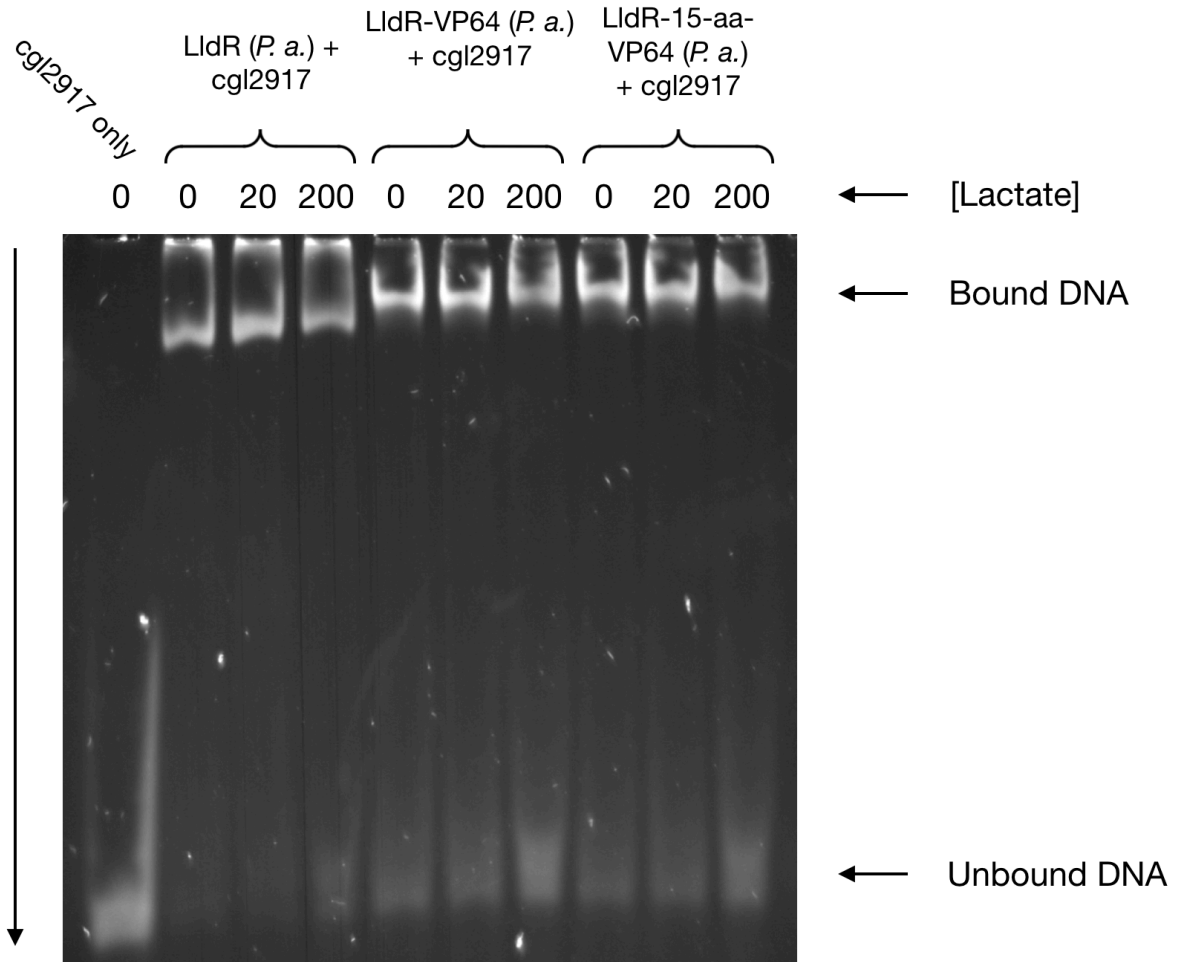


Figure 4.20. EMSA showing the incubation of LldR, LldR-VP64 and LldR-15aa-VP64 (all from *Pseudomonas aeruginosa*) with cgl2917 in the presence of 0, 20 and 200 mM of lactate. Although some unbinding in the presence of 200 mM lactate was apparent, the results here indicate much weaker unbinding behaviour than seen in the Gao et al. (2012) paper.

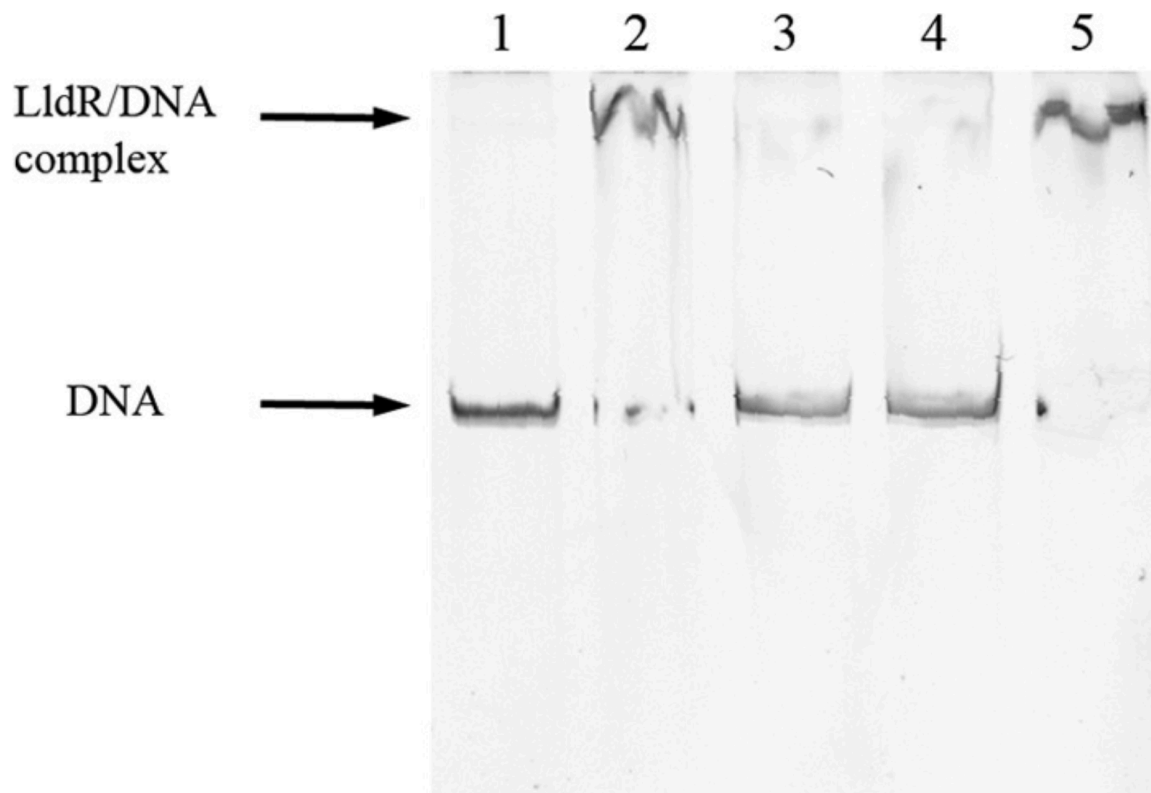
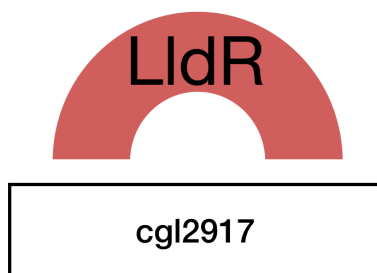


Figure 4.21. EMSA from Gao et al. (2012) showing the unbinding of LldR (*Pseudomonas aeruginosa*) from its operator in the presence of lactate. In this case, the DNA has been visualised. Lane 1 shows the operator DNA incubated without any protein. Lane 2 shows the incubation of operator DNA with LldR, with significant levels of binding apparent. Lane 3 shows the incubation of operator DNA with LldR and 40 mM D-lactate, with the LldR-operator unable to form. Lane 4 shows the incubation of operator DNA with LldR and 40 mM L-lactate, with the LldR-operator also unable to form. Lane 5 shows the incubation of operator DNA with LldR and 40 mM pyruvate, which had no impact on LldR-operator binding. LldR was present at a 10-fold molar excess in lanes 2 to 5. Annotations are as in the original figure. Figure is reproduced here with permission from the copyright holders, the American Society for Microbiology, detailed in appendix 9.14.

There are two differences between the experiments performed here and those of Georgi et al. (2008) and Gao et al. (2012). The LldR variants expressed here used a C-terminal fusion of the 6xHis-tag to help with purification, while it is likely that an N-terminal fusion of a His-tag was used by Georgi and Gao (it is not stated explicitly in either case but it would be the outcome of the cloning processes they describe). It is possible that the C-terminal fusion of the 6xHis-tag could interfere with ligand binding and consequent unbinding from the operator, as the lactate-binding domain is found at the C-terminus. Although instances of 6xHis-tag

interference with protein function have been reported previously (e.g. Majorek et al., 2014) it is considered rare given the small size of the tag (Crowe et al., 1994). Additionally, it is not clear how a C-terminal 6xHis-tag could interfere with lactate binding, given the polarity of histidine and that hydrophobic interactions between the ligand-binding domain of LldR and lactate predominate (Gao et al., 2008). Another difference is that the previously reported unbinding behaviour of LldR was from larger fragments of DNA, unlike the 20 bp cgl2917 tested here. For instance, Georgi et al. (2008) used a 331-bp fragment containing the cgl2917 sequence, while Gao et al. (2012) used a 136-bp fragment containing a similar operator sequence. Both were amplified from genomic DNA by PCR. Although there is no clear mechanism by which LldR would not demonstrate complete unbinding from a short operator-containing DNA fragment in contrast to a longer one, it is at least an obvious difference between the experimental approach in this chapter relative to that of Georgi et al. (2008) and Gao et al. (2012) and so is worth further investigation on this basis.

a.



b.



Figure 4.22. The cgl2917 operator tested in this *in vitro* optimisation chapter (panel a) and its use in subsequent *in vivo* experiments (panel b).

4.5 Summary of results; discussion

Table 4.3. Summary of fold molar excesses (FME) of LldR variants required to bind 50% of the available cgl2917 operator. VP64 refers to the VP64 transcriptional activator. 15aa refers to a 15-amino acid flexible protein linker. Asterisks indicate proteins that were used in subsequent *in vivo* testing in chapter 5.

Protein variant	FME
LldR (<i>Corynebacterium glutamicum</i>)*	3.6
LldR-VP64 (<i>Corynebacterium glutamicum</i>)	16.7
VP64-LldR (<i>Corynebacterium glutamicum</i>)	No binding observed
LldR-15aa-VP64 (<i>Corynebacterium glutamicum</i>)*	13.0
LldR (<i>Pseudomonas aeruginosa</i>)*	3.2
LldR-VP64 (<i>Pseudomonas aeruginosa</i>)*	3.9
LldR-15aa-VP64 (<i>Pseudomonas aeruginosa</i>)	4.2

Apart from LldR-KRAB and KRAB-LldR (both *Corynebacterium glutamicum*), it was possible to clone, express, purify and test all the desired constructs.

It was possible to determine the relative binding affinities of LldR and its effector fusion variants to various operator sequences (summarised in table 4.3). It was found that the use of cgl2917 was preferable to the operator used in the initial *in vivo* testing chapter. Additional spacing between multiple operator repeats was also found to improve its binding affinity for LldR. These variations in the operator design were employed in a subsequently tested *in vivo* lactate-inducible transgene expression system (see figure 4.22).

In keeping with the suggestion from Karlsson et al. (2011) that any effector domain fusion to a transcriptional biosensor should be away from the DNA-binding domain, fusion of VP64 to the N-terminal domain of LldR abolished any DNA-binding capacity of LldR. In subsequently tested systems, VP64 was fused to the C-terminus of LldR. The use of a flexible (GGGS)₃ fusion linker between the LldR from *Corynebacterium glutamicum* and VP64 also restored some of the DNA-binding capacity that was otherwise hindered by the presence of the fusion domain. VP64 did not seem to significantly impact the DNA-binding capacity of LldR from

Pseudomonas aeruginosa, and so the use of a flexible linker was not deemed necessary in this case.

Only partial unbinding of LldR from cgl2917 in the presence of lactate was apparent, in contrast to the near complete unbinding seen in the original work characterising this behaviour (Georgi et al., 2008; Gao et al., 2012). Two possible explanations for this have been given - the use of a 6xHis tag at the C-terminus here may be interfering with lactate-binding (an N-terminal fusion was used in the previously cited work); or the short length of the cgl2917 DNA fragment used here is having some impact on the ability of LldR to unbind from this operator in the presence of lactate. Although this lactate-dependent binding of LldR to operator DNA is of fundamental importance to its use in an inducible transgene expression system, the operator used in the *in vivo* system is of course part of a much larger DNA sequence (thus removing this as a possible concern for subsequent testing), while LldR can be tested without any effector domain fusions, also removing this as a possible concern in particular test conditions. It may still be the case that the presence of C-terminal fusions to LldR are hindering its ability to unbind from operator DNA in the presence of lactate, introducing a trade-off to their usage in an *in vivo* system; however, in previous work on another GntR-type protein, VanR, a C-terminal effector domain fusion had no such impact (Gitzinger et al., 2012, Folcher et al., 2013).

5 Testing of optimised *in vivo* constructs in CHO cells

5.1 Chapter aims

- The determination of an efficient approach to plasmid construction, allowing a sufficiently broad number of conditions to be tested (around 78 in total)
- To test de-repression and transactivation methods of lactate-inducible transgene expression in the following ways:
 - Using two plasmids, one encoding LldR variants, and the other encoding an appropriate response element, expressed via transient transfection
 - Using a single plasmid encoding both LldR variants and the response element, expressed via transient transfection

5.2 Chapter summary

In vitro testing of LldR fusion variants to a modified operator in the previous chapter meant that these optimised components could be employed in a subsequent round of *in vivo* testing in Chinese hamster ovary cells. Transient transfections found multiple functional lactate-inducible transgene expression systems, with a fold-induction of 3.46 for the top performing configuration, and the number of conditions tested here meant that certain variables could be compared to one another. In particular, the use of multiple operators was preferred; upstream operators gave high fold-induction ratios than downstream operators; minCMV outperformed YB_TATA; and LldR from *Corynebacterium glutamicum* was slightly more effective than LldR from *Pseudomonas aeruginosa*.

5.3 Cloning strategy for generation of optimised *in vivo* constructs in CHO cells

As in chapter 3, a straightforward approach for assembling multiple constructs with numerous variable elements was essential. Each combination of biosensor and responsive promoter construct could be tested as two co-transfected plasmids, and with both constructs on a single plasmid, as this can control for any discrepancies in plasmid uptake efficiency during transfection. In order to generate all of these constructs in a timely manner, the MXS Chaining approach (Sladitschek & Neveu, 2015) described in figure 3.1 was used in combination with Gibson assembly. Gibson assembly was used to rapidly and scarlessly assemble LldR with

fusion linkers, effector domains and the constitutive CMV promoter. MXS chaining was used to assemble the responsive elements in a stepwise fashion, allowing constructs such as minCMV-2x-mAzamiGreen-pA to be assembled in two rounds of cloning; this stepwise, restriction enzyme-based approach to construction of the responsive elements (and their combination with biosensor constructs) was considered necessary, as the palindromic nature of the operator repeat used means it would not be amenable to PCR-based cloning approaches such as Gibson assembly.

Once both the biosensors and responsive promoters were assembled on separate plasmids, it was possible to test their lactate-responsiveness by co-transfecting them in various combinations transiently in CHO-S cells, while simultaneously performing an additional round of cloning to assemble these plasmids together in order to perform the same experiments with single plasmids, in order to avoid any potential confounding impacts of plasmid uptake efficiency mismatches. These plasmid constructs are listed in table 9.6 in the appendix. A summary of variables relating to the lactate-inducible system itself are given in table 5.1.

This overall cloning approach is summarised in figure 5.1.

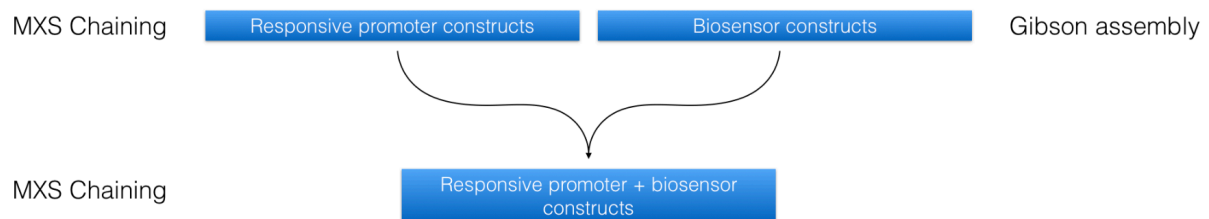


Figure 5.1. Flow chart showing the overall cloning strategy employed in this section. It shows the progression of cloning rounds, with inducible system components initially expressed from individual plasmids, to their combination into single plasmids.

Table 5.1. Variables employed in the *in vivo* testing of an optimised lactate-inducible transgene expression system.

Variables	Examples used
Operator number	1x; 2x; 1x and 1x (flanking a promoter); 2x and 2x (flanking a promoter)
Operator location, relative to a promoter	Upstream; downstream; flanking
LldR host origin	<i>Corynebacterium glutamicum</i> ; <i>Pseudomonas aeruginosa</i>
Configuration type	De-repression by steric hindrance; transactivation using VP64
Minimal promoter	minCMV; YB_TATA
Number of constructs co-transfected	LldR and responsive element on separate plasmids; both components on a single plasmid

5.4 Transient transfections of plasmids containing LldR variants and response elements into CHO-S

After cells were transiently co-transfected with given constructs in the absence and presence of lactate, they were incubated for 48 hours before being assayed by flow cytometry. To restrict subsequent analyses to the events of interest, the following process was carried out (see figure 3.2): cells which do not resemble the forward and side scatter profiles of a negative control are excluded as non-viable cells (Homann et al., 2017); then, single cells are selected for by looking at the area and height of forward scatter, excluding the cells which have a greater area than would be expected for their height; then, the RFP+ cells (those expressing mCherry as part of a P2A fusion to a transcriptional regulator construct) are selected. From these cells, the geometric mean of the GFP signal is taken. These geometric means are averaged for the three biological replicates of each condition, for both uninduced and induced conditions. A Student's T-test (one-tailed, two-sampled, equal variance) is performed to determine if there is any significant difference between the results. To determine the variability of the outcomes, each of these experiments was repeated three times in total. For every batch of experiments, 4 control experiments were also performed: a negative control where no DNA

or transfection reagents were used, so that any unexpected adverse cultivation could be detected; a negative control to which transfection reagents were added without any DNA, to gauge the impact of the reagents on cell morphology and to accurately gate for RFP+ populations; and two positive controls where either constitutively expressing GFP or RFP constructs were transfected into cells, to check the proper functioning of the flow cytometer. Six batches of experiments were performed, with these controls performed for each batch. For organisational purposes, each batch of experiments was labelled according to the order in which it was performed; subsequent mentions of, for example, NICE1, refer to “New *in vivo* construct experiment 1”. There was a significant difference in the uninduced and induced GFP signals of the positive controls used in two batches of experiments (see figure 9.26) and these were not analysed further. Certain configurations have been tested in 3 independent experiments (each composed of 3 biological replicates) while others were tested in 1 independent experiment (again, composed of 3 biological replicates).

The experiments performed and the results obtained are given in appendix 9.10. Owing to limitations of time, not all of the constructs given in table 9.6 were made/tested.

Of the 68 experiments that were successfully performed, 41 experiments returned a fold-change upon the addition of lactate with a p-value below 0.05, the threshold at which a significant difference between two populations is inferred. A total of 25 of the 36 constructs tested returned a significantly different fold-change in at least one of the independent experiments performed, all but one of which were transactivation configurations. A summary of the results is given in table 9.8, and the fold-changes of each experiment are shown in figure 5.2. In the following paragraphs, induction is calculated by dividing the GFP signal in the absence of lactate by the signal in the presence of lactate, following Gossen & Bujard, 1992. Additionally, co-transfections of LldR and responsive elements on separate plasmids are indicated by the use of a “+” symbol between these two components; the use of a “–” symbol in its place indicated that these components were encoded on a single plasmid.

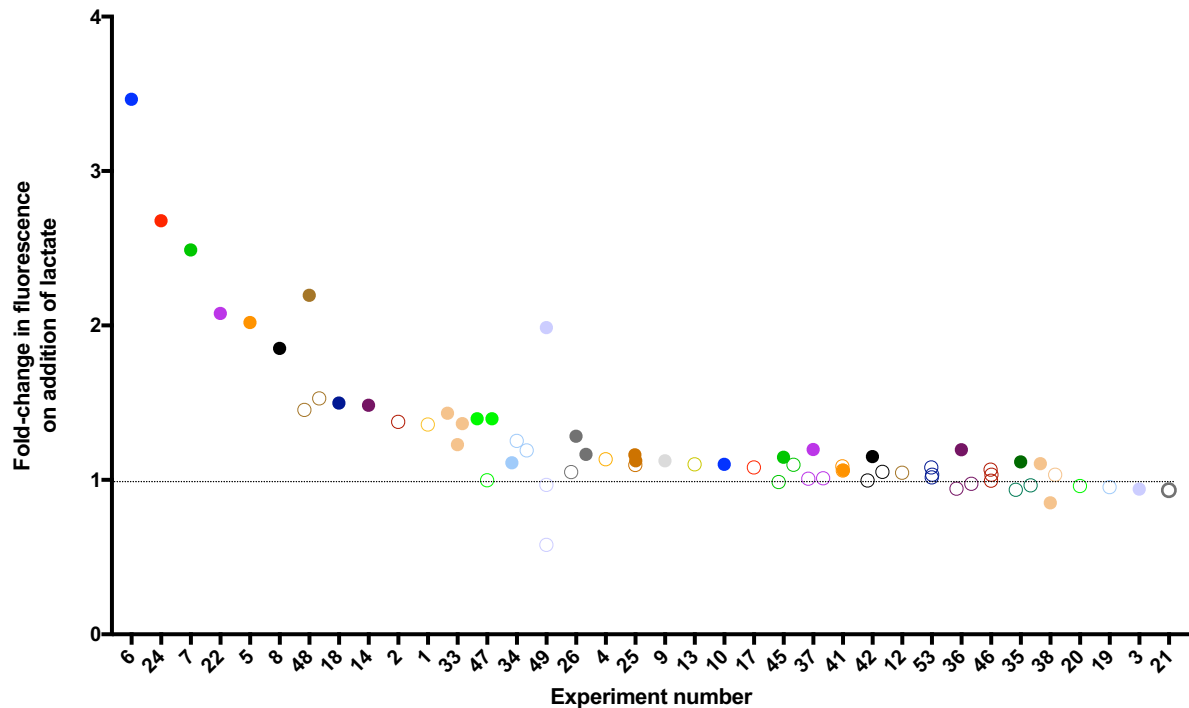


Figure 5.2. Conditions sorted by the average fold-change in geometric mean of the GFP signal. The horizontal dotted line shows a fold-change of 1, where the uninduced geometric mean of the GFP signal is equal to the induced signal. Filled circles indicate a p-value lower than 0.05, and hollow circles indicate a p value greater than 0.05. The identity of the biosensor construct and responsive element used in each condition is given in table 9.8, along with a detailed summary of the experimental results.

Table 5.2. Detailed summary of results from the transient transfection experiments conducted with optimized in vivo constructs. The p-value is calculated through the use of a Student's T-test (one-tailed, two-sampled, equal variance) to determine whether the uninduced and induced populations were significantly different. *C. g.* refers to the *Corynebacterium glutamicum*, and *P. a.* refers to *Pseudomonas aeruginosa*. The use of a “+” or a “-” in the “Construct name(s)” column refers to a co-transfection of two plasmids or the transfection of a single, larger plasmid, respectively. Full details are given in table 9.8, including the GFP signal for each biological replicate and standard deviations.

Experiment #	Construct name(s)	GFP signal (AU)		Fold induction (-lactate/+lactate)	p-value
		-lactate	+lactate		
1	L (<i>C. g.</i>) + CMV-1x	301377	221802	1.36	0.07

2	L (<i>C. g.</i>) + CMV-2x	284785	206855	1.38	0.08
3	L (<i>P. a.</i>) + CMV-1x	9009	9583	0.94	0.03
4	L (<i>P. a.</i>) + CMV-2x	79259	69890	1.13	0.34
5	L15V (<i>C. g.</i>) + minCMV-1x	29517	14617	2.02	0.00
6	L15V (<i>C. g.</i>) + minCMV - 2x	38326	11062	3.46	0.00
7	L15V (<i>C. g.</i>) + 1x-minCMV	22832	9174	2.49	0.01
8	L15V (<i>C. g.</i>) + 2x-minCMV	16027	8658	1.85	0.04
9	L15V (<i>C. g.</i>) + 1x- minCMV-1x	17295	15393	1.12	0.00
10	L15V (<i>C. g.</i>) + 2x- minCMV-2x	23953	21764	1.10	0.03
12	L15V (<i>C. g.</i>) +YB-2x	15757	15057	1.05	0.28
13	L15V (<i>C. g.</i>) +1x-YB	12422	11285	1.10	0.12
14	L15V (<i>C. g.</i>) +2x-YB	12970	8742	1.48	0.00
17	LV (<i>P. a.</i>) + minCMV-1x	19163	17733	1.08	0.34
18	LV (<i>P. a.</i>) + minCMV-2x	35629	23779	1.50	0.00
19	LV (<i>P. a.</i>) + 1x-minCMV	25390	26625	0.95	0.34
20	LV (<i>P. a.</i>) + 2x-minCMV	24306	25297	0.96	0.33
21	LV (<i>P. a.</i>) + 1x-minCMV- 1x	23768	25457	0.93	0.40
22	LV (<i>P. a.</i>) + 2x-minCMV- 2x	44188	21257	2.08	0.04
24	LV (<i>P. a.</i>) + YB-2x	23123	8635	2.68	0.01
25	LV (<i>P. a.</i>) + 1x-YB	16394	14585	1.12	0.00
		10792	9286	1.16	0.01
		18589	16953	1.10	0.07
26	LV (<i>P. a.</i>) + 2x-YB	19481	16707	1.17	0.01
		13459	10485	1.28	0.03
		18864	17955	1.05	0.18

33	L15V (C. g.) - minCMV-1x	14786	10833	1.36	0.00
		12716	10353	1.23	0.01
		15167	10588	1.43	0.02
34	L15V (C. g.) - minCMV-2x	12370	11131	1.11	0.03
		13884	11087	1.25	0.06
		14525	12187	1.19	0.26
35	L15V (C. g.) - 1x-minCMV	9489	8501	1.12	0.05
		9709	10067	0.96	0.10
		15738	16804	0.94	0.12
36	L15V (C. g.) - 2x-minCMV	10257	8582	1.20	0.01
		15391	16305	0.94	0.16
		12931	13265	0.97	0.21
37	L15V (C. g.) - 1x-minCMV-1x	10942	9138	1.20	0.01
		12534	12436	1.01	0.39
		14785	14635	1.01	0.41
38	L15V (C. g.) - 2x-minCMV-2x	19455	17609	1.10	0.01
		13193	15480	0.85	0.02
		10086	9764	1.03	0.24
41	L15V (C. g.) - 1x-YB	7604	7170	1.06	0.04
		7524	6913	1.09	0.06
		6056	5703	1.06	0.08
42	L15V (C. g.) - 2x-YB	6229	5413	1.15	0.00
		8406	7993	1.05	0.25
		7380	7408	1.00	0.48
45	LV (P. a.) - minCMV-1x	8468	7381	1.15	0.04
		15868	14450	1.10	0.07
		11875	12043	0.99	0.45

46	LV (<i>P. a.</i>) - minCMV-2x	9114	8548	1.07	0.08
		14140	13673	1.03	0.16
		12008	12074	0.99	0.47
47	LV (<i>P. a.</i>) - 1x-minCMV	13210	9463	1.40	0.03
		28289	20260	1.40	0.04
		22700	22735	1.00	0.49
48	LV (<i>P. a.</i>) - 2x-minCMV	50455	22984	2.20	0.01
		33884	22182	1.53	0.07
		21902	15074	1.45	0.11
49	LV (<i>P. a.</i>) - 1x-minCMV-1x	25439	12807	1.99	0.02
		9217	15924	0.58	0.12
		34905	36061	0.97	0.48
53	LV (<i>P. a.</i>) - 1x-YB	7229	6684	1.08	0.09
		7813	7683	1.02	0.16
		5645	5458	1.03	0.21

The best performing conditions are shown in figure 5.3. These were all transactivation configurations, with fold inductions ranging from 3.46 for L15V (*Corynebacterium glutamicum*) + minCMV – 2x, to 2.02 for L15V (*Corynebacterium glutamicum*) + minCMV – 1x. All showed statistically significant levels of induction, albeit only in one independent experiment. Figure 5.4 shows those configurations that demonstrated statistically significant levels of induction in more than one independent experiment. The best performer was the L15V (*Corynebacterium glutamicum*) – minCMV – 1x configuration, with three statistically significant results, and an average fold induction of 1.34.

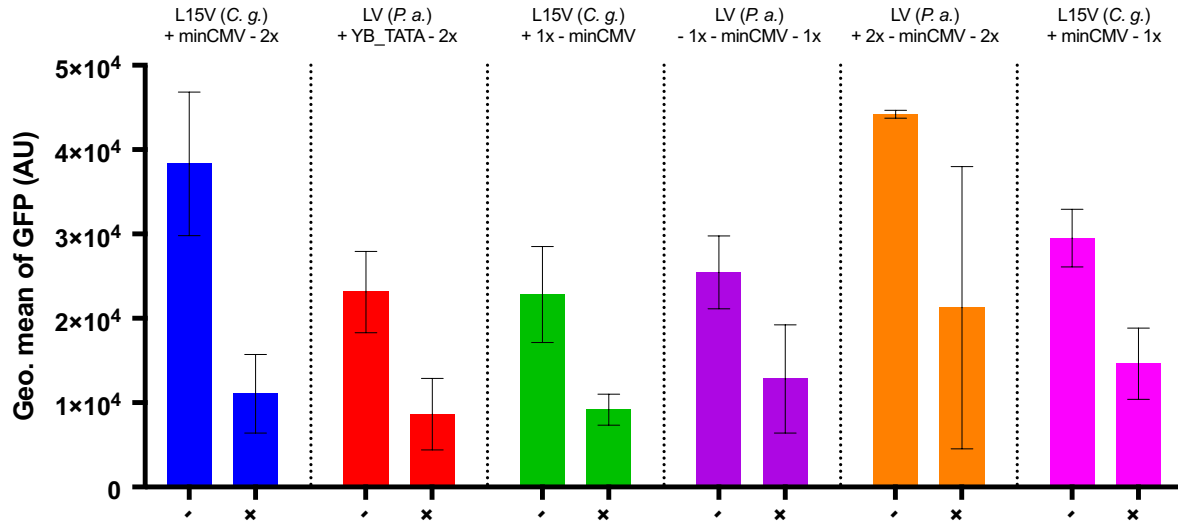


Figure 5.3. The configurations demonstrating the highest levels of induction in response to lactate. L15V (*Corynebacterium glutamicum*) + minCMV – 2x produced the highest induction observed, of 3.46. The plus and minus symbols along the x-axis indicate the presence or absence, respectively, of 20 mM lactate for the condition directly above. Independent experiments are shown, each consisting of three biological replicates. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

The number of experiments carried out here allowed the comparison of numerous variables to each other, including whether it is preferable to have 1 or 2 operators at a given location; if the location of those operators has a significant impact; if the LldR from *Corynebacterium glutamicum* is preferable to that from *Pseudomonas aeruginosa*; if the minCMV or YB_TATA minimal promoter is preferable for transactivation; and whether the use of large plasmids containing both a biosensor and a responsive promoter were preferable to co-transfections with plasmids containing either component. In outline, it was preferable to use LldR from *Corynebacterium glutamicum* with multiple operators adjacent to minCMV; no clear preference for the location of these operators relative to a promoter could be established. Additionally, the comparison between co-transfections and transfections of larger plasmids encoding the entire lactate-inducible system showed that the co-transfections resulted in greater levels of fold induction.

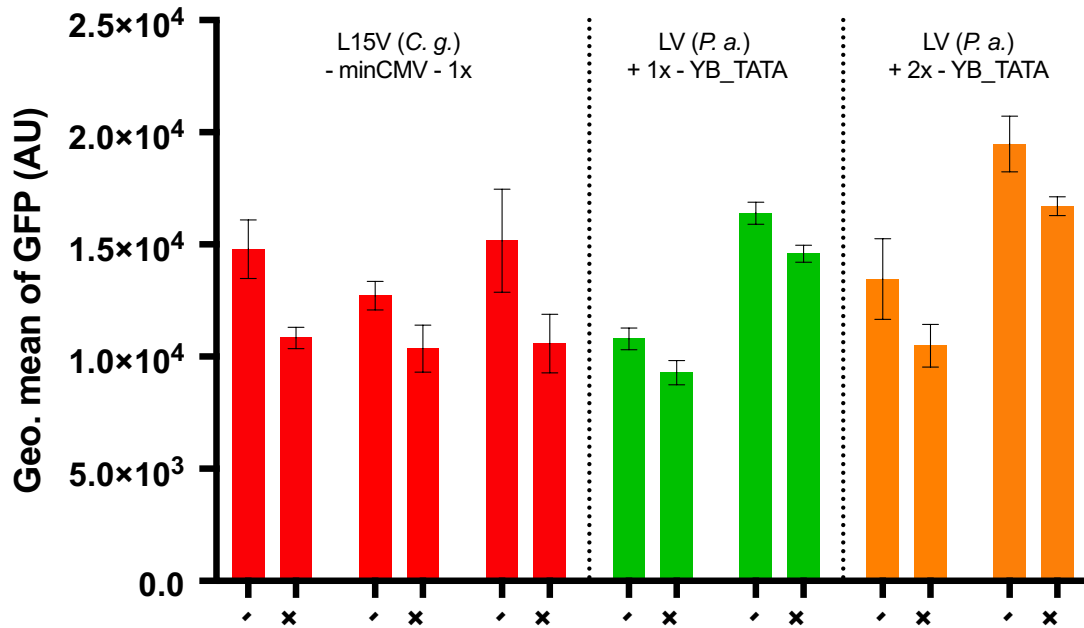


Figure 5.4. Configurations demonstrating significant levels of induction in multiple independent experiments. L15V (*Corynebacterium glutamicum*) – minCMV – 1x gave statistically significant levels of induction in 3 independent experiments, averaging a fold induction of 1.34. The plus and minus symbols along the x-axis indicate the presence or absence, respectively, of 20 mM lactate for the condition directly above. Independent experiments are shown, each consisting of three biological replicates. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

5.4.1 Comparison regarding the impact of operator number

It was possible to make 16 comparisons regarding the effect of operator number, between 32 different configurations which differed only on this basis. In 3 out of these 16 comparisons, neither configuration returned a statistically significant result in any of up to 3 independent experiments, and so were not compared. Of the remaining 13 comparisons where at least one significant experiment was observed between two configurations, it was instructive to first look at the operators placed upstream of a promoter. In this location, it was clearly preferable to have 2 operator repeats, as it gave a higher fold induction than only 1 repeat in 5 out of 6 comparisons (see figure 5.5). 4 comparisons were possible for consideration of downstream operators, where the preference is less clear (see figure 5.6). For LldR (*Pseudomonas aeruginosa*), a significant result was only seen for its de-repression configuration with a single operator repeat downstream of a CMV promoter (no other significant de-repression configurations were observed in these series of experiments), and a single downstream

operator was also preferable for L15V (*Corynebacterium glutamicum*) - minCMV, with an average fold induction of 1.34 in three statistically significant independent experiments, compared to a fold induction of 1.11 for only one statistically significant result with 2 downstream operators. Three comparisons were possible for configurations where the number of operators flanking both sides of a promoter was varied (see figure 5.7). For L15V (*Corynebacterium glutamicum*) and either 1x-minCMV-1x or 2x-minCMV-2x, both fold induction levels were significant and very close at 1.12 and 1.10 respectively; for LV (*Pseudomonas aeruginosa*) + either 1x-minCMV-1x or 2x-minCMV-2x, only the latter gave a significant fold induction (2.08); and for the single constructs L15V (*Corynebacterium glutamicum*) - 1x - minCMV - 1x and L15V (*Corynebacterium glutamicum*) - 2x - minCMV - 2x, both returned one significant fold induction indicating transactivation (1.20 and 1.10 respectively), while for the latter, a significant apparent de-repression result was also seen, with a fold induction of 0.85. This latter result indicates the difficulty of predicting the function of such a configuration, as the presence of operators both up and downstream of a promoter can conceivably have opposing effects on transcriptional activity.

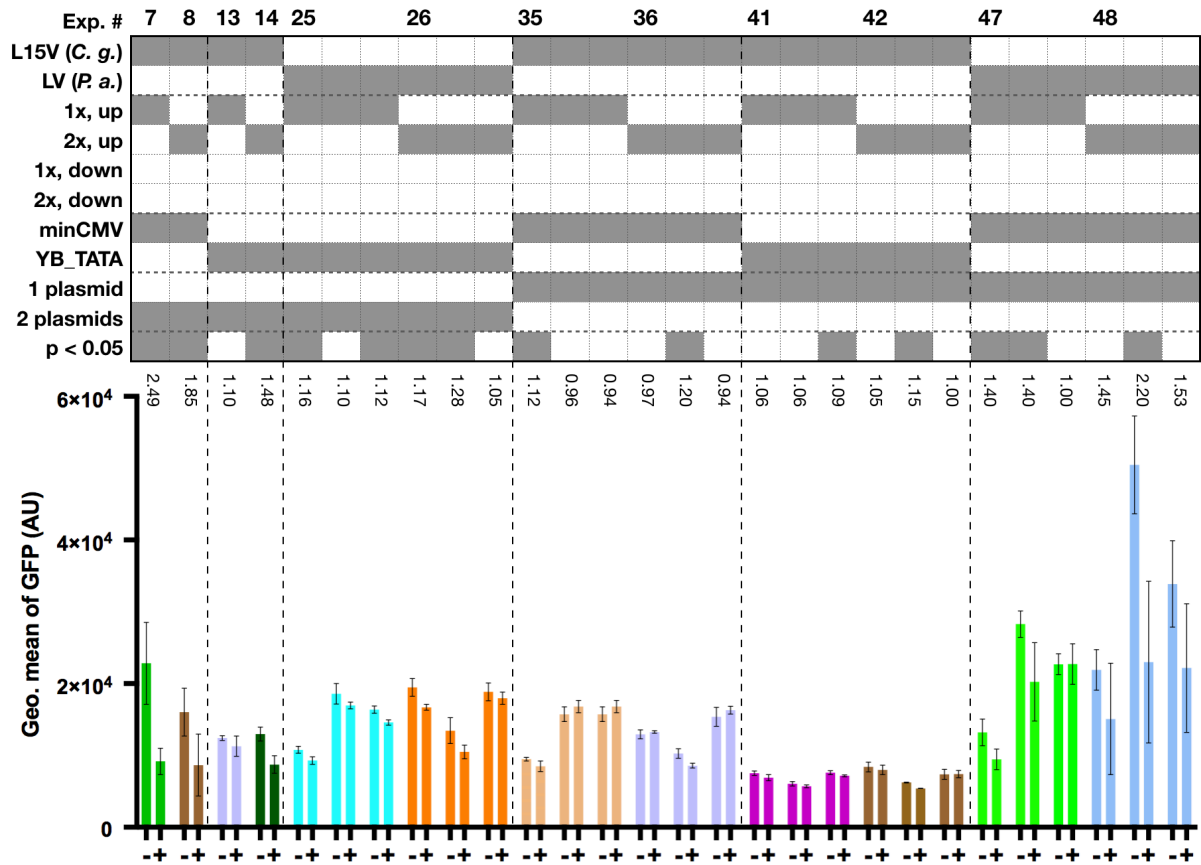


Figure 5.5. Comparison of configurations with either 1 or 2 upstream operators. The top half of the figure shows a legend that describes the configuration being tested, and the bottom half shows the GFP signal of transfected populations of CHO cells. The leftmost column of each pair represents the GFP signal from cells cultivated in the absence of lactate; the rightmost column of each pair is the GFP signal in the presence of 20 mM lactate. The number directly above each pair shows its fold-induction. Where more than one independent experiment was performed for a given configuration, the representative columns have been given the same colour. The leftmost comparison is between L15V (*Corynebacterium glutamicum*) + 1x-minCMV and L15V (*Corynebacterium glutamicum*) + 2x-minCMV, where the use of 1 operator repeat lead to greater fold induction in response to lactate (2.49 vs 1.85 for 2 operators repeats). This was the only exception to the general trend showing that 2 operator repeats led to greater fold induction. Comparisons are demarcated from one another by the vertical dotted lines. Each independent experiment consisted of three biological repeats. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

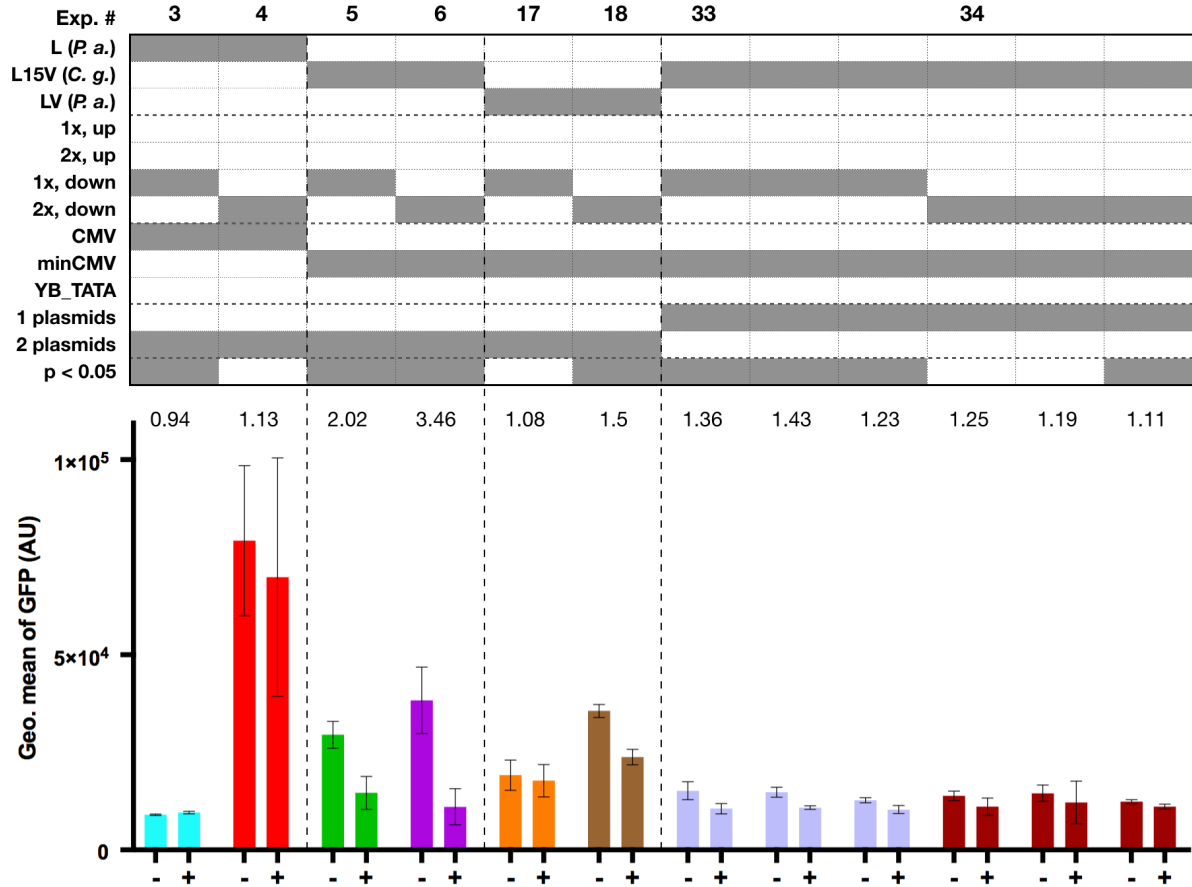


Figure 5.6. Comparison of configurations with either 1 or 2 downstream operators. No clear preference for the use of 1 or 2 downstream operators could be established from these experiments. The top half of the figure shows a legend that describes the configuration being tested, and the bottom half shows the GFP signal of transfected populations of CHO cells. The leftmost column of each pair represents the GFP signal from cells cultivated in the absence of lactate; the rightmost column of each pair is the GFP signal in the presence of 20 mM lactate. The number directly above each pair shows its fold-induction. Comparisons are demarcated from one another by the vertical dotted lines. Each independent experiment consisted of three biological repeats. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

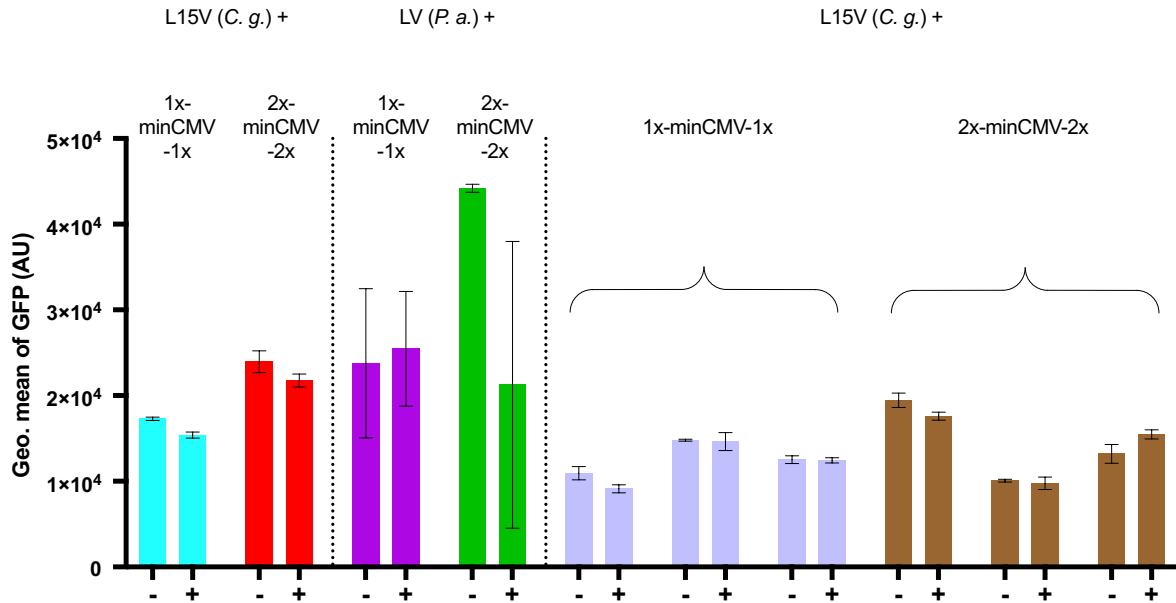


Figure 5.7. Comparison of configurations with either 1 or 2 operator repeats flanking both sides of a promoter. No clear preference for the use of 1 or 2 operators flanking the promoter could be established from these experiments. The leftmost column of each pair represents the GFP signal from cells cultivated in the absence of lactate; the rightmost column of each pair is the GFP signal in the presence of 20 mM lactate. The number directly above each pair shows its fold-induction. Where more than one independent experiment was performed for a given configuration, the representative columns have been given the same colour. Comparisons are demarcated from one another by the vertical dotted lines. Each independent experiment consisted of three biological repeats. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

5.4.2 Comparison regarding the use of LldR from either *Corynebacterium glutamicum* or *Pseudomonas aeruginosa*

16 comparisons can be made concerning the use of LldR from either *Corynebacterium glutamicum* or *Pseudomonas aeruginosa* (see figures 5.8 and 5.9). A slight preference for the use of the *Corynebacterium glutamicum* LldR can be observed: in 9 out of 16 of these comparisons, the LldR from *Corynebacterium glutamicum* gives a greater fold induction relative to that from *Pseudomonas aeruginosa*; for the remaining 7 comparisons, the converse is true. To give an example, for L15V (*Corynebacterium glutamicum*) co-transfected with minCMV - 2x, a fold induction of 3.46 was observed, whereas a fold induction of 1.50 was seen with the use of LV (*Pseudomonas aeruginosa*) on the same responsive element.

Conversely, the transfection of a single plasmid encoding LV (*Pseudomonas aeruginosa*) and 1x - minCMV gave a fold induction of 1.4 in 2 out of 3 independent experiments, whereas a comparison condition using L15V (*Corynebacterium glutamicum*) gave a significant fold induction of 1.12 in only 1 out of 3 independent experiments. There seemed to be a bias whereby configurations employing a minCMV promoter were more likely to see significant and greater induction using the LldR protein from *Corynebacterium glutamicum*, true for 7/11 comparisons. No other biases were observed.

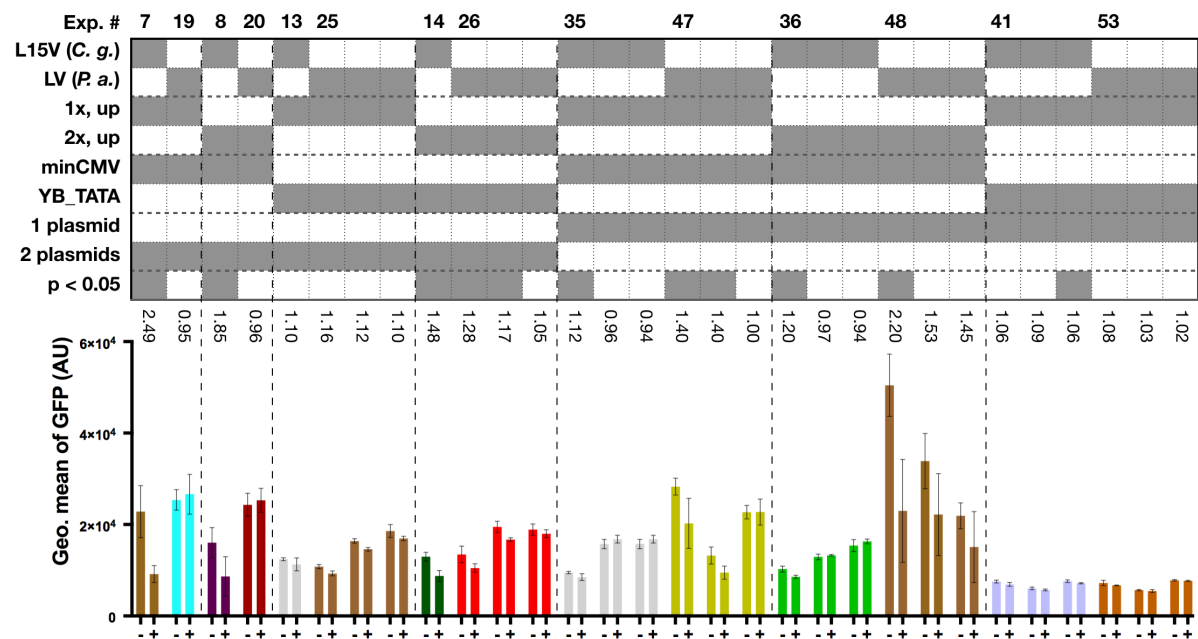


Figure 5.8. Comparison of configurations with upstream operators using LldR from either *Corynebacterium glutamicum* or *Pseudomonas aeruginosa*. A slight preference for the use of the *C. glut.* LldR is seen in these experiments. The LV (*Pseudomonas aeruginosa*) – 1x – minCMV construct gave a fold induction of 1.4 in two out of three independent experiments, whereas the use of L15V (*Corynebacterium glutamicum*) in its place gave a fold induction of 1.12 in only one out of three experiments. The top half of the figure shows a legend that describes the configuration being tested, and the bottom half shows the GFP signal of transfected populations of CHO cells. The leftmost column of each pair represents the GFP signal from cells cultivated in the absence of lactate; the rightmost column of each pair is the GFP signal in the presence of 20 mM lactate. The number directly above each pair shows its fold-induction. Comparisons are demarcated from one another by the vertical dotted lines. Each independent experiment consisted of three biological repeats. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

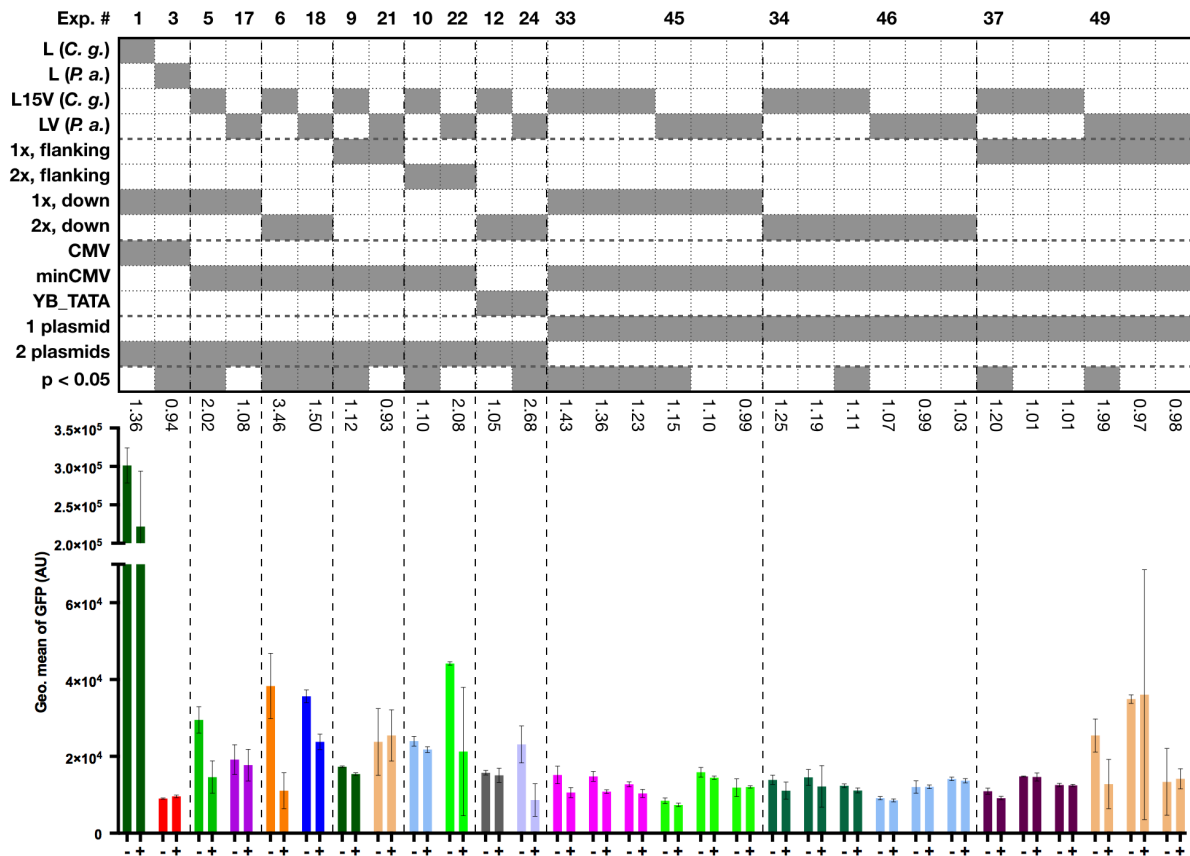


Figure 5.9. Comparison of configurations with downstream or flanking operators using LldR from either *Corynebacterium glutamicum* or *Pseudomonas aeruginosa*. A slight preference for the use of the *C. glut.* LldR is seen in these experiments. L15V (*Corynebacterium glutamicum*) co-transfected with minCMV-2x gave a fold induction of 3.46, whereas the use of LV (*Pseudomonas aeruginosa*) in the same configuration gave only a 1.50 fold induction. The top half of the figure shows a legend that describes the configuration being tested, and the bottom half shows the GFP signal of transfected populations of CHO cells. The leftmost column of each pair represents the GFP signal from cells cultivated in the absence of lactate; the rightmost column of each pair is the GFP signal in the presence of 20 mM lactate. The number directly above each pair shows its fold-induction. Comparisons are demarcated from one another by the vertical dotted lines. Each independent experiment consisted of three biological repeats. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

5.4.3 Comparison regarding the placement of operators upstream or downstream of a promoter

9 comparisons can be made concerning the placement of operators either upstream or downstream of a minimal promoter used in a transactivation configuration. In 5 out of 9 such comparisons, the upstream operator gives rise to greater fold induction, and in 4 out of 9 comparisons the downstream operator is preferable. The best level of induction for a configuration with an upstream operator was for L15V (*Corynebacterium glutamicum*) co-transfected with 1x-minCMV, with a fold induction of 2.49, relative to a fold induction of 2.02 when a downstream operator was used. However, the greatest level of induction seen was for a downstream operator, in the L15V (*Corynebacterium glutamicum*) co-transfection with minCMV-2x, which gave an induction of 3.46 compared to 1.85 when the operator was instead placed upstream. There were no configuration variables (e.g. the origin of the LldR or the number of operators used) that were particularly associated with the configurations where the upstream promoter led to greater levels of induction relative to a downstream one (i.e., it is not the case that wherever an upstream operator was preferable, the *Corynebacterium glutamicum* LldR was also part of the configurations being compared).

5.4.4 Comparison regarding the use of the minCMV or YB_TATA minimal promoters

9 comparisons can be made regarding the use of either a minCMV or YB_TATA minimal promoter in a transactivation configuration (see figure 5.10). minCMV gave rise to higher levels of induction in 6 out of these 9 comparisons, with YB_TATA the better performer in the other 3. The maximum induction seen with minCMV was L15V (*Corynebacterium glutamicum*) co-transfected with minCMV - 2x, with a fold induction of 3.46, whereas the use of YB_TATA in its place did not show a statistically significant level of induction. The maximum induction seen with YB_TATA was LV (*Pseudomonas aeruginosa*) co-transfected with YB_TATA - 2x, with a fold induction of 2.68, compared to 1.50 when minCMV was used in its place. In general, YB_TATA resulted in lower levels of expression compared to minCMV, in both uninduced and induced conditions. Thus, it may be preferable to use minCMV in any future transactivation configuration.

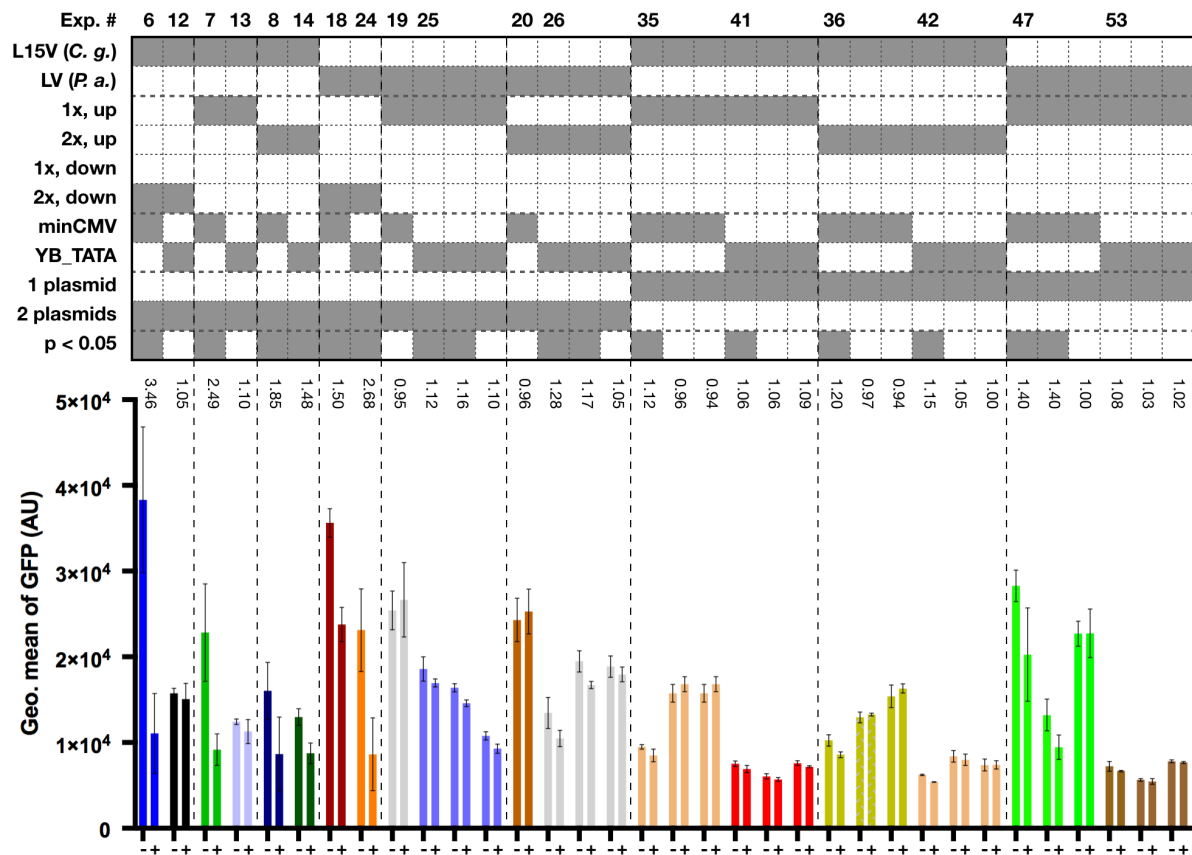


Figure 5.10. Comparison of configurations using either minCMV or YB_TATA. The top half of the figure shows a legend that describes the configuration being tested, and the bottom half shows the GFP signal of transfected populations of CHO cells. In general, the use of YB_TATA led to lower levels of GFP expression than minCMV. The leftmost column of each pair represents the GFP signal from cells cultivated in the absence of lactate; the rightmost column of each pair is the GFP signal in the presence of 20 mM lactate. The number directly above each pair shows its fold-induction. Comparisons are demarcated from one another by the vertical dotted lines. Each independent experiment consisted of three biological repeats. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

5.4.5 Comparison regarding co-transfections of two plasmids against the transfection of a single, larger plasmid

Co-transfections of two plasmids expressing either the lactate sensing protein (LldR) or containing a responsive element can be compared to the transfection of a single plasmid containing both of these elements in 13 comparison experiments (see figures 5.11 and 5.12). A higher fold induction was seen in 8 out of these 13 comparisons when co-transfecting two

plasmids. It may be the case that when a single plasmid containing both a lactate sensing protein and a responsive element was transfected, the proportion of these two elements relative to each other was suboptimal compared to when they were co-transfected as separate plasmids. This may imply that the design of an inducible system for genomic integration could benefit from modulating the promoter strength of one of these two elements, to optimise the amount of LldR expressed relative to the responsive element.

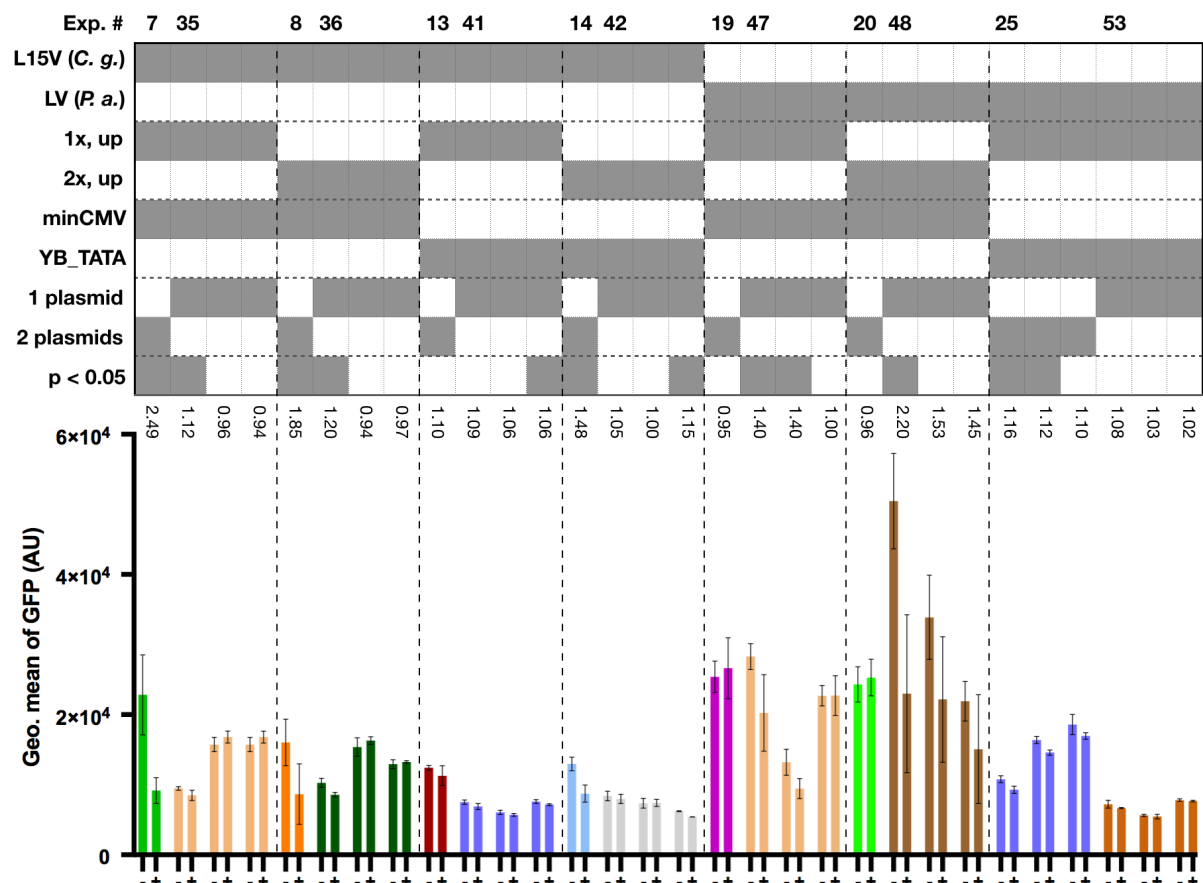


Figure 5.11. Comparison of configurations with upstream operators transfected either with separate constructs or on a single construct. In general, a greater fold induction was observed for co-transfected plasmids. The top half of the figure shows a legend that describes the configuration being tested, and the bottom half shows the GFP signal of transfected populations of CHO cells. The leftmost column of each pair represents the GFP signal from cells cultivated in the absence of lactate; the rightmost column of each pair is the GFP signal in the presence of 20 mM lactate. The number directly above each pair shows its fold-induction. Comparisons are demarcated from one another by the vertical dotted lines. Each independent experiment consisted of three biological repeats. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

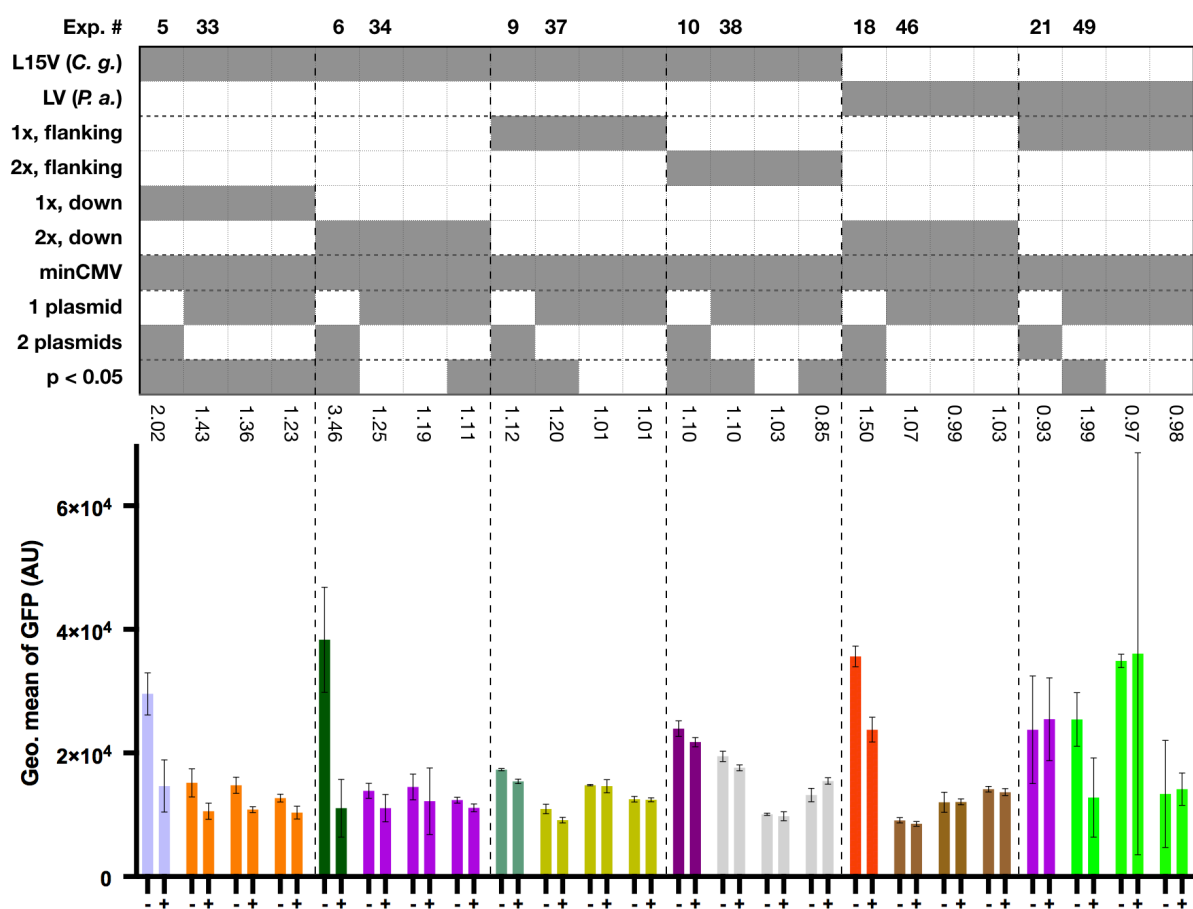


Figure 5.12. Comparison of configurations with downstream or flanking operators transfected either with separate constructs or on a single construct. In general, a greater fold induction was observed for co-transfected plasmids. This is particularly clear in the cases of L15V (*Corynebacterium glutamicum*) co-transfected with minCMV – 1x, minCMV – 2x, and LV (*Pseudomonas aeruginosa*) co-transfected with minCMV – 2x. The top half of the figure shows a legend that describes the configuration being tested, and the bottom half shows the GFP signal of transfected populations of CHO cells. The leftmost column of each pair represents the GFP signal from cells cultivated in the absence of lactate; the rightmost column of each pair is the GFP signal in the presence of 20 mM lactate. The number directly above each pair shows its fold-induction. Comparisons are demarcated from one another by the vertical dotted lines. Each independent experiment consisted of three biological repeats. Error bars represent standard deviation from the mean. AU is an abbreviation for arbitrary units.

5.5 Summary and discussion

Transient transfection experiments were able to show multiple constructs capable of significantly inducing a GFP signal upon the addition of lactate. 25 out of the 36 configurations tested returned a significant fold induction in response to lactate in at least one independent experiment; 5 of the 36 were also able to show a fold induction above 2 in at least one independent experiment (see figure 5.3). The best performing construct was L15V (*Corynebacterium glutamicum*) co-transfected with minCMV-2x, giving a fold induction of 3.46. The most consistently functional configuration was L15V (*Corynebacterium glutamicum*) – minCMV – 1x, which gave an average fold induction of 1.34 across three statistically significant independent experiments (see figure 5.4).

Induction was often inconsistently observed; additionally, some positive control experiments showed apparent induction in their GFP signals in response to lactate. To some extent, this will be an unavoidable impact caused by the addition of lactate to the cells. As described in section 1.2.3.3, lactate is detrimental to the health of CHO cells and may be limiting their expression of GFP, in combination with the stress caused by the transfection procedure. Beyond repeating the experiments performed here to confirm the functionality of various configurations, additional rounds of *in vitro* testing could also be performed in order to circumvent any confounding impact the addition of lactate has *in vivo*. In future *in vivo* experiments, the use of an intracellular pH sensor (Tantama et al., 2011) may be able to provide cell-by-cell information about intracellular lactate-induced pH change (if any) and its possible relationship to the induction observed.

A general preference for the use of multiple operators was seen relative to the use of only one. That this trend was especially evident for operators upstream of minimal promoters indicates that the addition of extra operator sites is likely to be a fruitful approach to the further optimisation of future lactate-inducible systems, as they will not interfere with the progression of RNA polymerase from the promoter to the transcription start site. Given that the use of multiple operators has been shown to improve inducibility of transgene expression systems (e.g. Gossen & Bujard, 1992), this result was not unexpected.

The use of upstream operators typically gave greater levels of induction than the use of downstream operators, although this was not particularly pronounced. Future work employing higher operator numbers would help to illuminate this. It may be the case that the use of few operator repeats meant that not only was the extent of transactivation caused by the action of VP64 limited, but any possible steric hindrance of a downstream operator was also limited.

The use of additional operators in the upstream location is likely to give cumulative improvements to inducibility, whereas further operators in the downstream location are likely to be comparatively detrimental.

minCMV seemed to be the preferred choice of minimal promoter despite its well-known leakiness and the touted superiority of YB_TATA in this regard (Ede et al., 2016). It may be the case that the low expression of YB_TATA means that it is hard to discriminate between its uninduced and induced outputs using GFP expression as a signal, given the high levels of autofluorescence in CHO cells (Aubin et al., 1979). Using luciferase as an output instead might be a better way of detecting expression (Troy et al., 2004), although as this typically necessitates the use of a plate reader, it would no longer be possible to distinguish between transfected and untransfected cells. If the overall inducibility could be improved (for example, through the use of additional operator repeats), it may be possible to discern finer differences between the merits of minCMV and YB_TATA, with a clear preference emerging for one over the other.

With regard to the host origin of the LldR used, there was a slight preference for that of *Corynebacterium glutamicum*. Although there is no published data which directly compares the binding activity of these two proteins, table 4.3 does show that LldR from *Pseudomonas aeruginosa* binds to cgl2917 more strongly than the LldR from *Corynebacterium glutamicum*. With this in mind, and the limited extent to which the LldR from *Corynebacterium glutamicum* was more successful in the experiments performed in this chapter, it may be prudent in any further attempts at optimising this inducible system to test both proteins.

The disparity between the observed levels of induction when co-transfecting two plasmids separately encoding the LldR fusion variant and the responsive element, and transfecting the same configuration on a single plasmid, may indicate that the proportion of these two elements to each other need to be optimised in their own right. This could be achieved by modulating the strength of the promoter expressing the LldR, using a large library of synthetic promoters (Brown et al., 2014).

6 Use of active Cas9-VPR to upregulate endogenous genes

6.1 Chapter aims

- Develop an efficient cloning strategy for constructing guide RNA-expressing plasmids
- Determine whether Cas9-VPR can be used to alter gene expression in CHO-S

6.2 Chapter summary

Cas9-VPR is a DNA-targeting protein fusion capable of, in theory, upregulation, downregulation, and excision of targeted genes. This chapter provides a preliminary demonstration of its use in Chinese hamster ovary cells, with a transient transfection of Cas9-VPR able to upregulate expression of GFP from a co-transfected plasmid powered by a minimal promoter.

6.3 Results

6.3.1 Constructing guide RNA-expressing plasmids for use with Cas9-VPR

In an attempt to rapidly demonstrate the utility of the Cas9-VPR system to upregulate gene expression in CHO cells, it was used to target a transiently expressed minCMV promoter expressing mAzamiGreen. In the absence of Cas9-VPR targeted to this minCMV promoter, only basal (but non-zero) levels of mAzamiGreen expression should be seen; when a guide RNA targeting the Cas9-VPR to the promoter is present, greater levels of mAzamiGreen expression should typically be seen. Given that the minCMV promoter is leaky and exhibits low levels of expression in the absence of a targeted transcriptional activator (see figure 6.2 and Ede et al., 2016), it is also possible to use this promoter to test the ability of the Cas9-VPR system to downregulate gene expression.

Table 6.1. Description of the guide RNAs designed against the minCMV-mAzamiGreen construct.

Guide	Sequence	Location (# bases upstream of the TSS)	Target type	On-target score	Geometric mean of GFP (arbitrary units *10 ⁴)	Significant increase over condition with no guide?
1	cgagtatgtcgagg	96	promoter	72.4	13.6	Yes
2	agctcactcaaagg	193	upstream of promoter	70.6	14.7	Yes
3	cacgctcaccatgg	2	across TSS	68.2	8.6	No
4	gcgtgtacggtggg	94	promoter	59.6	13.4	Yes
5	cacgctcaccatgg	35	promoter	56.2	13.6	Yes

The RNA-guided *Streptococcus pyogenes* Cas9 can theoretically target any stretch of DNA, as long as it contains the protospacer adjacent motif (PAM) NGG. Practically this does not impose major constraints on the possible targets of Cas9 - Cong et al. (2013) state that this means a PAM sequence is found on average every 8 bases in human cells, which is presumably similar for CHO cells as both species have a GC content of approximately 41% (Xu et al., 2011). In the 200 bases upstream of the transcription start site (TSS) of mAzamiGreen, there are 21 possible guide RNA targets. There are no clear rules with regard to the kinds of DNA sequence that are best to target with the aim of gene upregulation, beyond evidence that the closer the target is to the TSS the greater the likelihood and extent of upregulation, and high levels of upregulation can be seen within 200 bases of the TSS (Konermann et al., 2015). Downregulation with Cas9-VPR seems to require the presence of a TATA box within a promoter where a downstream guide RNA can be targeted, although this appears to be far less reliably achieved than upregulation with Cas9-VPR. Kiani et al. (2015), Gilbert et al. (2013) and Konermann et al. (2015) all show that the extent of upregulation can

vary greatly and somewhat unpredictably depending on the sequence of the guide RNA used, and so it is prudent to test multiple guide for each gene being targeted.

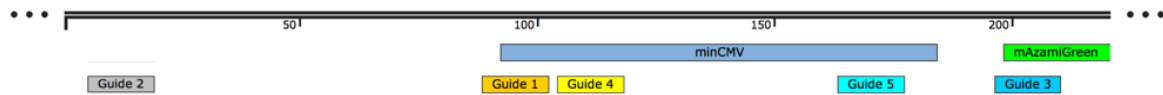


Figure 6.1. Location of guide RNA targets against the minCMV-mAzamiGreen plasmid. Only the relevant portion of the circular plasmid is shown, with the remainder of the plasmid designated by the flanking ellipses.

The 21 possible guide RNA targets upstream of the TSS of mAzamiGreen can be discriminated between on the basis of their on-target scores, predicted using an algorithm from Doench et al. (2016). Their work looked at the correlation of guide RNA features to the strength of an interaction with its target, to determine the probability that a given guide will be functional. Thus, the guides which were within 200 bases of the TSS and had the highest on-target scores were selected for testing, and are described in table 6.1.

Given the need to test multiple guides, it is essential that cloning these guides into a suitable vector is as straightforward as possible. Once designed in the Benchling web-application, 5' sequence overhangs can be added that are complementary to a BsaI-digested guide RNA entry vector. After ordering a pair of oligonucleotides encoding the guide of interest, they were phosphorylated with T4 polynucleotide kinase, annealed together and ligated into a BsaI-digested entry vector in a one-pot Golden Gate reaction (see figure 2.1, panels A to F). To streamline the selection of positively transformed *Escherichia coli* cells, superfolder GFP (sfGFP) containing the two BsaI sites was present in the guide RNA entry vector, such that a successful cloning reaction would excise the sfGFP and allow the ligation of the guide RNA construct of interest. Thus, it was possible to distinguish between non-transformed cells (which expressed a highly visible level of sfGFP, apparent to the naked eye and especially on transillumination equipment - see panel G of figure 2) and transformed cells, which did not express sfGFP and so appeared as normal colonies on selective media, resulting from the presence of a correctly integrated guide RNA construct.

6.3.2 Preliminary testing of the Cas9-VPR system in CHO-S

As preliminary work to establish whether the Cas9-VPR system could be used in Chinese hamster ovary cells, a transient co-transfection of plasmids encoding three components was carried out:

- Cas9-VPR, allowing the manipulation of the transcription rate of the targeted gene
- The green fluorescent protein mAzamiGreen controlled by the minimal promoter minCMV
- A guide RNA expressing construct targeting the Cas9-VPR to the minCMV promoter

This was tested by flow cytometry 48 hours after transfection.

The results of this experiment are seen in figure 6.2, showing that significant increases in the GFP signal were seen in 4 out of 5 instances, and no instances of downregulation. The greatest fold-change seen was 1.73 for guide 2, with an average fold-change of 1.63 seen for those capable of induction. Only guide 3, positioned partially over the TSS, did not lead to upregulation. This demonstrates that it is relatively straightforward to design, construct and test guide RNA constructs in CHO cells that lead to upregulation of a gene, but that finding a guide RNA able to downregulate gene expression will likely require the screening of many more guides and may not be possible in all cases.

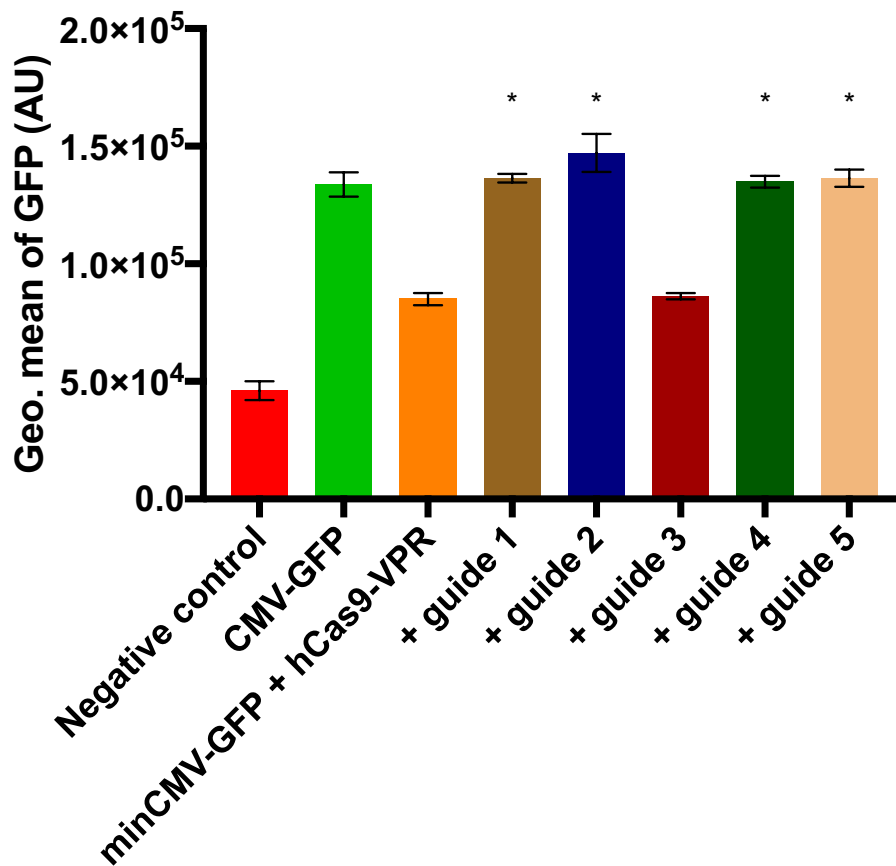


Figure 6.2. Testing the ability of Cas9-VPR to manipulate gene expression in a transient transfection experiment. The negative control is a CHO-S sample that is untransfected. The “minCMV-GFP + hCas9-VPR” is the condition against which the subsequent 5 “guide” columns have been compared, using a Student’s t-test to assess whether the differences in their values (from 3 biological replicates) is statistically significant, as indicated by the presence of an asterisk above the experimental column. Error bars represent standard deviation.

6.3.3 Targeting genes involved in lactate metabolism with Cas9-VPR

On the basis of the results showing that upregulation can be achieved using the Cas9-VPR system in CHO cells, the system was tested against genomic targets. There are many possible targets, such as those described previously in this chapter (e.g. lactate dehydrogenase C and the anti-apoptotic Bcl-2), or those found to be differentially expressed in a lactate consumption state versus a lactate production state (e.g. ATP synthase and cytochrome oxidase, from Mulukutla et al., 2012). It was decided to test the Cas9-VPR system on a previously untargeted gene, monocarboxylate transporter 1 (MCT1), which is characterised as a lactate importing

protein (Pisarsky et al., 2016). It was hypothesised that upregulating the expression of this gene could lead to earlier intracellular accumulation of lactate, reducing the flux to lactate via glycolysis and supporting oxidative phosphorylation.

Co-transfections of Cas9-VPR and various guide constructs targeting MCT1 were performed in CHO cells and assayed after 48 hours. RNA was extracted using the rapid and cost-effective RNASwift method (Nwokeoji et al., 2016), maintaining the integrity of the RNA (data not shown). To determine any difference in transcript levels of MCT1 in the presence of the guide constructs, RT-qPCR was used. The use of the Pfaffl method of relative quantification (Pfaffl, 2001) accounted for variations in primer amplification efficiencies, which are given in appendix 9.12. Actin-beta was used as a reference gene.

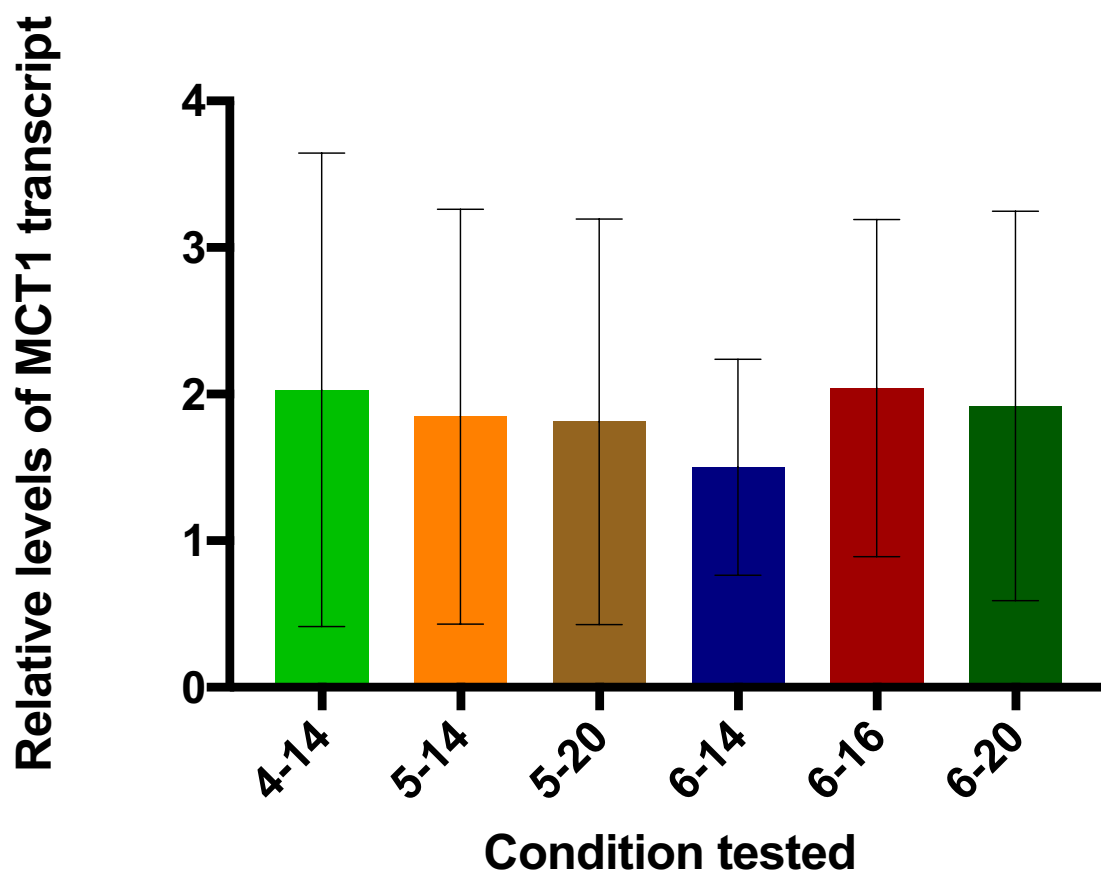


Figure 6.3. Testing the ability of Cas9-VPR to manipulate transcript levels of MCT1. None of the guides tested (shown on the x-axis) were able to reliably change the transcript levels of MCT1 above untreated levels. The y-axis shows the levels of MCT1 transcript relative to a negative control. Error bars represent standard deviation.

The results did not show consistent upregulation of MCT1 in the presence of the tested guide RNAs. Although in an initial experiment all the tested guides showed upregulation relative to a negative control (ranging from 2.4 to 3.9 fold increases in MCT1 transcript levels), two subsequent experiments could not reproduce these results.

6.4 Summary of results; discussion

The work in this chapter is a demonstration that Cas9-VPR can be used to upregulate gene expression in CHO, which was unreported at the commencement of this thesis. Only upregulation was demonstrated here, indicating that the use of Cas9-VPR to downregulate gene expression might be less reliable than indicated in the original paper (Kiani et al., 2015). The rules that indicate whether a guide will lead to downregulation are unclear, and it is possible that successful design relies on certain promoter features that will not be present for all genes. For example, Kiani et al. indicate that placement of a guide downstream of a TATA box can lead to downregulation of gene expression, but this feature is only found in 24% of promoters (Yang et al., 2007) and even when present, is likely to require the testing of many more guides than is necessary for achieving upregulation of the same target. The inability to reliably upregulate endogenous gene expression here may reflect a need to optimise the experimental conditions, rather than any fundamental barrier to the use of Cas9-VPR in this way – upregulation of transiently transfected DNA was reproducibly observed here (see figure 6.2), and work showing endogenous upregulation with Cas9-VP64 in CHO was published in October 2017 (Kleinjan et al., 2017).

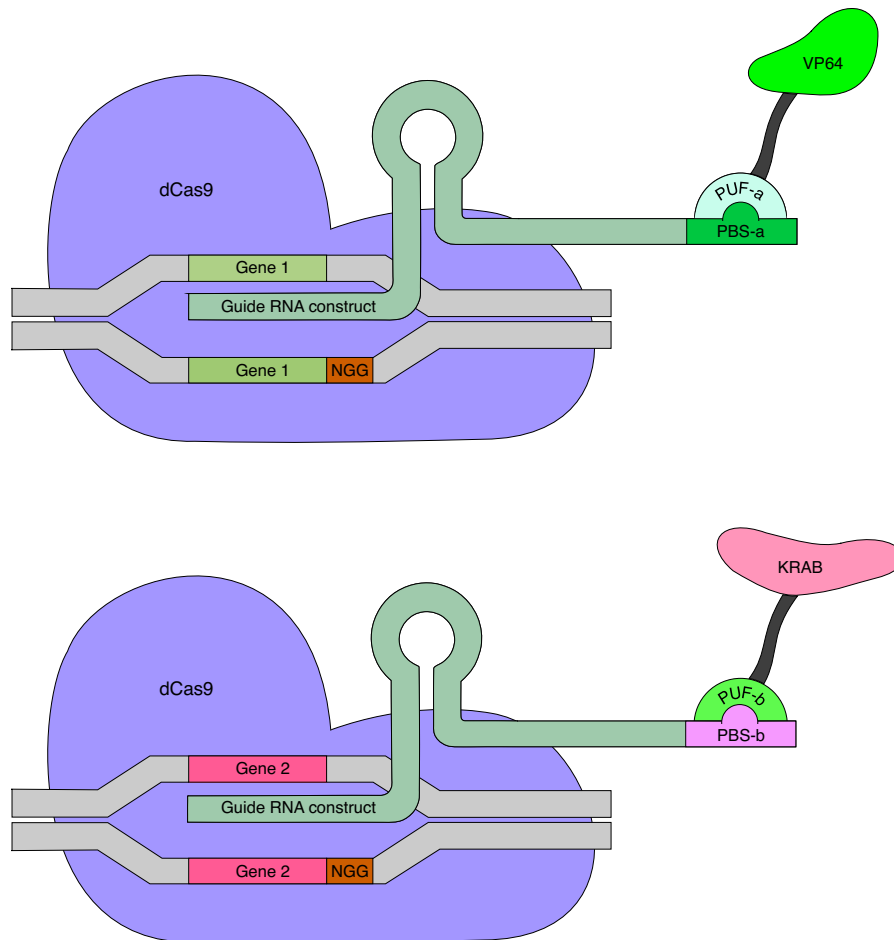


Figure 6.4. The Casilio system uses orthogonal pairs of Pumilio RNA-binding domains and cognate RNA octamer sequences to achieve simultaneous transactivation and transrepression at different genomic loci (Cheng et al., 2016). The catalytically inactive dCas9 (which has no fused effector domains) is targeted to a given gene via a guide RNA construct, which contains a downstream PUF-binding site (PBS). This PBS can be bound only by a specific Pumilio/fem-3 binding factor (PUF), which is fused to a specific effector domain. For example, the figure shows PUF-a fused to VP64, which is only able to bind to PBS-a, targeting gene 1 for upregulation (top). PUF-b, fused to the KRAB domain, is only able to bind to PBS-b and so will repress gene 2 (bottom).

Given the demonstrable utility of being able to simultaneously up- and downregulate gene expression with regards to lactate accumulation, it may be useful to explore alternative approaches to achieving this goal which would allow the rapid design, construction and testing of an RNA-guided gene targeting system. For instance, it could be possible to use a catalytically inactive form of the RNA-guided endonuclease Cpf1 (which typically require a T-rich protospacer adjacent motif for target binding - Zetsche et al., 2015) as an orthogonal

means of gene targeting, and for which transrepression and transactivation with KRAB and VPR has been shown, respectively (Liu et al., 2017) - such a system could be used in parallel with the Cas9-VPR system described here. Additionally, the Casilio system uses multiple programmable Pumilio RNA-binding domains fused to different effector domains (e.g. VP64, p65, KRAB), and due to the Pumilio domain's ability to recognise only specific RNA sequences, the type of effector targeted to a specific gene can be chosen by varying the guide RNA scaffold (see figure 6.4). Multiple simultaneous and orthogonal genomic manipulations are thus possible with this system (Cheng et al., 2016).

7 Discussion and future work

7.1 Summary of the results

This project aimed to develop a lactate-inducible transgene expression system for use in Chinese hamster ovary cells, based on the one component transcriptional regulator LldR from *Corynebacterium glutamicum* and *Pseudomonas aeruginosa*. Initial *in vivo* experiments did not demonstrate any lactate-dependent induction of reporter gene expression, although they did indicate that the use of a nuclear localisation signal (NLS) was not necessary for efficient nuclear trafficking. The absence of a positive result meant it was not possible at this stage to identify the basis on which the system was not functioning. It was thus necessary to further characterise and optimise the components of the inducible system *in vitro*, in novel contrast from previous work establishing mammalian inducible transgene expression systems from similar components. This *in vitro* work found that the binding affinity of LldR for its operator could be improved by altering the bases flanking the pseudopalindromic sequence to more closely resemble those found in the natural promoter. Increasing the distance between a pair of operators also improved its apparent binding affinity to LldR. A fusion of VP64 to the N-terminus of LldR from *Corynebacterium glutamicum* abolished its DNA-binding interaction, while it was maintained by a C-terminal fusion, albeit significantly weaker than with the wild-type protein. The DNA-binding interaction of LldR-VP64 from *Corynebacterium glutamicum* to its operator could be partially restored by the use of a flexible linker. The LldR from *Pseudomonas aeruginosa* was able to bind to the operator from *Corynebacterium glutamicum* with a similar affinity to that of its counterpart, and this interaction was not significantly diminished by the fusion of VP64 to its C-terminus. These optimisations were subsequently employed in an additional round of *in vivo* testing. Transient transfections were both able to show statistically significant levels of induction for trans-activating configurations in response to lactate over 3 independent experiments, each consisting of 3 biological replicates. No significant levels of induction were seen for de-repression by steric hindrance configurations. It was generally the case that designs using multiple operators upstream of the minCMV promoter gave better fold induction.

Work was carried out to establish that Cas9-VPR could be used to target genes known or hypothesised to have a beneficial impact on lactate accumulation when targeted for up/down-regulation, such that their dynamic regulation could later be tested with the lactate-inducible

transgene expression system. Cas9-VPR was successfully used to upregulate expression of a transiently transfected plasmid, showing that it was functional in principle, although the experimental conditions used did not reliably upregulate the endogenous MCT1 gene.

7.2 Future work

This section describes additional work that could be carried out over the short-term, i.e. manipulations that individually would be straightforward to implement, as well as complementary work to be performed over a longer timeframe with approaches beyond those employed in this project.

With regards to chapter 4, the DNA-binding capacity of LldR from *Corynebacterium glutamicum* was diminished by the fusion of an effector domain and was only partially restored by the introduction of a flexible linker between these two components. Given the importance of DNA-binding strength to enhance the induction of a transgene expression system (e.g. Krell et al., 2007), it could be worthwhile to optimise the sequence of the flexible linker in order to fully restore the interaction of LldR with its operator. Simply extending the length of the flexible linker may be sufficient.

A more granular investigation of the optimal inter-operator spacing could be carried out. While increasing the spacing from 11 bases to 23 improved the apparent binding affinity of LldR for this sequence, it remains to be seen whether it could have been improved further with additional bases. If fewer intervening bases are required, this would allow a greater density of operators between a promoter and a transcription start site, meaning that i. higher de-repressed expression from an LldR response element would be seen (essentially, the greater the number of bases between a promoter and a transcription start site, the weaker the expression of that gene), and ii. the multiple operator sites would maximise the steric hindrance of the promoter by LldR in the absence of lactate.

It may be helpful to characterise the *in vitro* binding behaviour of LldR to sequences containing three or more operator repeats, which may show diminishing returns in terms of apparent binding affinity and therefore provide valuable information concerning the optimal number of repeats.

Previous work from Georgi et al. (2008) and Gao et al. (2012) showing the lactate-dependent DNA-binding behaviour of LldR was only partially replicated, for reasons that are unclear. Based on experiments performed by Gao et al. (2008), the work carried out here used a shorter operator sequence than those used by Georgi et al. (2008) and Gao et al. (2012); additionally, *in vitro* experiments here used a C-terminal hexahistidine purification tag (compared to a likely N-terminal location of the same tag in Georgi et al. (2008) and Gao et al. (2008), which cannot be eliminated as a potential cause of the inability of LldR to fully unbind from its operator in the presence of lactate in experiments performed here. Although it is unclear how either point of difference can explain the observed partial unbinding behaviour, testing a larger operator construct or modifying the position of the purification tag may be illuminating.

The initial *in vivo* experiments employed the KRAB domain as a trans-repressor, in keeping with its use for this purpose in previous projects (e.g. Deuschle et al., 1995, Groner et al., 2010). Subsequent attempts to characterise the impact of its fusion to LldR failed, owing to the “profound insolubility” of the KRAB domain when specifically expressed in *Escherichia coli* (Peng et al., 2000), and as a result it was not used in later *in vivo* experiments. Although it might be possible to successfully isolate LldR fused to the KRAB domain by varying purification conditions or by expressing these constructs in mammalian cells (where there is no indication that there would be any solubility issues), a more fruitful alternative could be to use the HDAC4 protein as a trans-repression domain instead. Having been used in this context previously (e.g. Bockamp et al., 2007), with rapid and fully reversible gene silencing properties compared to KRAB, which is partially irreversible (Bintu et al., 2016), and with apparently straightforward expression in *Escherichia coli* (Bottomley et al., 2008), in the future it may be preferable to characterise and employ the HDAC4 domain instead of KRAB. The benefit of such a system is that it can readily be used to activate (or ‘de-repress’) transcription in response to lactate, as opposed to the transactivation configuration, which counterintuitively sees the down-regulation of transcription in the presence of lactate.

Multiple operators (3 or more) would likely be able to improve the extent of induction observed, as in previous work on inducible transgene expression systems (e.g. Gossen & Bujard, 1992), as well as the observation in chapter 5. This is because multiple operator sites increase the probability that the DNA-binding protein will be found adjacent to a promoter, where it will be able to exert some influence on gene expression. Although 3 or more operators were tested in the initial round of *in vivo* testing, time constraints meant that this was not repeated in the

later round - cloning multiple repeats of the pseudopalindromic operator sequence had to be performed stepwise in order to avoid misannealing and undesired cloning products.

The LldR proteins used here were chosen on the basis that they had been studied previously, with evidence that they exhibit the characteristics necessary for use in a lactate-inducible transgene expression system in mammalian cells. However, they only represent two out of many such LldR proteins, with 1166 sequences found on UniProt, a database of protein sequences (Bateman et al., 2016), and it is possible that another LldR would be a more suitable candidate for use in a lactate-inducible system on the basis of operator-binding affinity and/or responsiveness to lactate. A precedent for screening multiple previously uncharacterised proteins comes from Bayer et al. (2009), who synthesised a diverse set of methyl halide transferases and were able to discover enzymes with significantly improved rates of catalysis. Multiple LldR proteins could be screened rapidly using cell-free expression, which allows expression from linear PCR products without the need for lengthy cloning and purification steps (Rosenblum & Cooperman, 2014), and which could then be tested for activity against a library of operator-promoter constructs in *in vitro* experiments using GFP as an output (e.g. Chappell et al., 2013). This approach could be modified to assay the activity of randomised LldR sequences from *Corynebacterium glutamicum* and *Pseudomonas aeruginosa* currently present in our lab, by applying error-prone PCR or related techniques (e.g. McCullum et al., 2010).

It may be the case that the operator of LldR could be optimised to improve its protein-DNA affinity. It is unlikely that *Corynebacterium glutamicum* has been under any strong selection pressure to increase the affinity of this interaction, as the response element that LldR targets contains two adjacent sites (cgl2917-site1 and cgl2917-site2) so close to each other that they reduce the apparent affinity of LldR for this sequence by a factor of 14, relative to the affinity of LldR for cgl2917-site1 by itself (Gao et al., 2008). SELEX (systematic evolution of ligands by exponential enrichment) is one approach that could be employed to generate a more tightly-bound operator, which works by passing a pool of DNA sequences through a column containing immobilised protein, with the protein-bound sequences eluted under increasingly stringent conditions (Hasegawa et al., 2016).

In the longer-term, it would be sensible to test any highly-inducible system that is identified under conditions more closely resembling that of a bioproduction run. The *in vivo* experiments carried out were necessarily performed at a small-scale in order to rapidly test multiple

conditions, but this differs from the actual bioproduction conditions in which the inducible system would ultimately be used. For example, the CHO-S cells used here are not overexpressing a therapeutic protein in a high-glucose media, nor are they being subjected to the substantial stresses of cultivation in a bioreactor. These differences are known to have significant impacts on lactate accumulation phenotypes, which may only be partially replicated by the addition of lactate in the experimental set-up used here. For instance, the cell lines tested here may rapidly be consuming the additional lactate, unlike many production cell lines which will only do so after multiple days of cultivation. Once a robust and highly-inducible system is identified at a small-scale, it would be sensible to test it in a therapeutic-producing cell line that is known to accumulate lactate, alongside the dynamic regulation of genes that have previously been identified as being able to mitigate the impact of lactate.

It will be important to establish that up/down-regulating given endogenous genes via Cas9-VPR will have a beneficial impact on the lactate accumulation phenotype, either when expressed constitutively or when expressed via a dynamic lactate induction system. Hypotheses about the benefits of targeting multiple genes can be tested.

The use of LldR to detect and respond to lactate is promising, based on the results of chapter 5 (the second round of *in vivo* testing). However, if its use in such a system transpires to be only weakly inducible in mammalian cells after subsequent rounds of testing and optimisation, it may be prudent to consider alternative approaches to lactate-dependent gene expression.

Some mammalian genes have been identified as being lactate-responsive, and it may be possible to use the promoters from these genes to dynamically drive expression in CHO cells in response to lactate. For example, Formby & Stern (2003) identified 4 genes (CD44, caveolin-1, Hyal-1 and Hyal-2, all involved in wound repair) that were upregulated in response to 20 mM lactate, although they did not quantify this increase. In each of these genes they found multiple copies of the ets-1 and AP-1 response elements, which are targeted by Ets family of transcription factors and AP-1 for upregulation, respectively. Carbonic anhydrase IX (CA IX) expression has been shown elsewhere to be upregulated 2.4-fold in response to 10 mM lactate, containing multiple elements in its promoter protected in a DNaseI assay, including an AP-1 binding site, a hypoxia-response element (bound by the hypoxia inducible factor 1 α transcription factor), and an SP-1 binding site (bound by the SP-1 transcription factor) (Panisova et al., 2017). Other lactate-responsive genes could be identified using transcriptomic approaches - Hsu et al. (2017) used RNA-seq to find 3 genes that were

upregulated in response to low levels of lactate. Characterisation of as-yet unidentified lactate-responsive promoters could allow the randomisation of their response elements (alongside ets-1 and AP-1) in synthetic promoters, which could be screened for stronger lactate-response behaviour in CHO cells than that seen with natural promoters.

Certain mammalian proteins have been identified as interacting with lactate in non-enzymatic ways that could be leveraged in a lactate-inducible system. For example, the stabilisation of HIF1- α (hypoxia inducible factor 1 α) by 5 mM lactate has been shown to increase transcription of *Alp* and *Osteocalcin* 2-fold (Wu et al., 2017). The NDRG3 protein, which mediates the activation of the Raf-ERK signalling pathway, binds directly to lactate, stabilising NDRG3 by inducing a conformational change that conceals a ubiquitination site (Lee et al., 2015). This results in the accumulation of NDRG3 and allows it to phosphorylate key proteins in the Raf-ERK signalling pathway, resulting in the activation of AP-1 transcription factors that go on to upregulate associated genes. Beyond simply using the promoters targeted by NDRG3 and HIF1 α , a lactate-inducible transgene expression system could be built around these proteins themselves. NDRG3 could be fused to a DNA-binding domain and transcriptional activator, allowing it to interact with the promoter of a gene known to mitigate lactate accumulation. In the absence of lactate, the NDRG3 fusion will be targeted for ubiquitination and will not be able to upregulate expression; in the presence of lactate, the NDRG3 fusion will no longer be degraded and will be able to upregulate expression. In order to construct an orthogonal system that does not have an excessive impact on endogenous signalling pathways in CHO, it might be necessary to mutate NDRG3 so that it does not maintain its multiple protein-protein binding interactions, with a possible focus on its phosphorylation mechanism described by Melotte et al., 2010. It may also be possible to use HIF1 α in this way, although it is not clear that it is stabilised by lactate in the same way as NDRG3.

The SugR protein binds to multiple, unique operators, and will only unbind from them in the presence of a cognate sugar phosphate (including fructose-1-phosphate, fructose-6-phosphate, fructose-1,6-bisphosphate, and glucose-6-phosphate) (Gaigalat et al., 2007, Engels & Wendisch, 2007, Toyoda et al., 2009). Although the levels of these sugar phosphates tend to fall within a narrow range in CHO cells that are not overexpressing a protein of interest (e.g. Hsu et al., 2017), if it was determined that they varied throughout a bioprocessing run in a way that correlated with lactate, it would be possible to use SugR to regulate transcription as an indirect response to lactate.

Proteins that undergo a general conformational change in response to lactate could theoretically be used in a lactate-inducible transgene expression system, without necessitating a lactate-dependent interaction with DNA, if the conformational change presented a sufficiently distinct epitope. For example, Pook et al. (1998) were able to develop multiple antibodies to TetR that would either bind strongly to the TetR-operator complex (and not to free TetR or TetR complexed to tetracycline), or vice versa. Many proteins undergoing conformational changes in response to lactate have been described in this thesis, including LldR. The LldR from *Escherichia coli* remains bound to its operator whether lactate is present or not (Aguilera et al., 2008); if an antibody fused to a transcriptional effector domain could be specifically targeted to one of the lactate-dependent conformational states of this LldR, it would be possible to up- or downregulate expression from an adjacent promoter in response to lactate.

The accumulation of lactate during cultivation of mammalian cells for biopharmaceutical production is a longstanding issue affecting glycosylation quality and productivity. Many approaches exist to mitigate its impact, either through the replacement of glucose with slowly metabolised sugars, dynamic feeding strategies, or host cell engineering. The manipulation of genes in this latter approach is constitutive and may suboptimally respond to cellular needs. The LldR proteins from *Corynebacterium glutamicum* and *Pseudomonas aeruginosa* have been used in this project to create a lactate-inducible transgene expression system, which can be used subsequently to dynamically drive expression of proteins previously targeted to mitigate the accumulation of lactate. Some preliminary work was also carried out with Cas9-VPR, which will be a useful tool for upregulating multiple previously identified targets, as well as helping to find new beneficial targets. The general approach outlined here for the development of this lactate-inducible transgene expression system will be appropriate for any other such project where the ligand of interest is a central metabolite. Inherently weaker induction might be a feature of such a system, given the presence of the inducer at low and relatively benign or neutral concentrations throughout a period of interest, as well as its possible degradation; testing an unoptimised system in mammalian cells may return little or no detectable induction signal and therefore it will not be straightforward to optimise such a system solely through the use of *in vivo* experimentation. *In vitro* characterisation of the inducible system components (as performed here) can provide essential feedback regarding the impact of effector domain fusion and operator design on the DNA-binding affinity of the biosensor prior to *in vivo* testing.

8 References

- Aggarwal, R. S. 2014. What's fueling the biotech engine-2012 to 2013. *Nature Biotechnology*, 32(1), 32-39.
- Agha-Mohammadi, S., O'Malley, M., Etemad, A., Wang, Z., Xiao, X. & Lotze, M. T. 2004. Second-generation tetracycline-regulatable promoter: repositioned tet operator elements optimize transactivator synergy while shorter minimal promoter offers tight basal leakiness. *Journal of Gene Medicine*, 6(7), 817-28.
- Aguilera, L., Campos, E., Gimenez, R., Badia, J., Aguilar, J. & Baldoma, L. 2008. Dual role of LldR in regulation of the lldPRD operon, involved in L-lactate metabolism in *Escherichia coli*. *Journal of Bacteriology*, 190(8), 2997-3005.
- Ahn, W. S. & Antoniewicz, M. R. 2013. Parallel labeling experiments with [1, 2-¹³C] glucose and [U-¹³C] glutamine provide new insights into CHO cell metabolism. *Metabolic Engineering*, 15(1), 1534-47.
- Aitken, M., & Kleinrock, M. (2017). *Lifetime Trends in Biopharmaceutical Innovation: Recent Evidence and Implications*. [pdf] Parsippany, New Jersey: IQVIA Institute for Human Data Science. Available at <https://www.iqvia.com/institute/reports/lifetime-trends-in-biopharmaceutical-innovation-recent-evidence-and-implications> [Accessed 23 Jan. 2019]
- Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. 2005. Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences*, 102(36), 12678-12683.
- Altamirano, C., Illanes, A., Becerra, S., Cairo, J. J. & Godia, F. 2006. Considerations on the lactate consumption by CHO cells in the presence of galactose. *Journal of Biotechnology*, 125(4), 547-56.
- Altamirano, C., Paredes, C., Illanes, A., Cairo, J. & Godia, F. 2004. Strategies for fed-batch cultivation of t-PA producing CHO cells: substitution of glucose and glutamine and rational design of culture medium. *Journal of Biotechnology*, 110(2), 171-179.

- Amarzguioui, M. & Prydz, H. 2004. An algorithm for selection of functional siRNA sequences. *Biochemical and Biophysical Research Communications*, 316(4), 1050-8.
- An, Z. 2010. Monoclonal antibodies - a proven and rapidly expanding therapeutic modality for human diseases. *Protein & Cell*, 1(4), 319-330.
- Andersen, D. C. & Goochee, C. F. 1995. The effect of ammonia on the O-linked glycosylation of granulocyte colony-stimulating factor produced by Chinese hamster ovary cells. *Biotechnology and Bioengineering*, 47(1), 96-105.
- Anesiadis, N., Cluett, W. R. & Mahadevan, R. 2008. Dynamic metabolic engineering for increasing bioprocess productivity. *Metabolic Engineering*, 10(5), 255-66.
- Antonetti, F., Finocchiaro, O., Mascia, M., Terlizze, M. G. & Jaber, A. 2002. Short Communication: A Comparison of the Biologic Activity of Two Recombinant IFN- β Preparations Used in the Treatment of Relapsing-Remitting Multiple Sclerosis. *Journal of Interferon & Cytokine Research*, 22(12), 1181-1184.
- Aranda, P. S., Lajoie, D. M. & Jorcyk, C. L. 2012. Bleach gel: a simple agarose gel for analyzing RNA quality. *Electrophoresis*, 33(2), 366-9.
- Athmaram, T. N., Singh, A. K., Saraswat, S., Srivastava, S., Misra, P., Kameswara Rao, M., Gopalan, N. & Rao, P. V. 2013. A simple *Pichia pastoris* fermentation and downstream processing strategy for making recombinant pandemic Swine Origin Influenza a virus Hemagglutinin protein. *Journal of Industrial Microbiology and Biotechnology*, 40(2), 245-55.
- Aubin, J. E. 1979. Autofluorescence of viable cultured mammalian cells. *Journal of Histochemistry & Cytochemistry*, 27(1), 36-43.
- Bacchus, W., Lang, M., El-Baba, M. D., Weber, W., Stelling, J. & Fussenegger, M. 2012. Synthetic two-way communication between mammalian cells. *Nature Biotechnology*, 30(10), 991-6.

- Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M., Ramadan, H. A., Saini, K. S., Baeshen, N. A. & Redwan, E. M. 2015. Production of Biopharmaceuticals in *E. coli*: Current Scenario and Future Perspectives. *Journal of Microbiology and Biotechnology*, 25(7), 953-62.
- Baneyx, F. & Mujacic, M. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology*, 22(11), 1399-408.
- Barnes, L. M., Bentley, C. M. & Dickson, A. J. 2003. Stability of protein production from recombinant mammalian cells. *Biotechnology and Bioengineering*, 81(6), 631-9.
- Barrangou, R. & Doudna, J. A. 2016. Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*, 34(9), 933-941.
- Bayer, T. S., Widmaier, D. M., Temme, K., Mirsky, E. A., Santi, D. V. & Voigt, C. A. 2009. Synthesis of methyl halides from biomass using engineered microbes. *Journal of the American Chemical Society*, 131(18), 6508-15.
- Berli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F. 1998. Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences*, 95(25), 14628-14633.
- Benchling [Biology Software]. 2019. Retrieved from <https://benchling.com>.
- Berlec, A. & Strukelj, B. 2013. Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *Journal of Industrial Microbiology and Biotechnology*, 40(3-4), 257-74.
- Berrios, J., Altamirano, C., Osses, N. & Gonzalez, R. 2011. Continuous CHO cell cultures with improved recombinant protein productivity by using mannose as carbon source: Metabolic analysis and scale-up simulation. *Chemical Engineering Science*, 66(11), 2431-2439.

- Bintu, L., Yong, J., Antebi, Y. E., Mccue, K., Kazuki, Y., Uno, N., Oshimura, M. & Elowitz, M. B. 2016. Dynamics of epigenetic regulation at the single-cell level. *Science*, 351(6274), 720-4.
- Bockamp, E., Christel, C., Hameyer, D., Khobta, A., Maringer, M., Reis, M., Heck, R., Cabezas-Wallscheid, N., Epe, B., Oesch-Bartlomowicz, B., Kaina, B., Schmitt, S. & Eshkind, L. 2007. Generation and characterization of tTS-H4: a novel transcriptional repressor that is compatible with the reverse tetracycline-controlled TET-ON system. *Journal of Gene Medicine*, 9(4), 308-18.
- Borys, M. C., Linzer, D. I. & Papoutsakis, E. T. 1994. Ammonia affects the glycosylation patterns of recombinant mouse placental lactogen-I by Chinese hamster ovary cells in a pH-dependent manner. *Biotechnology and Bioengineering*, 43(6), 505-514.
- Boshart, M., Weber, F., Jahn, G., Dorschler, K., Fleckenstein, B. & Schaffner, W. 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell*, 41(2), 521-530.
- Bottomley, M. J., Lo Surdo, P., Di Giovine, P., Cirillo, A., Scarpelli, R., Ferrigno, F., Jones, P., Neddermann, P., De Francesco, R., Steinkuhler, C., Gallinari, P. & Carfi, A. 2008. Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain. *Journal of Biological Chemistry*, 283(39), 26694-704.
- Brenowitz, M., Senear, D. F. & Kingston, R. E. 2001. DNase I footprint analysis of protein-DNA binding. *Current Protocols in Molecular Biology*, Chapter 12 Unit 12.4.
- Brinkmann, U., Mattes, R. E. & Buckel, P. 1989. High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product. *Gene*, 85(1), 109-114.
- Brockman, I. M. & Prather, K. L. 2015. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnology Journal*, 10(9), 1360-9.

- Brown, A. J., Sweeney, B., Mainwaring, D. O. & James, D. C. 2014. Synthetic promoters for CHO cell engineering. *Biotechnology & Bioengineering*, 111(8), 1638-47.
- Brutinel, E. D. & Gralnick, J. A. 2012. Preferential utilization of D-lactate by *Shewanella oneidensis*. *Applied and Environmental Microbiology*, 78(23), 8474-6.
- Bylund, F., Collet, E., Enfors, S. O. & Larsson, G. 1998. Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. *Bioprocess Engineering*, 18(3), 171.
- Carey, B. W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M. & Jaenisch, R. 2009. Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proceedings of the National Academy of Sciences*, 106(1), 157-62.
- Cedar, H. & Bergman, Y. 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics*, 10(5), 295.
- Chappell, J., Jensen, K. & Freemont, P. S. 2013. Validation of an entirely *in vitro* approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Research*, 41(5), 3471-81.
- Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., E, P. R. I., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B. E., Perrimon, N., Weiss, R., Aach, J., Collins, J. J. & Church, G. M. 2015. Highly efficient Cas9-mediated transcriptional programming. *Nature Methods*, 12(4), 326-8.
- Chee Fung Wong, D., Tin Kam Wong, K., Tang Goh, L., Kiat Heng, C. & Gek Sim Yap, M. 2005. Impact of dynamic online fed-batch strategies on metabolism, productivity and N-glycosylation quality in CHO cell cultures. *Biotechnology and Bioengineering*, 89(2), 164-77.
- Chen, C. & Chasin, L. A. 1998. Cointegration of DNA Molecules Introduced into Mammalian Cells by Electroporation. *Somatic Cell and Molecular Genetics*, 24(4), 249-256.

- Chen, C. D. & Sawyers, C. L. 2002. NF- B Activates Prostate-Specific Antigen Expression and Is Upregulated in Androgen-Independent Prostate Cancer. *Molecular and Cellular Biology*, 22(8), 2862-2870.
- Chen, P. & Harcum, S. W. 2007. Differential display identifies genes in Chinese hamster ovary cells sensitive to elevated ammonium. *Applied Biochemistry and Biotechnology*, 141(2-3), 349-359.
- Chen, X., Zaro, J. L. & Shen, W. C. 2013. Fusion protein linkers: property, design and functionality. *Advanced Drug Delivery Reviews*, 65(10), 1357-69.
- Cheng, A. W., Jillette, N., Lee, P., Plaskon, D., Fujiwara, Y., Wang, W., Taghbalout, A. & Wang, H. 2016. Casilio: a versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling. *Cell Research*, 26(2), 254-7.
- Chiang, J., Gloff, C. A., Yoshizawa, C. N. & Williams, G. J. 1993. Pharmacokinetics of recombinant human interferon-beta ser in healthy volunteers and its effect on serum neopterin. *Pharmaceutical Research*, 10(4), 567-572.
- Chong, W. P., Reddy, S. G., Yusufi, F. N., Lee, D. Y., Wong, N. S., Heng, C. K., Yap, M. G. & Ho, Y. S. 2010. Metabolomics-driven approach for the improvement of Chinese hamster ovary cell growth: overexpression of malate dehydrogenase II. *Journal of Biotechnology*, 147(2), 116-21.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. & Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), 819-23.
- Consortium, T. U. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45(D1), D158-D169.
- Corish, P. & Tyler-Smith, C. 1999. Attenuation of green fluorescent protein half-life in mammalian cells. *Protein engineering*, 12(12), 1035-1040.

- Cossons, N., Hayter, P., Tuite, M. & Jenkins, N. 1991. Stability of amplified DNA in Chinese hamster ovary cells. *Production of biologicals from animal cells in culture*. Elsevier.
- Cost, G. J., Freyvert, Y., Vafiadis, A., Santiago, Y., Miller, J. C., Rebar, E., Collingwood, T. N., Snowden, A. & Gregory, P. D. 2010. BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. *Biotechnology and Bioengineering*, 105(2), 330-40.
- Courtois, F., Schneider, C. P., Agrawal, N. J. & Trout, B. L. 2015. Rational Design of Biobetters with Enhanced Stability. *Journal of Pharmaceutical Sciences*, 104(8), 2433-40.
- Cregg, J. M. 2007. Introduction: distinctions between *Pichia pastoris* and other expression systems. *Methods in Molecular Biology*, 3891-10.
- Cronin, C. A., Gluba, W. & Scrable, H. 2001. The lac operator-repressor system is functional in the mouse. *Genes & Development*, 15(12), 1506-17.
- Cuthbertson, L. & Nodwell, J. R. 2013. The TetR family of regulators. *Microbiology and Molecular Biology Reviews*, 77(3), 440-75.
- Dalton, A. C. & Barton, W. A. 2014. Over-expression of secreted proteins from mammalian cell lines. *Protein Science*, 23(5), 517-25.
- Das, A. T., Tenenbaum, L. & Berkhout, B. 2016. Tet-On Systems For Doxycycline-inducible Gene Expression. *Current Gene Therapy*, 16(3), 156-167.
- De Los Santos, E. L., Meyerowitz, J. T., Mayo, S. L. & Murray, R. M. 2016. Engineering Transcriptional Regulator Effector Specificity Using Computational Design and *In Vitro* Rapid Prototyping: Developing a Vanillin Sensor. *ACS Synthetic Biology*, 5(4), 287-95.
- De Ruijter, A. J., Van Gennip, A. H., Caron, H. N., Kemp, S. & Van Kuilenburg, A. B. 2003. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochemical Journal*, 370(Pt 3), 737-49.

- Dean, N. 1999. Asparagine-linked glycosylation in the yeast Golgi. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1426(2), 309-322.
- Dekker, L. & Polizzi, K. M. 2017. Sense and sensitivity in bioprocessing-detecting cellular metabolites with biosensors. *Current Opinion in Chemical Biology*, 4031-36.
- Desguin, B., Goffin, P., Bakouche, N., Diman, A., Viaene, E., Dandoy, D., Fontaine, L., Hallet, B. & Hols, P. 2015. Enantioselective regulation of lactate racemization by LarR in *Lactobacillus plantarum*. *Journal of Bacteriology*, 197(1), 219-30.
- Deuschle, U., Meyer, W. & Thiesen, H. 1995. Tetracycline-reversible silencing of eukaryotic promoters *Molecular and Cellular Biology*, 15(4), 1907-1914.
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J. & Root, D. E. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), 184-191.
- Doerks, T., Copley, R. R., Schultz, J., Ponting, C. P. & Bork, P. 2002. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Research*, 12(1), 47-56.
- Ducrest, A.-L., Amacker, M., Lingner, J. & Nabholz, M. 2002. Detection of promoter activity by flow cytometric analysis of GFP reporter expression. *Nucleic Acids Research*, 30(14), e65.
- Dzyadevych, S. V., Arkhypova, V. N., Soldatkin, A. P., El'skaya, A. V., Martelet, C. & Jaffrezic-Renault, N. 2008. Amperometric enzyme biosensors: Past, present and future. *LRBM*, 29(2-3), 171-180.
- Ede, C., Chen, X., Lin, M. Y. & Chen, Y. Y. 2016. Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. *ACS Synthetic Biology*, 5(5), 395-404.

- Ellis, T., Evans, D. A., Martin, C. R. & Hartley, J. A. 2007. A 96-well DNase I footprinting screen for drug-DNA interactions. *Nucleic Acids Research*, 35(12), e89.
- Engels, V. & Wendisch, V. F. 2007. The DeoR-type regulator SugR represses expression of ptsG in *Corynebacterium glutamicum*. *Journal of Bacteriology*, 189(8), 2955-66.
- England, J. L. 2013. Statistical physics of self-replication. *The Journal of Chemical Physics*, 139(12), 121923.
- Engohang-Ndong, J., Baillat, D., Aumercier, M., Bellefontaine, F., Besra, G. S., Locht, C. & Baulard, A. R. 2003. EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. *Molecular Microbiology*, 51(1), 175-188.
- Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjostedt, E., Lundberg, E., Szigyarto, C. A., Skogs, M., Takanen, J. O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J. M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., Von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F. & Uhlen, M. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics*, 13(2), 397-406.
- Feng, W., Huth, J. R., Norton, S. E. & Ruddon, R. W. 1995. Asparagine-linked oligosaccharides facilitate human chorionic gonadotropin beta-subunit folding but not assembly of prefolded beta with alpha. *Endocrinology*, 136(1), 52-61.
- Fischer, S., Handrick, R. & Otte, K. 2015. The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology Advances*, 33(8), 1878-96.
- Folcher, M., Xie, M., Spinner, A. & Fussenegger, M. 2013. Synthetic mammalian trigger-controlled bipartite transcription factors. *Nucleic Acids Research*, 41(13), e134.

- Formby, B. & Stern, R. 2003. Lactate-sensitive response elements in genes involved in hyaluronan catabolism. *Biochemical and Biophysical Research Communications*, 305(1), 203-208.
- Franceschini, A., Meier, R., Casanova, A., Kreibich, S., Daga, N., Andritschke, D., Dilling, S., Ramo, P., Emmenlauer, M., Kaufmann, A., Conde-Alvarez, R., Low, S. H., Pelkmans, L., Helenius, A., Hardt, W. D., Dehio, C. & Von Mering, C. 2014. Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proceedings of the National Academy of Sciences*, 111(12), 4548-53.
- Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1), 92-105.
- Fu, T., Zhang, C., Jing, Y., Jiang, C., Li, Z., Wang, S., Ma, K., Zhang, D., Hou, S., Dai, J., Kou, G. & Wang, H. 2016. Regulation of cell growth and apoptosis through lactate dehydrogenase C over-expression in Chinese hamster ovary cells. *Applied Microbiology and Biotechnology*, 100(11), 5007-16.
- Furbish, F. S., Steer, C. J., Barranger, J. A., Jones, E. A. & Brady, R. O. 1978. The uptake of native and desialylated glucocerebrosidase by rat hepatocytes and Kupffer cells. *Biochemical and Biophysical Research Communications*, 81(3), 1047-1053.
- Furlong, I. J., Ascaso, R., Rivas, A. L. & Collins, M. 1997. Intracellular acidification induces apoptosis by stimulating ICE-like protease activity. *Journal of Cell Science*, 110(5), 653-661.
- Fussenegger, M., Morris, R. P., Fux, C., Rimann, M., Von Stockar, B., Thompson, C. J. & Bailey, J. E. 2000. Streptogramin-based gene regulation systems for mammalian cells. *Nature Biotechnology*, 18(11), 1203-8.
- Gagnon, M., Hiller, G., Luan, Y. T., Kittredge, A., Defelice, J. & Drapeau, D. 2011. High-end pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures. *Biotechnology and Bioengineering*, 108(6), 1328-37.

- Gaigalat, L., Schluter, J. P., Hartmann, M., Mormann, S., Tauch, A., Puhler, A. & Kalinowski, J. 2007. The DeoR-type transcriptional regulator SugR acts as a repressor for genes encoding the phosphoenolpyruvate:sugar phosphotransferase system (PTS) in *Corynebacterium glutamicum*. *BMC Molecular Biology*, 8104.
- Gao, C., Hu, C., Zheng, Z., Ma, C., Jiang, T., Dou, P., Zhang, W., Che, B., Wang, Y., Lv, M. & Xu, P. 2012. Lactate utilization is regulated by the FadR-type regulator LldR in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 194(10), 2687-92.
- Gao, Y. G., Suzuki, H., Itou, H., Zhou, Y., Tanaka, Y., Wachi, M., Watanabe, N., Tanaka, I. & Yao, M. 2008. Structural and functional characterization of the LldR from *Corynebacterium glutamicum*: a transcriptional repressor involved in L-lactate and sugar utilization. *Nucleic Acids Research*, 36(22), 7110-23.
- Garriga-Canut, M., Agustin-Pavon, C., Herrmann, F., Sanchez, A., Dierssen, M., Fillat, C. & Isalan, M. 2012. Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proceedings of the National Academy of Sciences*, 109(45), E3136-45.
- Gasser, B. & Mattanovich, D. 2007. Antibody production with yeasts and filamentous fungi: on the road to large scale? *Biotechnology Letters*, 29(2), 201-12.
- Gasser, B., Maurer, M., Gach, J., Kunert, R. & Mattanovich, D. 2006. Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnology and Bioengineering*, 94(2), 353-61.
- Gemmill, T. R. & Trimble, R. B. 1999. Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1426(2), 227-237.
- Georgi, T., Engels, V. & Wendisch, V. F. 2008. Regulation of L-lactate utilization by the FadR-type regulator LldR of *Corynebacterium glutamicum*. *Journal of Bacteriology*, 190(3), 963-71.

- Ghorbaniaghdam, A., Chen, J., Henry, O. & Jolicoeur, M. 2014. Analyzing clonal variation of monoclonal antibody-producing CHO cell lines using an in silico metabolomic platform. *PLoS One*, 9(3), e90832.
- Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S. & Qi, L. S. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2), 442-51.
- Gitzinger, M., Kemmer, C., El-Baba, M. D., Weber, W. & Fussenegger, M. 2009. Controlling transgene expression in subcutaneous implants using a skin lotion containing the apple metabolite phloretin. *Proceedings of the National Academy of Sciences*, 106(26), 10638-43.
- Gitzinger, M., Kemmer, C., Fluri, D. A., El-Baba, M. D., Weber, W. & Fussenegger, M. 2012. The food additive vanillic acid controls transgene expression in mammalian cells and mice. *Nucleic Acids Research*, 40(5), e37.
- Glazyrina, J., Materne, E.-M., Dreher, T., Storm, D., Junne, S., Adams, T., Grellner, G. & Neubauer, P. 2010. High cell density cultivation and recombinant protein production with *Escherichia coli* in a rocking-motion-type bioreactor. *Microbial Cell Factories*, 9(1), 42.
- Goers, L., Ainsworth, C., Goey, C. H., Kontoravdi, C., Freemont, P. S. & Polizzi, K. M. 2017. Whole-cell *Escherichia coli* lactate biosensor for monitoring mammalian cell cultures during biopharmaceutical production. *Biotechnology and Bioengineering*, 114(6), 1290-1300.
- Goers, L., Kyllis, N., Tomazou, M., Yan Wen, K., Freemont, P. & Polizzi, K. 2013. Engineering Microbial Biosensors. 40119-156.
- Gogolok, S., Garcia-Diaz, C. & Pollard, S. M. 2016. STAR: a simple TAL effector assembly reaction using isothermal assembly. *Scientific Reports*, 633209.

- Gossen, M. & Bujard, H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*, 89(12), 5547-5551.
- Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. & Bujard, H. 1995. Transcriptional activation by tetracyclines in mammalian cells. *Science*, 268(5218), 1766-1769.
- Goulas, T., Cuppari, A., Garcia-Castellanos, R., Snipas, S., Glockshuber, R., Arolas, J. L. & Gomis-Ruth, F. X. 2014. The pCri System: a vector collection for recombinant protein expression and purification. *PLoS One*, 9(11), e112643.
- Greber, D. & Fussenegger, M. 2010. An engineered mammalian band-pass network. *Nucleic Acids Research*, 38(18), e174.
- Greisman, H. A. 1997. A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites. *Science*, 275(5300), 657-661.
- Groner, A. C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P. & Trono, D. 2010. KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLOS Genetics*, 6(3), e1000869.
- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science*, 293(5532), 1146-50.
- Hansen, H. A. & Emborg, C. 1994. Influence of ammonium on growth, metabolism, and productivity of a continuous suspension Chinese hamster ovary cell culture. *Biotechnology progress*, 10(1), 121-124.
- Hardwick, J. M. & Soane, L. 2013. Multiple functions of BCL-2 family proteins. *Cold Spring Harbor Perspectives in Biology*, 5(2).

- Hartenbach, S., Daoud-El Baba, M., Weber, W. & Fussenegger, M. 2007. An engineered L-arginine sensor of *Chlamydia pneumoniae* enables arginine-adjustable transcription control in mammalian cells and mice. *Nucleic Acids Research*, 35(20), e136.
- Hartley, F., Walker, T., Chung, V. & Morten, K. 2018. Mechanisms driving the lactate switch in Chinese hamster ovary cells. *Biotechnology and Bioengineering*, 115(8), 1890-1903.
- Hasegawa, H., Savory, N., Abe, K. & Ikebukuro, K. 2016. Methods for Improving Aptamer Binding Affinity. *Molecules*, 21(4), 421.
- Hashizume, R., Maki, Y., Mizutani, K., Takahashi, N., Matsubara, H., Sugita, A., Sato, K., Yamaguchi, S. & Mikami, B. 2011. Crystal structures of protein-glutaminase and its pro forms converted into enzyme-substrate complex. *Journal of Biological Chemistry*, jbc.M111. 255133.
- Heller-Harrison, R., Crowe, K., Cooley, C., Hone, M., Mccarthy, K. & Leonard, M. 2009. Managing cell line instability and its impact during cell line development. *BioPharm International*.
- Hellman, L. M. & Fried, M. G. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, 2(8), 1849-61.
- Hodel, A. E., Harreman, M. T., Pulliam, K. F., Harben, M. E., Holmes, J. S., Hodel, M. R., Berland, K. M. & Corbett, A. H. 2006. Nuclear localization signal receptor affinity correlates with *in vivo* localization in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 281(33), 23545-56.
- Homann, S., Hofmann, C., Gorin, A. M., Nguyen, H. C. X., Huynh, D., Hamid, P., Maithel, N., Yacoubian, V., Mu, W., Kossyvakis, A., Sen Roy, S., Yang, O. O. & Kelesidis, T. 2017. A novel rapid and reproducible flow cytometric method for optimization of transfection efficiency in cells. *PLoS One*, 12(9), e0182941.

- Houde, D., Peng, Y., Berkowitz, S. A. & Engen, J. R. 2010. Post-translational modifications differentially affect IgG1 conformation and receptor binding. *Molecular & Cellular Proteomics*, 9(8), 1716-28.
- Hsu, H. H., Araki, M., Mochizuki, M., Hori, Y., Murata, M., Kahar, P., Yoshida, T., Hasunuma, T. & Kondo, A. 2017. A Systematic Approach to Time-series Metabolite Profiling and RNA-seq Analysis of Chinese hamster ovary Cell Culture. *Scientific Reports*, 743518.
- Hu, W., Berdugo, C. & Chalmers, J. J. 2011. The potential of hydrodynamic damage to animal cells of industrial relevance: current understanding. *Cytotechnology*, 63(5), 445-60.
- Ike, K., Arasawa, Y., Koizumi, S., Mihashi, S., Kawai-Noma, S., Saito, K. & Umeno, D. 2015. Evolutionary Design of Choline-Inducible and -Repressible T7-Based Induction Systems. *ACS Synthetic Biology*, 4(12), 1352-60.
- Inoue, H., Nojima, H. & Okayama, H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, 96(1), 23-28.
- Jackson, A. L., Burchard, J., Schelter, J., Chau, B. N., Cleary, M., Lim, L. & Linsley, P. S. 2006. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA*, 12(7), 1179-87.
- Jaffe, S. R., Strutton, B., Levarski, Z., Pandhal, J. & Wright, P. C. 2014. *Escherichia coli* as a glycoprotein production host: recent developments and challenges. *Current Opinion in Biotechnology*, 30205-10.
- Jahic, M., Rotticci-Mulder, J., Martinelle, M., Hult, K. & Enfors, S.-O. 2002. Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. *Bioprocess and Biosystems Engineering*, 24(6), 385-393.
- Jain, D. 2015. Allosteric control of transcription in GntR family of transcription regulators: A structural overview. *IUBMB Life*, 67(7), 556-63.

- Jayapal, K. P., Wlaschin, K. F., Hu, W. S. & Yap, M. G. S. 2007. Recombinant protein therapeutics from CHO Cells - 20 years and counting. *Chemical Engineering Progress*, 103(10), 40-47.
- Jeon, M. K., Yu, D. Y. & Lee, G. M. 2011. Combinatorial engineering of Idh-a and bcl-2 for reducing lactate production and improving cell growth in dihydrofolate reductase-deficient Chinese hamster ovary cells. *Applied Microbiology and Biotechnology*, 92(4), 779-790.
- Jeong, D., Kim, T. S., Lee, J. W., Kim, K. T., Kim, H. J., Kim, I. H. & Kim, I. Y. 2001. Blocking of acidosis-mediated apoptosis by a reduction of lactate dehydrogenase activity through antisense mRNA expression. *Biochemical and Biophysical Research Communications*, 289(5), 1141-9.
- Jeong, D. W., Cho, I. T., Kim, T. S., Bae, G. W., Kim, I. H. & Kim, I. Y. 2006. Effects of lactate dehydrogenase suppression and glycerol-3-phosphate dehydrogenase overexpression on cellular metabolism. *Molecular and Cellular Biochemistry*, 284(1-2), 1-8.
- Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P. & Breitsprecher, D. 2014. MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, 1077101-113.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816-21.
- Jones, W. & Bianchi, K. 2015. Aerobic glycolysis: beyond proliferation. *Frontiers in immunology*, 6227.
- Jun, S. C., Kim, M. S., Hong, H. J. & Lee, G. M. 2006. Limitations to the development of humanized antibody producing Chinese hamster ovary cells using glutamine synthetase-mediated gene amplification. *Biotechnology progress*, 22(3), 770-780.

- Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. 1984. A short amino acid sequence able to specify nuclear location. *Cell*, 39(3), 499-509.
- Kane, J. F. 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology*, 6(5), 494-500.
- Karasawa, S., Araki, T., Yamamoto-Hino, M. & Miyawaki, A. 2003. A green-emitting fluorescent protein from *Galaxeidae* coral and its monomeric version for use in fluorescent labeling. *Journal of Biological Chemistry*, 278(36), 34167-71.
- Karlsson, M., Weber, W. & Fussenegger, M. 2011. De novo design and construction of an inducible gene expression system in mammalian cells. *Methods in Enzymology*, 497:239-53.
- Kelly, P. S., Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N. T., Meleady, P., O'gorman, D., Clynes, M. & Barron, N. 2015. Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype. *Biotechnology Journal*, 10(7), 1029-40.
- Kemmer, C., Gitzinger, M., Daoud-El Baba, M., Djonov, V., Stelling, J. & Fussenegger, M. 2010. Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Nature Biotechnology*, 28(4), 355-60.
- Khan, O. A., Xia, Q., Bever, C. T., Johnson, K. P., Panitch, H. S. & Dhib-Jalbut, S. S. 1996. Interferon beta-1b serum levels in multiple sclerosis patients following subcutaneous administration. *Neurology*, 46(6), 1639-1643.
- Khawli, L. A., Goswami, S., Hutchinson, R., Kwong, Z. W., Yang, J., Wang, X., Yao, Z., Sreedhara, A., Cano, T., Tesar, D., Nijem, I., Allison, D. E., Wong, P. Y., Kao, Y. H., Quan, C., Joshi, A., Harris, R. J. & Motchnik, P. 2010. Charge variants in IgG1: Isolation, characterization, *in vitro* binding properties and pharmacokinetics in rats. *MAbs*, 2(6), 613-24.
- Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B. W., Beal, J., Vora, S., Buchthal, J., Kowal, E. J., Ebrahimkhani, M. R., Collins, J. J.,

- Weiss, R. & Church, G. 2015. Cas9 gRNA engineering for genome editing, activation and repression. *Nature Methods*, 12(11), 1051-4.
- Kim, B. G. & Park, H. W. 2016. Tetrahydrofolate increases suspension growth of dihydrofolate reductase-deficient Chinese hamster ovary DG44 cells in chemically defined media. *Biotechnology progress*, 32(6), 1539-1546.
- Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., Kim, M. K., Shin, B. A. & Choi, S. Y. 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One*, 6(4), e18556.
- Kim, L. K., Esplugues, E., Zorca, C. E., Parisi, F., Kluger, Y., Kim, T. H., Galjart, N. J. & Flavell, R. A. 2014. Oct-1 regulates IL-17 expression by directing interchromosomal associations in conjunction with CTCF in T cells. *Molecular Cell*, 54(1), 56-66.
- Kim, N. S., Kim, S. J. & Lee, G. M. 1998. Clonal variability within dihydrofolate reductase-mediated gene amplified Chinese hamster ovary cells: Stability in the absence of selective pressure. *Biotechnology and Bioengineering*, 60(6), 679-688.
- Kim, S. H. & Lee, G. M. 2007. Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. *Applied Microbiology and Biotechnology*, 74(1), 152-9.
- Kim, S. H. & Lee, G. M. 2007. Functional expression of human pyruvate carboxylase for reduced lactic acid formation of Chinese hamster ovary cells (DG44). *Applied Microbiology and Biotechnology*, 76(3), 659-65.
- Kimura, R. & Miller, W. M. 1996. Effects of elevated pCO₂ and/or osmolality on the growth and recombinant tPA production of CHO cells. *Biotechnology and Bioengineering*, 52(1), 152-160.
- Kingston, R. E., Kaufman, R. J., Bebbington, C. R. & Rolfe, M. R. 2002. Amplification using CHO cell expression vectors. *Current Protocols in Molecular Biology*, Chapter 16 Unit 16 23.

- Kis, Z., Pereira, H. S., Homma, T., Pedrigi, R. M. & Krams, R. 2015. Mammalian synthetic biology: emerging medical applications. *Journal of the Royal Society Interface*, 12(106).
- Kiss, Z., Elliott, S., Jedynasty, K., Tesar, V. & Szegedi, J. 2010. Discovery and basic pharmacology of erythropoiesis-stimulating agents (ESAs), including the hyperglycosylated ESA, darbepoetin alfa: an update of the rationale and clinical impact. *European Journal of Clinical Pharmacology*, 66(4), 331-40.
- Kleinjan, D. A., Wardrope, C., Nga Sou, S. & Rosser, S. J. 2017. Drug-tunable multidimensional synthetic gene control using inducible degron-tagged dCas9 effectors. *Nature Communications*, 8(1), 1191.
- Kompella, U. B. & Lee, V. H. L. 1991. Pharmacokinetics of peptide and protein drugs. *In*: Lee, V. H. L. (ed.) *Peptide and Protein Drug Delivery, Advances in Parenteral Sciences*. New York, USA: Marcel Dekker.
- Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., Nureki, O. & Zhang, F. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517(7536), 583-8.
- Koslowski, M., Türeci, Ö., Bell, C., Krause, P., Lehr, H. A., Brunner, J., Seitz, G., Nestle, F. O., Huber, C. & Sahin, U. 2002. Multiple splice variants of lactate dehydrogenase C selectively expressed in human cancer. *Cancer Research*, 62(22), 6750-6755.
- Krainer, F. W., Gmeiner, C., Neutsch, L., Windwarder, M., Pletzenauer, R., Herwig, C., Altmann, F., Glieder, A. & Spadiut, O. 2013. Knockout of an endogenous mannosyltransferase increases the homogeneity of glycoproteins produced in *Pichia pastoris*. *Scientific Reports*, 33279.
- Kramer, B. P., Fischer, C. & Fussenegger, M. 2004. BioLogic gates enable logical transcription control in mammalian cells. *Biotechnology and Bioengineering*, 87(4), 478-84.

- Krebs, H. & Bellamy, D. 1960. The interconversion of glutamic acid and aspartic acid in respiring tissues. *Biochemical Journal*, 75(3), 523.
- Krell, T., Teran, W., Mayorga, O. L., Rivas, G., Jimenez, M., Daniels, C., Molina-Henares, A. J., Martinez-Bueno, M., Gallegos, M. T. & Ramos, J. L. 2007. Optimization of the palindromic order of the TtgR operator enhances binding cooperativity. *Journal of Molecular Biology*, 369(5), 1188-99.
- Kunert, R., Gach, J. & Katinger, H. 2008. Expression of a Fab Fragment in CHO and *Pichia pastoris*. *BioProcess International*, 6(6).
- Kurano, N., Leist, C., Messi, F., Kurano, S. & Fiechter, A. 1990. Growth behavior of Chinese hamster ovary cells in a compact loop bioreactor. 2. Effects of medium components and waste products. *Journal of Biotechnology*, 15(1-2), 113-128.
- Kyriakopoulos, S. & Kontoravdi, C. 2014. A framework for the systematic design of fed-batch strategies in mammalian cell culture. *Biotechnology and Bioengineering*, 111(12), 2466-2476.
- Lai, L., Hui, C.-K., Leung, N. & Lau, G. K. 2006. Pegylated interferon alpha-2a (40 kDa) in the treatment of chronic hepatitis B. *International Journal of Nanomedicine*, 1(3), 255-262.
- Larsen, S., Weaver, J., De Sa Campos, K., Bulahan, R., Nguyen, J., Grove, H., Huang, A., Low, L., Tran, N., Gomez, S., Yau, J., Ilustrisimo, T., Kawilarang, J., Lau, J., Tranphung, M., Chen, I., Tran, C., Fox, M., Lin-Cereghino, J. & Lin-Cereghino, G. P. 2013. Mutant strains of *Pichia pastoris* with enhanced secretion of recombinant proteins. *Biotechnology Letters*, 35(11), 1925-35.
- Le, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., Karypis, G. & Hu, W. S. 2012. Multivariate analysis of cell culture bioprocess data--lactate consumption as process indicator. *Journal of Biotechnology*, 162(2-3), 210-23.
- Le, H., Vishwanathan, N., Kantardjieff, A., Doo, I., Srienc, M., Zheng, X., Somia, N. & Hu, W. S. 2013. Dynamic gene expression for metabolic engineering of mammalian cells in culture. *Metabolic Engineering*, 20212-20.

- Leader, B., Baca, Q. J. & Golan, D. E. 2008. Protein therapeutics: a summary and pharmacological classification. *Nature Reviews Drug Discovery*, 721.
- Lederer, T., Takahashi, M. & Hillen, W. 1995. Thermodynamic analysis of tetracycline-mediated induction of Tet repressor by a quantitative methylation protection assay. *Analytical Biochemistry*, 232(2), 190-6.
- Lee, D. C., Sohn, H. A., Park, Z. Y., Oh, S., Kang, Y. K., Lee, K. M., Kang, M., Jang, Y. J., Yang, S. J., Hong, Y. K., Noh, H., Kim, J. A., Kim, D. J., Bae, K. H., Kim, D. M., Chung, S. J., Yoo, H. S., Yu, D. Y., Park, K. C. & Yeom, Y. I. 2015. A lactate-induced response to hypoxia. *Cell*, 161(3), 595-609.
- Lee, J. S., Ha, T. K., Park, J. H. & Lee, G. M. 2013. Anti-cell death engineering of CHO cells: co-overexpression of Bcl-2 for apoptosis inhibition, Beclin-1 for autophagy induction. *Biotechnology and Bioengineering*, 110(8), 2195-207.
- Lim, S. F., Chuan, K. H., Liu, S., Loh, S. O., Chung, B. Y., Ong, C. C. & Song, Z. 2006. RNAi suppression of Bax and Bak enhances viability in fed-batch cultures of CHO cells. *Metabolic Engineering*, 8(6), 509-22.
- Lim, Y., Wong, N. S., Lee, Y. Y., Ku, S. C., Wong, D. C. & Yap, M. G. 2010. Engineering mammalian cells in bioprocessing - current achievements and future perspectives. *Biotechnology and Applied Biochemistry*, 55(4), 175-89.
- Lin, A. A., Kimura, R. & Miller, W. M. 1993. Production of tPA in recombinant CHO cells under oxygen-limited conditions. *Biotechnology and Bioengineering*, 42(3), 339-50.
- Lindsley, C. W. 2017. New 2016 Data and Statistics for Global Pharmaceutical Products and Projections through 2017. *ACS Chemical Neuroscience*, 8(8), 1635-1636.
- Liu, B., Spearman, M., Doering, J., Lattova, E., Perreault, H. & Butler, M. 2014. The availability of glucose to CHO cells affects the intracellular lipid-linked oligosaccharide distribution, site occupancy and the N-glycosylation profile of a monoclonal antibody. *Journal of Biotechnology*, 17017-27.

- Liu, Y., Han, J., Chen, Z., Wu, H., Dong, H. & Nie, G. 2017. Engineering cell signaling using tunable CRISPR-Cpf1-based transcription factors. *Nature Communications*, 8(1), 2095.
- Longo, P. A., Kavran, J. M., Kim, M. S. & Leahy, D. J. 2013. Generating mammalian stable cell lines by electroporation. *Methods in Enzymology*, 529209-26.
- Luo, M., Pang, C. W., Gerken, A. E. & Brock, T. G. 2004. Multiple nuclear localization sequences allow modulation of 5-lipoxygenase nuclear import. *Traffic*, 5(11), 847-54.
- Ma, N., Ellet, J., Okediadi, C., Hermes, P., McCormick, E. & Casnocha, S. 2009. A single nutrient feed supports both chemically defined NS0 and CHO fed-batch processes: Improved productivity and lactate metabolism. *Biotechnology progress*, 25(5), 1353-63.
- Mack, L., Brill, B., Delis, N. & Groner, B. 2014. Endotoxin depletion of recombinant protein preparations through their preferential binding to histidine tags. *Analytical Biochemistry*, 46683-8.
- Maddocks, S. E. & Oyston, P. C. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology*, 154(Pt 12), 3609-23.
- Majorek, K. A., Kuhn, M. L., Chruszcz, M., Anderson, W. F. & Minor, W. 2014. Double trouble- Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Science*, 23(10), 1359-68.
- Malphettes, L., Weber, C. C., El-Baba, M. D., Schoenmakers, R. G., Aubel, D., Weber, W. & Fussenegger, M. 2005. A novel mammalian expression system derived from components coordinating nicotine degradation in arthrobacter nicotinovorans pAO1. *Nucleic Acids Research*, 33(12), e107.
- Mannini, R., Riviaccio, V., D'auria, S., Tanfani, F., Ausili, A., Facchiano, A., Pedone, C. & Grimaldi, G. 2006. Structure/function of KRAB repression domains: structural

properties of KRAB modules inferred from hydrodynamic, circular dichroism, and FTIR spectroscopic analyses. *Proteins*, 62(3), 604-16.

Marks, D. M. 2003. Equipment design considerations for large scale cell culture. *Cytotechnology*, 42(1), 21-33.

Marschall, L., Sagmeister, P. & Herwig, C. 2017. Tunable recombinant protein expression in *E. coli*: promoter systems and genetic constraints. *Applied Microbiology and Biotechnology*, 101(2), 501-512.

Martinez, V. S., Dietmair, S., Quek, L. E., Hodson, M. P., Gray, P. & Nielsen, L. K. 2013. Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. *Biotechnology and Bioengineering*, 110(2), 660-6.

Mathur, M., Xiang, J. S. & Smolke, C. D. 2017. Mammalian synthetic biology for studying the cell. *Journal of Cell Biology*, 216(1), 73-82.

Matthews, T. E., Berry, B. N., Smelko, J., Moretto, J., Moore, B. & Wiltberger, K. 2016. Closed loop control of lactate concentration in mammalian cell culture by Raman spectroscopy leads to improved cell density, viability, and biopharmaceutical protein production. *Biotechnology and Bioengineering*, 113(11), 2416-24.

McCullum, E. O., Williams, B. A., Zhang, J. & Chaput, J. C. 2010. Random mutagenesis by error-prone PCR. *Methods in Molecular Biology*, 634103-9.

Melotte, V., Qu, X., Ongenaert, M., Van Crielinge, W., De Bruine, A. P., Baldwin, H. S. & Van Engeland, M. 2010. The N-myc downstream regulated gene (NDRG) family: diverse functions, multiple applications. *FASEB Journal*, 24(11), 4153-66.

Mignon, C., Sodoyer, R. & Werle, B. 2015. Antibiotic-free selection in biotherapeutics: now and forever. *Pathogens*, 4(2), 157-81.

Moscou, M. J. & Bogdanove, A. J. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science*, 326(5959), 1501.

- Mullick, A., Xu, Y., Warren, R., Koutroumanis, M., Guilbault, C., Broussau, S., Malenfant, F., Bourget, L., Lamoureux, L., Lo, R., Caron, A. W., Pilotte, A. & Massie, B. 2006. The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC Biotechnology*, 643.
- Mulukutla, B. C., Gramer, M. & Hu, W. S. 2012. On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. *Metabolic Engineering*, 14(2), 138-49.
- Neddermann, P., Gargioli, C., Muraglia, E., Sambucini, S., Bonelli, F., De Francesco, R. & Cortese, R. 2003. A novel, inducible, eukaryotic gene expression system based on the quorum-sensing transcription factor TraR. *EMBO Reports*, 4(2), 159-65.
- Nivitchanyong, T., Martinez, A., Ishaque, A., Murphy, J. E., Konstantinov, K., Betenbaugh, M. J. & Thrift, J. 2007. Anti-apoptotic genes Aven and E1B-19K enhance performance of BHK cells engineered to express recombinant factor VIII in batch and low perfusion cell culture. *Biotechnology and Bioengineering*, 98(4), 825-41.
- Nwokeoji, A. O., Kilby, P. M., Portwood, D. E. & Dickman, M. J. 2016. RNASwift: A rapid, versatile RNA extraction method free from phenol and chloroform. *Analytical Biochemistry*, 51236-46.
- Nylen, A. & Chen, M. T. 2018. Production of Full-Length Antibody by *Pichia pastoris*. *Methods in Molecular Biology*, 167437-48.
- O'Gorman, S., Fox, D. & Wahl, G. 1991. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science*, 251(4999), 1351-1355.
- Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. 2000. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature Structural & Molecular Biology*, 7(3), 215-9.
- Osterlehner, A., Simmeth, S. & Gopfert, U. 2011. Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. *Biotechnology and Bioengineering*, 108(11), 2670-81.

- Overton, T. W. 2014. Recombinant protein production in bacterial hosts. *Drug Discovery Today*, 19(5), 590-601.
- Ozturk, S. S., Riley, M. R. & Palsson, B. O. 1992. Effects of ammonia and lactate on hybridoma growth, metabolism, and antibody production. *Biotechnology and Bioengineering*, 39(4), 418-31.
- Pais-Chanfrau, J. & Trujillo-Toledo, L. 2016. Optimization of culture medium for large-scale production of heterologous proteins in *Pichia pastoris* to be used in nanoscience and other biotechnological fields. *Biology and Medicine*, 8(3), 1.
- Panisova, E., Kery, M., Sedlakova, O., Brisson, L., Debreova, M., Sboarina, M., Sonveaux, P., Pastorekova, S. & Svastova, E. 2017. Lactate stimulates CA IX expression in normoxic cancer cells. *Oncotarget*, 8(44), 77819-77835.
- Pasquinelli, A. E. 2012. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nature Reviews Genetics*, 13(4), 271-82.
- Pavletich, N. & Pabo, C. 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science*, 261(5129), 1701-1707.
- Peng, H., Begg, G. E., Harper, S. L., Friedman, J. R., Speicher, D. W. & Rauscher, F. J., 3rd 2000. Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. *Journal of Biological Chemistry*, 275(24), 18000-10.
- Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. 2013. Enhancers: five essential questions. *Nature Reviews Genetics*, 14(4), 288-95.
- Pfaffl, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29(9), 45e-45.
- Pfeiffer, F. & Mayer, G. 2016. Selection and Biosensor Application of Aptamers for Small Molecules. *Frontiers in Chemistry*, 425.

- Piatkevich, K. D. & Verkhusha, V. V. 2011. Guide to red fluorescent proteins and biosensors for flow cytometry. *Methods in Cell Biology*, 102431-61.
- Piirainen, M. A., De Ruijter, J. C., Koskela, E. V. & Frey, A. D. 2014. Glycoengineering of yeasts from the perspective of glycosylation efficiency. *Nature Biotechnology*, 31(6), 532-7.
- Pilbrough, W., Munro, T. P. & Gray, P. 2009. Intracloonal protein expression heterogeneity in recombinant CHO cells. *PLoS One*, 4(12), e8432.
- Pisarsky, L., Bill, R., Fagiani, E., Dimeloe, S., Goosen, R. W., Hagmann, J., Hess, C. & Christofori, G. 2016. Targeting Metabolic Symbiosis to Overcome Resistance to Anti-angiogenic Therapy. *Cell Reports*, 15(6), 1161-74.
- Pook, E., Grimm, S., Bonin, A., Winkler, T. & Hillen, W. 1998. Affinities of mAbs to Tet repressor complexed with operator or tetracycline suggest conformational changes associated with induction. *European Journal of Biochemistry*, 258(3), 915-922.
- Pope, B. & Kent, H. M. 1996. High efficiency 5 min transformation of *Escherichia coli*. *Nucleic Acids Research*, 24(3), 536-537.
- Porter, A. J., Racher, A. J., Preziosi, R. & Dickson, A. J. 2010. Strategies for selecting recombinant CHO cell lines for cGMP manufacturing: improving the efficiency of cell line generation. *Biotechnology progress*, 26(5), 1455-64.
- Raj, A. & Van Oudenaarden, A. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2), 216-26.
- Ramirez, C. L., Foley, J. E., Wright, D. A., Muller-Lerch, F., Rahman, S. H., Cornu, T. I., Winfrey, R. J., Sander, J. D., Fu, F., Townsend, J. A., Cathomen, T., Voytas, D. F. & Joung, J. K. 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods*, 5(5), 374-5.

- Recillas-Targa, F. 2006. Multiple Strategies for Gene Transfer, Expression, Knockdown, and Chromatin Influence in Mammalian Cell Lines and Transgenic Animals. *Molecular Biotechnology*, 34(3), 337-354.
- Reichert, J. M. 2003. Trends in development and approval times for new therapeutics in the United States. *Nature Reviews Drug Discovery*, 2(9), 695-702.
- Rohatgi, A. 2016. *WebPlotDigitizer: Web based tool to extract data from plots, images, and maps* [Online].
- Romanova, N. & Noll, T. 2018. Engineered and Natural Promoters and Chromatin-Modifying Elements for Recombinant Protein Expression in CHO Cells. *Biotechnology Journal*, 13(3), 1700232.
- Rosenblum, G. & Cooperman, B. S. 2014. Engine out of the chassis: cell-free protein synthesis and its uses. *FEBS Letters*, 588(2), 261-8.
- Rossger, K., Charpin-El-Hamri, G. & Fussenegger, M. 2014. Bile acid-controlled transgene expression in mammalian cells and mice. *Metabolic Engineering*, 2181-90.
- Sadeghi, H. M. M., Rabbani, M., Rismani, E., Moazen, F., Khodabakhsh, F., Dormiani, K. & Khazaei, Y. 2011. Optimization of the expression of reteplase in *Escherichia coli*. *Research in Pharmaceutical Sciences*, 6(2), 87–92.
- Saito, K., Koshiba, S., Inoue, M., Kigawa, T. & Yokoyama, S. N.d. Solution structure of the Kruppel-associated box (KRAB) domain. *To be published*.
- Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. & Yamamoto, T. 2014. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Scientific Reports*, 45400.
- San Martin, A., Ceballo, S., Ruminot, I., Lerchundi, R., Frommer, W. B. & Barros, L. F. 2013. A genetically encoded FRET lactate sensor and its use to detect the Warburg effect in single cancer cells. *PLoS One*, 8(2), e57712.

- Sandu, C., Chiribau, C. B. & Brandsch, R. 2003. Characterization of HdnR, the transcriptional repressor of the 6-hydroxy-D-nicotine oxidase gene of *Arthrobacter nicotinovorans* pAO1, and its DNA-binding activity in response to L- and D-nicotine Derivatives. *Journal of Biological Chemistry*, 278(51), 51307-15.
- Sawaki, N., Tsujimoto, R., Shigyo, M., Konishi, M., Toki, S., Fujiwara, T. & Yanagisawa, S. 2013. A nitrate-inducible GARP family gene encodes an auto-repressible transcriptional repressor in rice. *Plant and Cell Physiology*, 54(4), 506-17.
- Scalcinati, G., Knuf, C., Partow, S., Chen, Y., Maury, J., Schalk, M., Daviet, L., Nielsen, J. & Siewers, V. 2012. Dynamic control of gene expression in *Saccharomyces cerevisiae* engineered for the production of plant sesquiterpene alpha-santalene in a fed-batch mode. *Metabolic Engineering*, 14(2), 91-103.
- Schein, C. H. 1990. Solubility as a Function of Protein Structure and Solvent Components. *Nature Biotechnology*, 8(4), 308-317.
- Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671-675.
- Schumacher, M. A., Miller, M. C. & Brennan, R. G. 2004. Structural mechanism of the simultaneous binding of two drugs to a multidrug-binding protein. *EMBO Journal*, 23(15), 2923-30.
- Serrato, J. A., Palomares, L. A., Meneses-Acosta, A. & Ramírez, O. T. 2004. Heterogeneous conditions in dissolved oxygen affect N-glycosylation but not productivity of a monoclonal antibody in hybridoma cultures. *Biotechnology and Bioengineering*, 88(2), 176-188.
- Sezonov, G., Joseleau-Petit, D. & D'ari, R. 2007. *Escherichia coli* physiology in Luria-Bertani broth. *Journal of Bacteriology*, 189(23), 8746-9.
- Sherman, F. 1991. [1] Getting started with yeast. *Methods in Enzymology*. Elsevier.

- Shiloach, J. & Fass, R. 2005. Growing *E. coli* to high cell density--a historical perspective on method development. *Biotechnology Advances*, 23(5), 345-57.
- Simon, A. E., Taylor, M. W., Bradley, W. E. & Thompson, L. H. 1982. Model involving gene inactivation in the generation of autosomal recessive mutants in mammalian cells in culture. *Molecular and Cellular Biology*, 2(9), 1126-1133.
- Skjoedt, M. L., Snoek, T., Kildegaard, K. R., Arsovska, D., Eichenberger, M., Goedecke, T. J., Rajkumar, A. S., Zhang, J., Kristensen, M., Lehka, B. J., Siedler, S., Borodina, I., Jensen, M. K. & Keasling, J. D. 2016. Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nature Chemical Biology*, 12(11), 951-958.
- Sladitschek, H. L. & Neveu, P. A. 2015. MXS-Chaining: A Highly Efficient Cloning Platform for Imaging and Flow Cytometry Approaches in Mammalian Systems. *PLoS One*, 10(4), e0124958.
- Stanton, B. C., Siciliano, V., Ghodasara, A., Wroblewska, L., Clancy, K., Trefzer, A. C., Chesnut, J. D., Weiss, R. & Voigt, C. A. 2014. Systematic transfer of prokaryotic sensors and circuits to mammalian cells. *ACS Synthetic Biology*, 3(12), 880-91.
- Strandberg, L. & Enfors, S.-O. 1991. Batch and fed batch cultivations for the temperature induced production of a recombinant protein in *Escherichia coli*. *Biotechnology Letters*, 13(8), 609-614.
- Striedner, G., Cserjan-Puschmann, M., Potschacher, F. & Bayer, K. 2003. Tuning the transcription rate of recombinant protein in strong *Escherichia coli* expression systems through repressor titration. *Biotechnology progress*, 19(5), 1427-32.
- Struhl, K. 1999. Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell*, 98(1), 1-4.
- Stryer, L., Berg, J. & Tymoczko, J. 2006. Biochemistry: International Edition. New York, WH Freeman & Co Ltd.

- Suvorova, I. A., Korostelev, Y. D. & Gelfand, M. S. 2015. GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PLoS One*, 10(7), e0132618.
- Szperalski, B., Jung, C., Shao, Z., Kantardjieff, A. & Hu, W. S. 2011. LDH-C can be differentially expressed during fermentation of CHO cells. *BMC Proceedings*, 5 Suppl 8P107.
- Tantama, M., Hung, Y. P. & Yellen, G. 2011. Imaging intracellular pH in live cells with a genetically encoded red fluorescent protein sensor. *Journal of the American Chemical Society*, 133(26), 10034-7.
- Taher, T. E. I. 2017. Monitoring Promoter Activity by Flow Cytometry. *Methods in Molecular Biology*, 165165-73.
- Tjio, J. H. & Puck, T. T. 1958. Genetics of somatic mammalian cells: II. Chromosomal constitution of cells in tissue culture. *The Journal of experimental medicine*, 108(2), 259.
- Toussaint, C., Henry, O. & Durocher, Y. 2016. Metabolic engineering of CHO cells to alter lactate metabolism during fed-batch cultures. *Journal of Biotechnology*, 217122-31.
- Toyoda, K., Teramoto, H., Inui, M. & Yukawa, H. 2009. Molecular mechanism of SugR-mediated sugar-dependent expression of the *ldhA* gene encoding L-lactate dehydrogenase in *Corynebacterium glutamicum*. *Applied Microbiology and Biotechnology*, 83(2), 315-27.
- Troy, T., Jekic-Mcmullen, D., Sambucetti, L. & Rice, B. 2004. Quantitative Comparison of the Sensitivity of Detection of Fluorescent and Bioluminescent Reporters in Animal Models. *Molecular Imaging*, 3(1), 153535002004031.
- Ulloa-Aguirre, A., Timossi, C., Damián-Matsumura, P. & Dias, J. A. 1999. Role of Glycosylation in Function of Follicle-Stimulating Hormone. *Endocrine*, 11(3), 205-216.

- Ulrich, L. E., Koonin, E. V. & Zhulin, I. B. 2005. One-component systems dominate signal transduction in prokaryotes. *Trends in Microbiology*, 13(2), 52-6.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115.
- Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. 2009. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*, 324(5930), 1029-33.
- Vera, A., Gonzalez-Montalban, N., Aris, A. & Villaverde, A. 2007. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and Bioengineering*, 96(6), 1101-6.
- Wahrheit, J., Nicolae, A. & Heinzle, E. 2014. Dynamics of growth and metabolism controlled by glutamine availability in Chinese hamster ovary cells. *Applied Microbiology and Biotechnology*, 98(4), 1771-83.
- Walls, P. L. L., Mcrae, O., Natarajan, V., Johnson, C., Antoniou, C. & Bird, J. C. 2017. Quantifying the potential for bursting bubbles to damage suspended cells. *Scientific Reports*, 7(1), 15102.
- Walsh, G. 2005. Biopharmaceuticals: recent approvals and likely directions. *Trends in Biotechnology*, 23(11), 553-8.
- Walsh, G. 2006. Biopharmaceutical benchmarks 2006. *Nature Biotechnology*, 24(10), 2476-9.
- Walsh, G. 2010. Biopharmaceutical benchmarks 2010. *Nature Biotechnology*, 28(9), 917.
- Walsh, G. & Jefferis, R. 2006. Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnology*, 24(10), 1241-52.
- Wang, H., Ye, H., Xie, M., Daoud El-Baba, M. & Fussenegger, M. 2015. Cosmetics-triggered percutaneous remote control of transgene expression in mice. *Nucleic Acids Research*, 43(14), e91.

- Wang, R. & Brattain, M. G. 2007. The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. *FEBS Letters*, 581(17), 3164-70.
- Wang, Z., Zhang, L. & Tan, T. 2010. High cell density fermentation of *Saccharomyces cerevisiae* GS2 for selenium-enriched yeast production. *Korean Journal of Chemical Engineering*, 27(6), 1836-1840.
- Weber, E., Gruetzner, R., Werner, S., Engler, C. & Marillonnet, S. 2011. Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One*, 6(5), e19722.
- Weber, W. 2003. Streptomyces-derived quorum-sensing systems engineered for adjustable transgene expression in mammalian cells and mice. *Nucleic Acids Research*, 31(14), 71e-71.
- Weber, W. 2003. Conditional human VEGF-mediated vascularization in chicken embryos using a novel temperature-inducible gene regulation (TIGR) system. *Nucleic Acids Research*, 31(12), 69e-69.
- Weber, W., Daoud-El Baba, M. & Fussenegger, M. 2007. Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proceedings of the National Academy of Sciences*, 104(25), 10435-40.
- Weber, W., Fux, C., Daoud-El Baba, M., Keller, B., Weber, C. C., Kramer, B. P., Heinzen, C., Auel, D., Bailey, J. E. & Fussenegger, M. 2002. Macrolide-based transgene control in mammalian cells and mice. *Nature Biotechnology*, 20(9), 901-7.
- Weber, W., Lienhart, C., Baba, M. D. & Fussenegger, M. 2009. A biotin-triggered genetic switch in mammalian cells and mice. *Metabolic Engineering*, 11(2), 117-24.
- Weber, W., Link, N. & Fussenegger, M. 2006. A genetic redox sensor for mammalian cells. *Metabolic Engineering*, 8(3), 273-80.
- Weber, W., Schoenmakers, R., Keller, B., Gitzinger, M., Grau, T., Daoud-El Baba, M., Sander, P. & Fussenegger, M. 2008. A synthetic mammalian gene circuit reveals

- antituberculosis compounds. *Proceedings of the National Academy of Sciences*, 105(29), 9994-8.
- Werten, M. W. T., Van Den Bosch, T. J., Wind, R. D., Mooibroek, H. & De Wolf, F. A. 1999. High-yield secretion of recombinant gelatins by *Pichia pastoris*. *Yeast*, 15(11), 1087-1096.
- Wiebe, M. E., Becker, F., Lazar, R., May, L., Casto, B., Semense, M. & Fautz, C. 1989. A multifaceted approach to assure that recombinant tPA is free of adventitious virus. *In: Spier, R. E., Griffiths, J. B., Stephenne, J. & Crooy, P. J. (eds.) Advances in animal cell biology and technology*. London, U.K.: Butterworth-Heinemann.
- Williams, A., Collard, T. & Paraskeva, C. 1999. An acidic environment leads to p53 dependent induction of apoptosis in human adenoma and carcinoma cell lines: implications for clonal selection during colorectal carcinogenesis. *Oncogene*, 18(21), 3199.
- Wong, D. C., Wong, K. T., Nissom, P. M., Heng, C. K. & Yap, M. G. 2006. Targeting early apoptotic genes in batch and fed-batch CHO cell cultures. *Biotechnology and Bioengineering*, 95(3), 350-61.
- Wu, P., Ray, N. G. & Shuler, M. L. 1993. A computer model for intracellular pH regulation in Chinese hamster ovary cells. *Biotechnology progress*, 9(4), 374-84.
- Wu, Y., Wang, M., Feng, H., Peng, Y., Sun, J., Qu, X. & Li, C. 2017. Lactate induces osteoblast differentiation by stabilization of HIF1alpha. *Molecular and Cellular Endocrinology*, 45284-92.
- Wysocka, J. & Herr, W. 2003. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends in Biochemical Sciences*, 28(6), 294-304.
- Xie, L., Nyberg, G., Gu, X., Li, H., Möllborn, F. & Wang, D. I. 1997. Gamma-interferon production and quality in stoichiometric fed-batch cultures of Chinese hamster ovary (CHO) cells under serum-free conditions. *Biotechnology and Bioengineering*, 56(5), 577-582.

- Xie, M., Ye, H., Hamri, G. C. & Fussenegger, M. 2014. Antagonistic control of a dual-input mammalian gene switch by food additives. *Nucleic Acids Research*, 42(14), e116.
- Xin, B., Wu, G., Zhang, K., He, Y., Tang, H., Gao, C., Xu, P. & Ma, C. 2016. Sequence similarity network analysis, crystallization, and X-ray crystallographic analysis of the lactate metabolism regulator LldR from *Pseudomonas aeruginosa*. *Bioresources and Bioprocessing*, 3(1).
- Xu, S., Gavin, J., Jiang, R. & Chen, H. 2017. Bioreactor productivity and media cost comparison for different intensified cell culture processes. *Biotechnology progress*, 33(4), 867-878.
- Xu, S., Hoshan, L. & Chen, H. 2016. Improving lactate metabolism in an intensified CHO culture process: productivity and product quality considerations. *Bioprocess and Biosystems Engineering*, 39(11), 1689-702.
- Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M. R., Neff, N., Passarelli, B., Koh, W., Fan, H. C., Wang, J., Gui, Y., Lee, K. H., Betenbaugh, M. J., Quake, S. R., Famili, I., Palsson, B. O. & Wang, J. 2011. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology*, 29(8), 735-41.
- Xu, Y., Kiningham, K. K., Devalaraja, M. N., Yeh, C. C., Majima, H., Kasarskis, E. J. & St Clair, D. K. 1999. An intronic NF-kappaB element is essential for induction of the human manganese superoxide dismutase gene by tumor necrosis factor-alpha and interleukin-1beta. *DNA and Cell Biology*, 18(9), 709-22.
- Yamaguchi, H. & Miyazaki, M. 2014. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules*, 4(1), 235-51.
- Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. 2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, 389(1), 52-65.

- Yang, M. & Butler, M. 2000. Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation. *Biotechnology and Bioengineering*, 68(4), 370-380.
- Yoshimoto, N. & Kuroda, S. 2014. Single-cell-based breeding: Rational strategy for the establishment of cell lines from a single cell with the most favorable properties. *Journal of Bioscience and Bioengineering*, 117(4), 394-400.
- Young, J. D. 2013. Metabolic flux rewiring in mammalian cell cultures. *Current Opinion in Biotechnology*, 24(6), 1108-15.
- Yuk, I. H., Russell, S., Tang, Y., Hsu, W. T., Mauger, J. B., Aulakh, R. P., Luo, J., Gawlitzek, M. & Joly, J. C. 2015. Effects of copper on CHO cells: cellular requirements and product quality considerations. *Biotechnology progress*, 31(1), 226-38.
- Yun, C. Y., Liu, S., Lim, S. F., Wang, T., Chung, B. Y., Jiat Teo, J., Chuan, K. H., Soon, A. S., Goh, K. S. & Song, Z. 2007. Specific inhibition of caspase-8 and -9 in CHO cells enhances cell viability in batch and fed-batch cultures. *Metabolic Engineering*, 9(5-6), 406-18.
- Zagari, F., Jordan, M., Stettler, M., Broly, H. & Wurm, F. M. 2013. Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity. *Nature Biotechnology*, 30(2), 238-45.
- Zago, P., Baralle, M., Ayala, Y. M., Skoko, N., Zacchigna, S., Buratti, E. & Tisminetzky, S. 2009. Improving human interferon-beta production in mammalian cell lines by insertion of an intronic sequence within its naturally uninterrupted gene. *Biotechnology and Applied Biochemistry*, 52(Pt 3), 191-8.
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E. V. & Zhang, F. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163(3), 759-71.
- Zhao, H. F., Boyd, J., Jolicoeur, N. & Shen, S. H. 2003. A coumermycin/novobiocin-regulated gene expression system. *Human Gene Therapy*, 14(17), 1619-29.

Zheng, J. 2012. Energy metabolism of cancer: Glycolysis versus oxidative phosphorylation. *Oncology letters*, 4(6), 1151-1157.

Zhou, M., Crawford, Y., Ng, D., Tung, J., Pynn, A. F., Meier, A., Yuk, I. H., Vijayasankaran, N., Leach, K., Joly, J., Snedecor, B. & Shen, A. 2011. Decreasing lactate level and increasing antibody production in Chinese hamster ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. *Journal of Biotechnology*, 153(1-2), 27-34.

Zohar, M., Mesika, A. & Reich, Z. 2001. Analysis of genetic control elements in eukaryotes: transcriptional activity or nuclear hitchhiking? *Bioessays*, 23(12), 1176-9.

9 Appendices

9.1 All biopharmaceuticals approved for sale by the Europeans Medicines Agency, from 1990 to July 2018

Table 9.1. Biopharmaceuticals approved for sale by the EMA from 1990 to July 2018. This includes the production host used, the company responsible for marketing, and date of authorization by the EMA.

Medicine Name	Active Substance	Production host	Marketing Authorisation Holder	Authorisation date
Procrit	epoetin	CHO	Amgen	1990
OncoScint	satumomab	CHO	Cytogen	1992
Gonal-F	follitropin alfa	CHO	Serono Pharma	1995
Betaferon	interferon beta-1b	<i>Escherichia coli</i>	Chiron Corporation	1995
NovoSeven	eptacog alfa	BHK	Novo Nordisk	1996
Puregon	follitropin beta	CHO	Organon	1996
Humalog	insullin lispro	<i>Escherichia coli</i>	Lilly France	1996
Rapilysin	reteplase	<i>Escherichia coli</i>	Roche	1996
Avonex	interferon beta-1a	CHO	Biogen	1997
BeneFIX	nonacog alfa	CHO	Baxter	1997
Cerezyme	imiglucerase	CHO	Genzyme	1997
NeoRecormon	epoetin beta	CHO	Roche	1997
Refludan	lerpirudin	<i>Saccharomyces cerevisiae</i>	Schering	1997
Revasc	desirudin	<i>Saccharomyces cerevisiae</i>	Ciba/Aventis	1997

Rituxan	rituximab	CHO	IDEC Pharma	1998
Rebif	interferon beta-1a	CHO	Serono Pharma	1998
HumaSPECT	votumumab	CHO	Organon Teknika	1998
Insuman	insulin	<i>Escherichia coli</i>	Aventis	1998
Simulect	basiliximab	Sp2/0	Novartis	1998
ReFacto	morotcocog alfa	CHO	Wyeth Labs	1999
Beromun	tasonermin	<i>Escherichia coli</i>	Boehringer Ingelheim	1999
Forcaltonin	salmon calcitonin	<i>Escherichia coli</i>	Unigene	1999
Infergen	interferon alfacon-1	<i>Escherichia coli</i>	Amgen	1999
Synagis	palivizumab	NS0	MedImmune	1999
Zenapax	daclizumab	NS0	Hoffman-La Roche	1999
NovoRapid	insulin aspart	<i>Saccharomyces cerevisiae</i>	Novo Nordisk	1999
Regranex	becaplermin	<i>Saccharomyces cerevisiae</i>	Ortho-McNeil	1999
Enbrel	etanercept	CHO	Wyeth Labs and Immunex	2000
Helixate NexGen	octocog alfa	CHO	Bayer	2000
Herceptin	trastuzumab	CHO	Hoffman-La Roche	2000
Luveris	lutropin alfa	CHO	Serono Pharma	2000
Thyrogen	thyrotropin alfa	CHO	Genzyme	2000
Intron A	interferon alfa-2b	<i>Escherichia coli</i>	Schering-Plough	2000

Lantus	insulin glargine	<i>Escherichia coli</i>	Aventis	2000
Optisulin	insulin glargine	<i>Escherichia coli</i>	Aventis	2000
PegIntron	peginterferon alfa-2b	<i>Escherichia coli</i>	Schering-Plough	2000
Viraferon	interferon alfa- 2b	<i>Escherichia coli</i>	Schering-Plough	2000
Remicade	infliximab	Sp2/0	Centocor	2000
Fabrazyme	agalsidase beta	CHO	Genzyme	2001
MabCampath	alemtuzumab	CHO	Schering	2001
TNKase	tenecteplase	CHO	Genentech	2001
Nespo	darbepoetin alfa	CHO	Amgen	2001
Osigraft	eptotermin alfa	CHO	Stryker Biotech	2001
Ovidrel	choriogonadotr opin	CHO	Serono Pharma	2001
NutropinAQ	somatropin	<i>Escherichia coli</i>	Schwarz Pharma	2001
Replagal	agalsidase alfa	HT-1080	TKT Europe	2001
Fasturect	rasburicase	<i>Saccharomyces cerevisiae</i>	Sanofi-Aventis	2001
HBV AXPRO	hepatitis B vaccine	<i>Saccharomyces cerevisiae</i>	Merck Sharp and Dohme	2001
Inductos	dibotermin alfa	CHO	Wyeth Labs	2002
Kineret	anakinra	<i>Escherichia coli</i>	Amgen	2002
Neulasta	pegfilgrastim	<i>Escherichia coli</i>	Amgen	2002
Somavert	pegvisomant	<i>Escherichia coli</i>	Pharmacia	2002
Xigris	drotrecogin alfa	HEK293	Lilly	2002

Dynepo	epoetin delta	HT-1080	Shire Pharmaceutical	2002
Novolin	insulin	<i>Saccharomyces cerevisiae</i>	Novo Nordisk	2002
Aldurazyme	Laronidase	CHO	BioMarin and Genzyme	2003
Humira	adalimumab	CHO	Abbott Laboratories	2003
Forsteo	teriparatide	<i>Escherichia coli</i>	Eli Lilly	2003
Raptiva	efalizumab	CHO	Serono	2004
Zevalin	ibritumomab	CHO	IDEC Pharmaceuticals	2004
Advate	octocog alfa	CHO	Baxter	2004
Apidra	insulin glulisine	<i>Escherichia coli</i>	Aventis	2004
Levemir	insulin detemir	<i>Saccharomyces cerevisiae</i>	Novo Nordisk	2004
Erbix	cetuximab	Sp2/0	Merck	2004
Naglazyme	bevacizumab	CHO	Genentech	2005
Avastin	bevacizumab	CHO	Roche	2005
Xolair	omalizumab	CHO	Genentech	2005
Kepivance	palifermin	<i>Escherichia coli</i>	Amgen	2005
Myozyme	acid-alpha-glucosidase	CHO	Genzyme	2006
Naglazyme	galsulfase	CHO	BioMarin	2006
Omnitrope	human growth hormone	<i>Escherichia coli</i>	Sandoz	2006
Exubera	insulin	<i>Escherichia coli</i>	Pfizer	2006
Preolach	parathyroid hormone	<i>Escherichia coli</i>	Nycomed	2006

Tysabri	natalizumab	NS0	Biogen	2006
Valtropin	human growth hormone	<i>Saccharomyces cerevisiae</i>	BioPartners	2006
Atryn	antithrombin	transgenic goat milk	LEO Pharma	2006
Orencia	abatacept	CHO	Bristol-Myers Squibb	2007
Retacrit	epoetin zeta	CHO	Hospira Enterprises	2007
Silapo	epoetin zeta	CHO	Stada Arzneimittel	2007
Epoetin alfa Hexal	erythropoietin alfa	CHO	Hexal Biotech	2007
Abseamed	erythropoietin alfa	CHO	Medice Arzneimittel Putter	2007
Binocrit	erythropoietin alfa	CHO	Sandoz	2007
Pergoveris	follitropin alfa	CHO	Serono	2007
Vectibix	panitumumab	CHO	Amgen	2007
Mircera	PEGylated epoetin beta	CHO	Roche	2007
Increlex	mecasermin	<i>Escherichia coli</i>	Tercica/Baxter	2007
Lucentis	ranibizumab	<i>Escherichia coli</i>	Genentech	2007
Elaprase	iduronate-2-sulfatase	HT-1080	Shire Pharmaceuticals	2007
Seliris	eculizumab	NS0	Alexion	2007
Filgrastim ratiopharm	filgrastin	<i>Escherichia coli</i>	Ratiopharm	2008
Extavia	interferon beta-1B	<i>Escherichia coli</i>	Novartis	2008
Biopoin	epoetin theta	CHO	CT Arzneimittel	2009

Eporatio	epoetin theta	CHO	Ratiopharm	2009
Opgenra	eptotermin alfa	CHO	Howmedica	2009
Fertavid	follitropin beta	CHO	Schering Plough	2009
Arcalyst	rilonacept	CHO	Regeneron	2009
RoActemra	tocilizumab	CHO	Roche	2009
Cimzia	certolizumab pegol	<i>Escherichia coli</i>	UCB	2009
Filgrastim hexal	filgrastim	<i>Escherichia coli</i>	Hexal Biotech	2009
Zarzio	filgrastim	<i>Escherichia coli</i>	Sandoz	2009
Nplate	romiplostim	<i>Escherichia coli</i>	Amgen	2009
Removab	catumaxomab	rat-mouse hybrid- hybridoma	Fresenius Biotech	2009
Victoza/Saxenda	liraglutide	<i>Saccharomyces cerevisiae</i>	Novo Nordisk (approved 2009 and 2015)	2009
Ilaris	canakinumab	Sp2/0	Novartis	2009
Simponi	golimumab	Sp2/0	Centocor	2009
Stelara	ustekinumab	Sp2/0	Centocor	2009
Elonva	corifollitropin alfa	CHO	Organon	2010
Prolia/Xgeva	denosumab	CHO	Amgen (approved 2010 and 2011)	2010
Nivestim	filgrastim	<i>Escherichia coli</i>	Hospira Enterprises	2010
VPRIV	velaglucerase alfa	HT-1080	Shire Pharmaceuticals	2010
Arzerra	ofatumumab	NS0	GlaxoSmithKline	2010

Scintimun	besilesomab	PTA6291	CIS Bio International	2010
Ruconest	conestat alfa	transgenic rabbit milk	Pharming	2010
Nulojix	belatacept	CHO	Bristol-Myers Squibb	2011
Yervoy	ipilimumab	CHO	Bristol-Myers Squibb	2011
Benlysta	bellimumab	NS0	GlaxoSmithKline	2011
Eylea	aflibercept	CHO	Bayer	2012
Adcetris	brentuximab vedotin	CHO	Takeda	2012
Lonquex	lipegfilgrastim	<i>Escherichia coli</i>	Teva Pharmaceuticals	2012
Revestive	teduglutide	<i>Escherichia coli</i>	Nycomed	2012
NovoThirteen	catridecog	<i>Saccharomyces cerevisiae</i>	Novo Nordisk	2012
Zaltrap	aflibercept	CHO	Regeneron/Sanofi-Aventis	2013
Ovaleap	follitropin alfa	CHO	Teva Pharmaceuticals	2013
Perjeta	pertuzumab	CHO	Roche	2013
Kadcyla	trastuzumab	CHO	Roche	2013
NovoEight	turoctotog alfa	CHO	Novo Nordisk	2013
Grastofil	filgrastim	<i>Escherichia coli</i>	Apotex	2013
Krystexxa	pegloticase	<i>Escherichia coli</i>	Savient	2013
Jetrea	ocriplasmin	<i>Pichia pastoris</i>	ThromboGenics	2013
Tresiba	insulin degludec	<i>Saccharomyces cerevisiae</i>	Novo Nordisk	2013

Ryzodeg	insulin degludec /insulin aspart	<i>Saccharomyces cerevisiae</i>	Novo Nordisk	2013
Somatropin BioPartners	somatropin	<i>Saccharomyces cerevisiae</i>	BioPartners	2013
Provenge	sipuleucel-T	Sf21	Dendreon	2013
Inflectra/Remsi ma	infliximab	Sp2/0	Hospira Enterprises/Celltrion	2013
Gazyvaro	obinutuzumab	CHO	Genentech/Roche	2014
Plegridy	peginterferon beta 1a	CHO	Biogen	2014
Sylvant	siltuximab	CHO	Janssen Biotech	2014
Entyvio	vedolizumab	CHO	Taastrup	2014
Trulicity	dulaglutide	CHO	Eli Lilly	2014
Rixubis	nonacog gamma	CHO	Baxalta Innovations	2014
Abasaglar (previously Abasria)	insulin glargine	<i>Escherichia coli</i>	Eli Lilly	2014
Accofil	filgrastim	<i>Escherichia coli</i>	Accord Healthcare	2014
Nuwiq/Vihuma	simoctocog alfa	HEK293	Octapharma (approved 2014 and 2017)	2014
Cyramza	ramucirumab	NS0	Eli Lilly	2014
Xultophy	insulin degludec / liraglutide	<i>Saccharomyces cerevisiae</i>	Novo Nordisk	2014
Obizur	susoctocog alfa	BHK	Baxalta Innovations	2015

Cosentyx	secukinumab	CHO	Novartis	2015
Opdivo/Nivolumab BMS	nivolumab	CHO	Bristol-Myers Squibb	2015
Repatha	evolocumab	CHO	Amgen	2015
Keytruda	pembrolizumab	CHO	Merck	2015
Unituxin	dinutuximab	CHO	United Therapeutics	2015
Strensiq	asfotase alfa	CHO	Alexion	2015
Praluent	alirocumab	CHO	Sanofi-Aventis Groupe	2015
Praxbind	idarucizumab	CHO	Boehringer Ingelheim	2015
Blinicyto	blinatumomab	CHO	Amgen	2015
Nucala	mepolizumab	CHO	GlaxoSmithKline	2015
Ristempa	pegfilgrastim	<i>Escherichia coli</i>	Amgen	2015
Kanuma	sebelipase alfa	<i>Gallus gallus</i>	Alexion	2015
Elocta	efmoroctocog alfa	HEK293	Swedish Orphan Biovitrum	2015
Iblias/Kovaltry	octocog alfa	BHK	Bayer	2016
Benepali	etanercept	CHO	Samsung Bioepis	2016
Taltz	ixekizumab	CHO	Eli Lilly	2016
Idelvion	albutrepenonacog alfa	CHO	CSL Behring	2016
Darzalex	daratumumab	CHO	Janssen-Cilag International	2016
Flixabi	infliximab	CHO	Samsung Bioepis	2016

Spectrila	asparaginase	<i>Escherichia coli</i>	Medac Gesellschaft fuer klinische Spezialpraeparate mbH	2016
Oncaspar	pegaspargase	<i>Escherichia coli</i>	Baxalta Innovations	2016
Alprolix	eftrenonacog alfa	HEK293	Swedish Orphan Biovitrum	2016
Portrazza	necitumumab	NS0	Eli Lilly	2016
Empliciti	elotuzumab	NS0	Bristol-Myers Squibb	2016
Zinbryta	daclizumab	NS0	Biogen	2016
Cinquaero	reslizumab	NS0	Teva Pharmaceuticals	2016
Lartruvo	olaratumab	NS0	Eli Lilly	2016
Rekovelte	follitropin delta	PER.C6	Ferring Pharmaceuticals	2016
Afstyla	lonoctocog alfa	CHO	CSL Behring	2017
Lifmior	etanercept	CHO	Pfizer	2017
Truxima	rituximab	CHO	Celltrion	2017
Amgevita/Soly mbic	adalimumab	CHO	Amgen	2017
Qarziba (previously Dinutuximab beta EUSA and Dinutuximab beta Apeiron)	dinutuximab beta	CHO	EUSA Pharma	2017
Brineura	cerliponase alfa	CHO	BioMarin	2017

Refixia	nonacog beta pegol	CHO	Novo Nordisk	2017
Rixathon/Rixi myo	rituximab	CHO	Sandoz	2017
Erelzi	etanercept	CHO	Sandoz	2017
Kevzara	sarilumab	CHO	Sanofi-Aventis Groupe	2017
Besponsa	inotuzumab ozogamicin	CHO	Pfizer	2017
Blitzima/Ritem via/Rituzena	rituximab	CHO	Celltrion	2017
Kyntheum	brodalumab	CHO	LEO Pharma	2017
Imraldi	adalimumab	CHO	Samsung Bioepis	2017
Bavencio	avelumab	CHO	Merck	2017
Tecentriq	atezolizumab	CHO	Roche	2017
Dupixent	dupilumab	CHO	Sanofi-Aventis Groupe	2017
Cyltezo	adalimumab	CHO	Boehringer Ingelheim	2017
Tremfya	guselkumab	CHO	Janssen-Cilag International	2017
Ontruzant	trastuzumab	CHO	Samsung Bioepis	2017
Lusduna	insulin glargine	<i>Escherichia coli</i>	Merck	2017
Suliqua	insulin glargine / lixisenatide	<i>Escherichia coli</i>	Sanofi-Aventis Groupe	2017
Natpar	parathyroid hormone	<i>Escherichia coli</i>	Shire Pharmaceuticals	2017
Oxervate	nerve growth factor	<i>Escherichia coli</i>	Dompe Farmaceutici	2017

Insulin lispro Sanofi	insulin lispro	<i>Escherichia coli</i>	Sanofi-Aventis Groupe	2017
Fiasp	insulin aspart	<i>Saccharomyces cerevisiae</i>	Novo Nordisk	2017
Ocrevus	ocrelizumab	CHO	Roche	2018
Adynovi	rurioctocog alfa pegol	CHO	Baxalta Innovations	2018
Herzuma	trastuzumab	CHO	Celltrion	2018
Crysvita	burosumab	CHO	Kyowa Kirin	2018
Hemlibra	emicizumab	CHO	Roche	2018
Lamzede	velmanase alfa	CHO	Chiesi Farmaceutici	2018
Mylotarg	gemtuzumab ozogamicin	CHO	Pfizer	2018
Kanjinti	trastuzumab	CHO	Amgen, Breda	2018
Zessly	infliximab	CHO	Sandoz	2018
Semglee	insulin glargine	<i>Pichia pastoris</i>	Mylan	2018

9.2 Bacterial strains used

Table 9.2. Bacterial strains used for cloning and protein expression. The Δ denotes a gene knockout. Allele designators are given after gene/loci names, and can be numbers or superscript letters (e.g. *lacI^q*). Plasmids are preceded by “p”.

Strain	Description	Origin
E. coli Turbo	F' <i>proA⁺B⁺ lacI^q ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)Tet^S endA1 thi-1 Δ(hsdS-mcrB)5</i>	NEB
E. coli Rosetta2	F- <i>ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (CamR)</i>	Novagen

Δ (hsdS-mcrB)5 – hsdS and mcrB are responsible for recognition and degradation of unmethylated DNA.

Δ (lac-proAB) – Deletion from the lac operon of genes required for proline synthesis.

Δ lacZM15 / fhuA2 – Abolition of β -galactosidase activity and deletion of the fhuA2 iron uptake receptor confers resistance to phage T1.

dcm (DE3) – Methylation at CCWGG sequences is abolished.

endA1 – Non-specific endonuclease, the deletion of which helps to maintain plasmid integrity.

F' – Carries an F plasmid, which expresses the lac operon and proline biosynthetic genes.

F- – Does not carry the F plasmid.

gal – Unable to metabolise galactose. Associated with higher competence

galE15 – Allele of galE with mutation to active site, producing inability to metabolise galactose. Associated with higher competence.

galK16 – Allele of galK unable to metabolise galactose. Associated with higher competence.

glnV – Required for growth of some phage vectors.

hsdSB(rB- mB-) – Unable to restrict or methylate certain DNA sequences.

lacI^q – Overproduction of the lac repressor, reducing expression from lac promoters.

ompT – Mutation in outer membrane protease, reducing proteolysis of expressed proteins.

pRARE2 (CamR) – Plasmid expressing rarely used tRNAs.

proA⁺B⁺ – Requires the provision of proline.

R(zgb-210::Tn10)Tet^S – Insertion of a transposon into the zgb-210 gene.

thi-1 – Unable to synthesise thiamine.

9.3 Constructs used in chapter 3; experimental results from chapter 3

Table 9.3. Constructs used in chapter 3. This includes lactate detection constructs (i.e. those expressing LldR), lactate response constructs (i.e. those designed to interact with LldR and give a detectable output), constructs to test nuclear localisation of LldR in CHO cells, and positive control plasmids. *C. glut.* refers to *Corynebacterium glutamicum*.

Plasmid name	Encodes LldR or responsive element (RE)?	Promoter	LldR origin	Operator number	Operator location relative to promoter	Polyadenylation signal	Plasmid origin
MXS-CMV-L-P2A-mCherry-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study
	LldR is unfused to an effector domain.						
MXS-CMV-LN-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study
	LldR has a C-terminal fusion of the SV40 nuclear localisation signal.						
MXS-CMV-NL-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study
	LldR has an N-terminal fusion of the SV40 nuclear localisation signal.						
MXS-CMV-LK-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study

	LldR has a C-terminal fusion of the KRAB repressor domain.						
MXS-CMV-KL-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study
	LldR has an N-terminal fusion of the KRAB repressor domain.						
MXS-CMV-NKL-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study
	LldR has an N-terminal fusion of the SV40 nuclear localisation signal and the KRAB repressor domain.						
MXS-CMV-LV-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study
	LldR has a C-terminal fusion of the VP64 transactivation domain.						
MXS-CMV-VL-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study
	LldR has an N-terminal fusion of the VP64 transactivation domain.						
MXS-CMV-LVN-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study
	LldR has a C-terminal fusion of the VP64 transactivation domain and the SV40 nuclear localisation signal.						
MXS-CMV-VNL-P2A-mC-SV40pA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	SV40	This study

	LldR has an N-terminal fusion of the VP64 transactivation domain and the SV40 nuclear localisation signal.						
MXS-CMV-LldR-mAzamiGreen-bGpA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	bGpA	This study
	This plasmid was used to investigate nuclear localisation of LldR in CHO cells.						
MXS-CMV-LldR-NLS-mAzamiGreen-bGpA	LldR	CMV	<i>C. glut.</i>	N/a	N/a	bGpA	This study
	This plasmid was used to investigate nuclear localisation of LldR in CHO cells.						
MXS-CMV-mAzamiGreen-bGpA	RE	CMV	N/a	N/a	N/a	bGpA	This study
	This plasmid was used as a positive control in transient transfection experiments.						
MXS-CMV-LOx2-mAzamiGreen-bGpA	RE	CMV	N/a	2	Downstream	bGpA	This study
MXS-LOx2-CMV-mAzamiGreen-bGpA	RE	CMV	N/a	2	Upstream	bGpA	This study
MXS-LOx6-CMV-mAzamiGreen-bGpA	RE	CMV	N/a	6	Upstream	bGpA	This study
MXS-LOx12-CMV-mAzamiGreen-bGpA	RE	CMV	N/a	12	Upstream	bGpA	This study

MXS-LOx2-CMV-LOx2-mAzamiGreen-bGpA	RE	CMV	N/a	2 and 2	Up- and downstream	bGpA	This study
MXS-LOx2-minCMV-mAzamiGreen-bGpA	RE	minCMV	N/a	2	Upstream	bGpA	This study
MXS-LOx6-minCMV-mAzamiGreen-bGpA	RE	minCMV	N/a	6	Upstream	bGpA	This study
MXS-LOx12-minCMV-mAzamiGreen-bGpA	RE	minCMV	N/a	12	Upstream	bGpA	This study

Table 9.4. Results from the transient transfections of the initial designs of the lactate-inducible system. The p-value is calculated through the use of a Student's T-test (one-tailed, two-sampled, equal variance) to determine whether the uninduced and induced populations were significantly different. *C. g.* refers to *Corynebacterium glutamicum*.

Exp. #	Construct name(s)	Average geometric mean of GFP (AU)		Standard deviation		FC induction (+lactate /-lactate)	p-value	Configuration type?
		-lactate	+lactate	-lactate	+lactate			
1	L (<i>C. g.</i>) + CMV-2x	5279	5057	733	568	0.96	0.37	Derepression by steric hindrance
2	L (<i>C. g.</i>) + 2x-CMV-2x	7600	7259	145	350	0.96	0.15	Derepression by steric hindrance
3	LN (<i>C. g.</i>) + CMV-2x	1483	1461	30	113	1.02	0.41	Derepression by steric hindrance
4	LN (<i>C. g.</i>) + 2x-CMV-2x	5754	5632	107	126	0.98	0.09	Derepression by steric hindrance
5	NL (<i>C. g.</i>) + CMV-2x	8116	7751	362	1468	0.96	0.37	Derepression by steric hindrance
6	NL (<i>C. g.</i>) + 2x-CMV-2x	6689	7091	1364	1524	1.06	0.24	Derepression by steric hindrance

7	LK (<i>C. g.</i>) + CMV-2x	5988	6199	745	456	1.04	0.39	Derepression by heterochromatin recruitment
8	LK (<i>C. g.</i>) + 2x-CMV	110805	138157	10724	46324	1.25	0.17	Derepression by heterochromatin recruitment
9	LK (<i>C. g.</i>) + 6x-CMV	50608	56429	15026	5039	1.12	0.30	Derepression by heterochromatin recruitment
10	LK (<i>C. g.</i>) + 12x-CMV	48024	50858	3149	500	1.06	0.16	Derepression by heterochromatin recruitment
11	LK (<i>C. g.</i>) + 2x-CMV-2x	5435	5782	737	619	1.06	0.32	Derepression by heterochromatin recruitment
12	KL (<i>C. g.</i>) + CMV-2x	7469	5256	864	2131	0.70	0.16	Derepression by heterochromatin recruitment

13	KL (<i>C. g.</i>) + 2x-CMV	16051	22801	271	6771	1.42	0.11	Derepression by heterochromatin recruitment
14	KL (<i>C. g.</i>) + 6x-CMV	17794	17354	2659	845	0.98	0.39	Derepression by heterochromatin recruitment
15	KL (<i>C. g.</i>) + 12x-CMV	22252	17001	4214	590	0.76	0.10	Derepression by heterochromatin recruitment
16	KL (<i>C. g.</i>) + 2x-CMV-2x	4941	4796	297	278	0.97	0.33	Derepression by heterochromatin recruitment
17	LKN (<i>C. g.</i>) + CMV-2x	6160	5768	173	167	0.94	0.08	Derepression by heterochromatin recruitment
18	LKN (<i>C. g.</i>) + 2x-CMV	14424	11898	1109	526	0.82	0.06	Derepression by heterochromatin recruitment

19	LKN (<i>C. g.</i>) + 6x-CMV	13164	12868	518	522	0.98	0.31	Derepression by heterochromatin recruitment
20	LKN (<i>C. g.</i>) + 12x-CMV	223393	214609	13306	12497	1.05	0.27	Derepression by heterochromatin recruitment
21	LKN (<i>C. g.</i>) + 2x-CMV-2x	5497	5161	171	491	0.94	0.12	Derepression by heterochromatin recruitment
22	NKL (<i>C. g.</i>) + CMV-2x	8595	6936	547	856	0.81	0.08	Derepression by heterochromatin recruitment
23	NKL (<i>C. g.</i>) + 2x-CMV	13240	12108	1268	958	0.91	0.12	Derepression by heterochromatin recruitment
24	NKL (<i>C. g.</i>) + 6x-CMV	13831	11288	732	2481	0.82	0.07	Derepression by heterochromatin recruitment

25	NKL (<i>C. g.</i>) + 12x-CMV	190548	190146	2238	3513	1.00	0.45	Derepression by heterochromatin recruitment
26	NKL (<i>C. g.</i>) + 2x-CMV-2x	8288	6701	299	796	0.81	0.06	Derepression by heterochromatin recruitment
27	LV (<i>C. g.</i>) + 2x-minCMV	9900	10017	826	958	1.01	0.46	Transactivation
28	LV (<i>C. g.</i>) + 6x-minCMV	8502	7326	537	543	0.86	0.08	Transactivation
29	LV (<i>C. g.</i>) + 12x-minCMV	8967	7892	2512	1070	0.88	0.33	Transactivation
30	VL (<i>C. g.</i>) + 2x-minCMV	6828	7283	445	434	1.07	0.18	Transactivation
31	VL (<i>C. g.</i>) + 6x-minCMV	6859	7034	1327	341	1.03	0.44	Transactivation
32	VL (<i>C. g.</i>) + 12x-minCMV	9497	10426	1458	672	1.10	0.26	Transactivation

33	LVN (<i>C. g.</i>) + 2x-minCMV	8968	5068	2611	261	0.57	0.07	Transactivation
34	LVN (<i>C. g.</i>) + 6x-minCMV	12130	12452	1026	937	1.03	0.31	Transactivation
35	LVN (<i>C. g.</i>) + 12x-minCMV	488	535	37	95	0.91	0.27	Transactivation
36	VNL (<i>C. g.</i>) + 2x-minCMV	12605	10676	1711	926	0.85	0.11	Transactivation
37	VNL (<i>C. g.</i>) + 6x-minCMV	10708	10980	732	73	1.03	0.30	Transactivation
38	VNL (<i>C. g.</i>) + 12x-minCMV	9497	10426	1458	672	1.10	0.26	Transactivation

9.4 Solubility testing of LldR and variants by SDS-PAGE

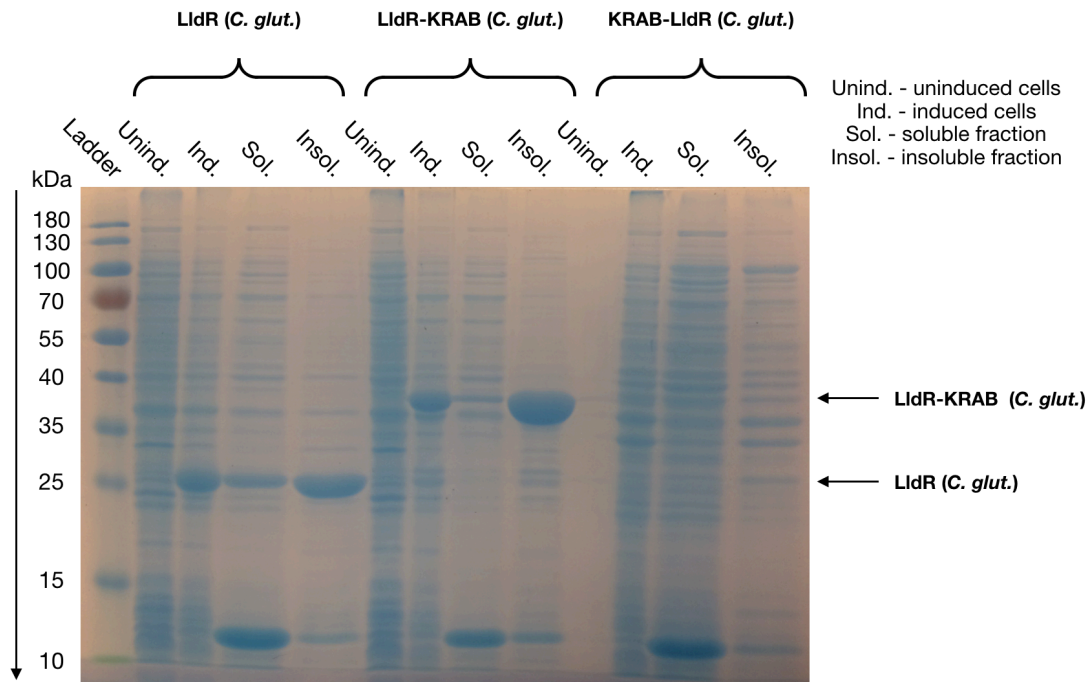


Figure 9.1. Initial solubility testing of LldR, LldR-KRAB and KRAB-LldR (all from *Corynebacterium glutamicum*) with SDS-PAGE.

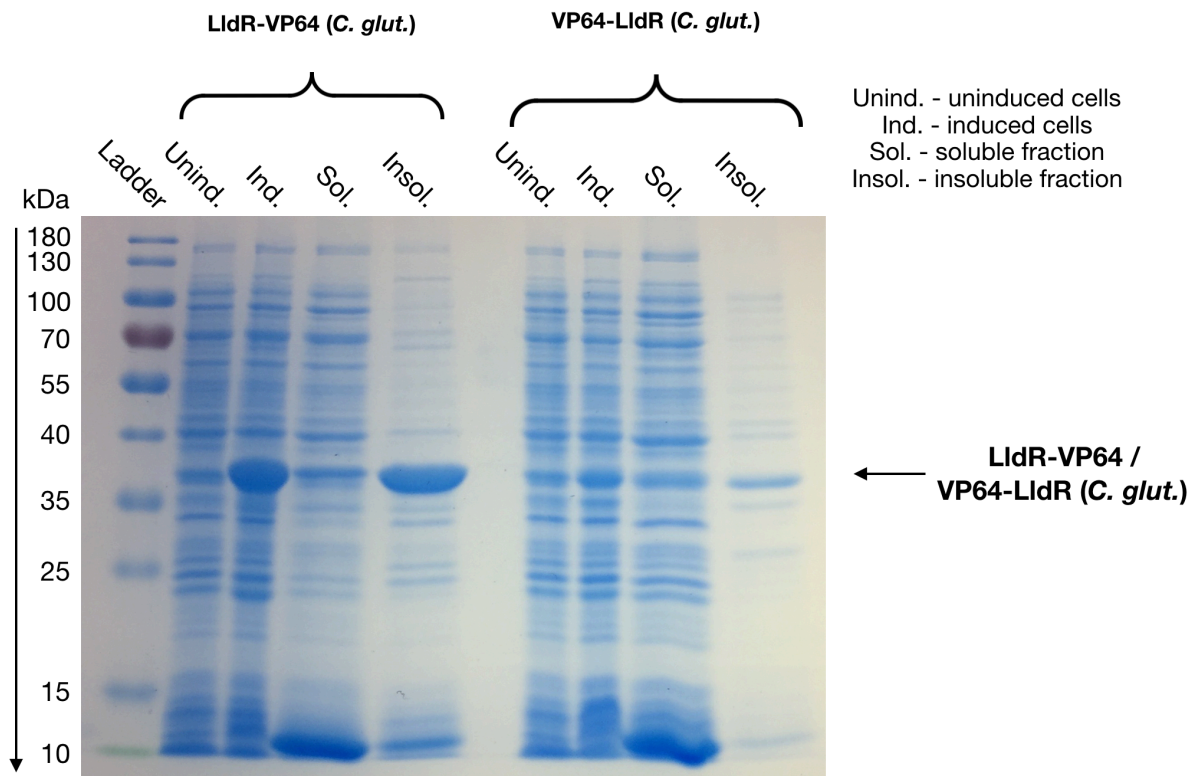


Figure 9.2. Solubility testing of LldR-VP64 and VP64-LldR (all from *Corynebacterium glutamicum*) with SDS-PAGE.

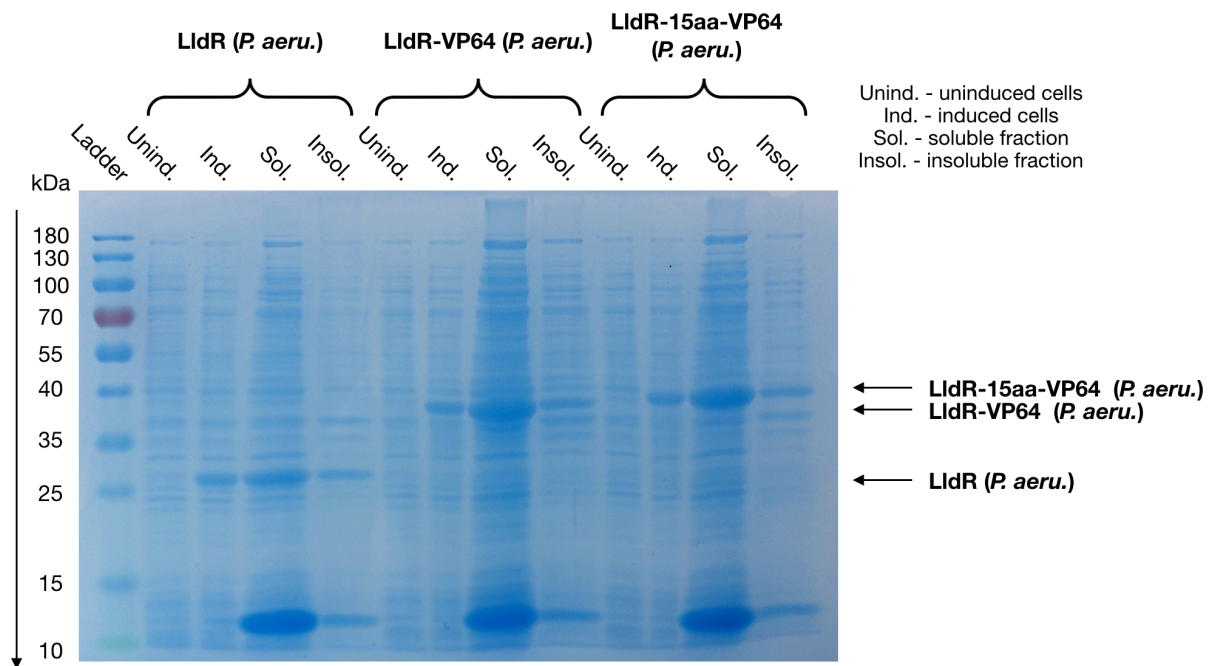


Figure 9.3. Solubility testing of LldR, LldR-VP64 and LldR-15aa-VP64 (all from *Pseudomonas aeruginosa*) with SDS-PAGE.

9.5 Chromatograms showing purification profiles of LldR and variants

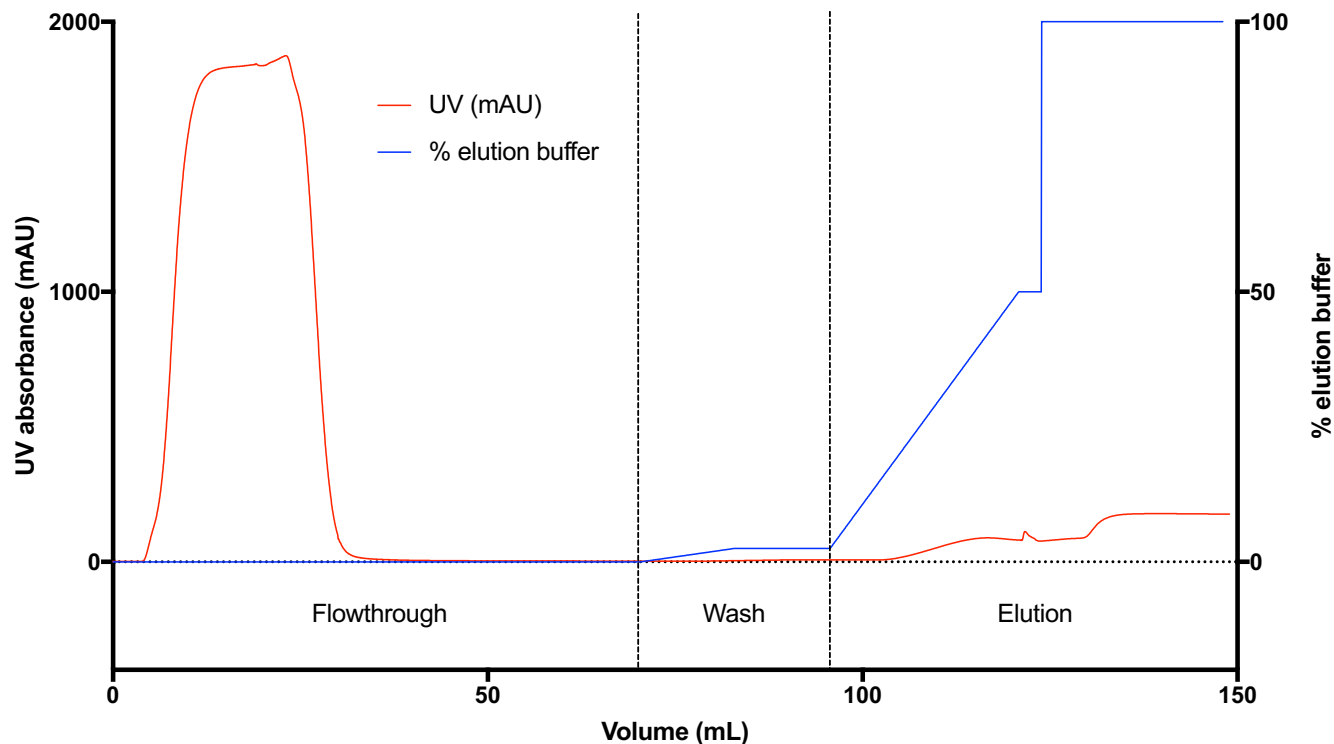


Figure 9.4. Chromatogram produced during the purification of LldR-KRAB (*Corynebacterium glutamicum*). The x-axis shows the volume of buffer progressively passed through the nickel-NTA column. The left y-axis shows the ultraviolet (UV) absorbance, given in milli-arbitrary units (mUA), which is indicative of protein content. The right y-axis shows the percentage of the elution buffer (containing 500 mM imidazole) present in the buffer being passed through the nickel-NTA column. The flowthrough, wash and elution stages of the purification run are demarcated in the figure by the vertical dotted lines.

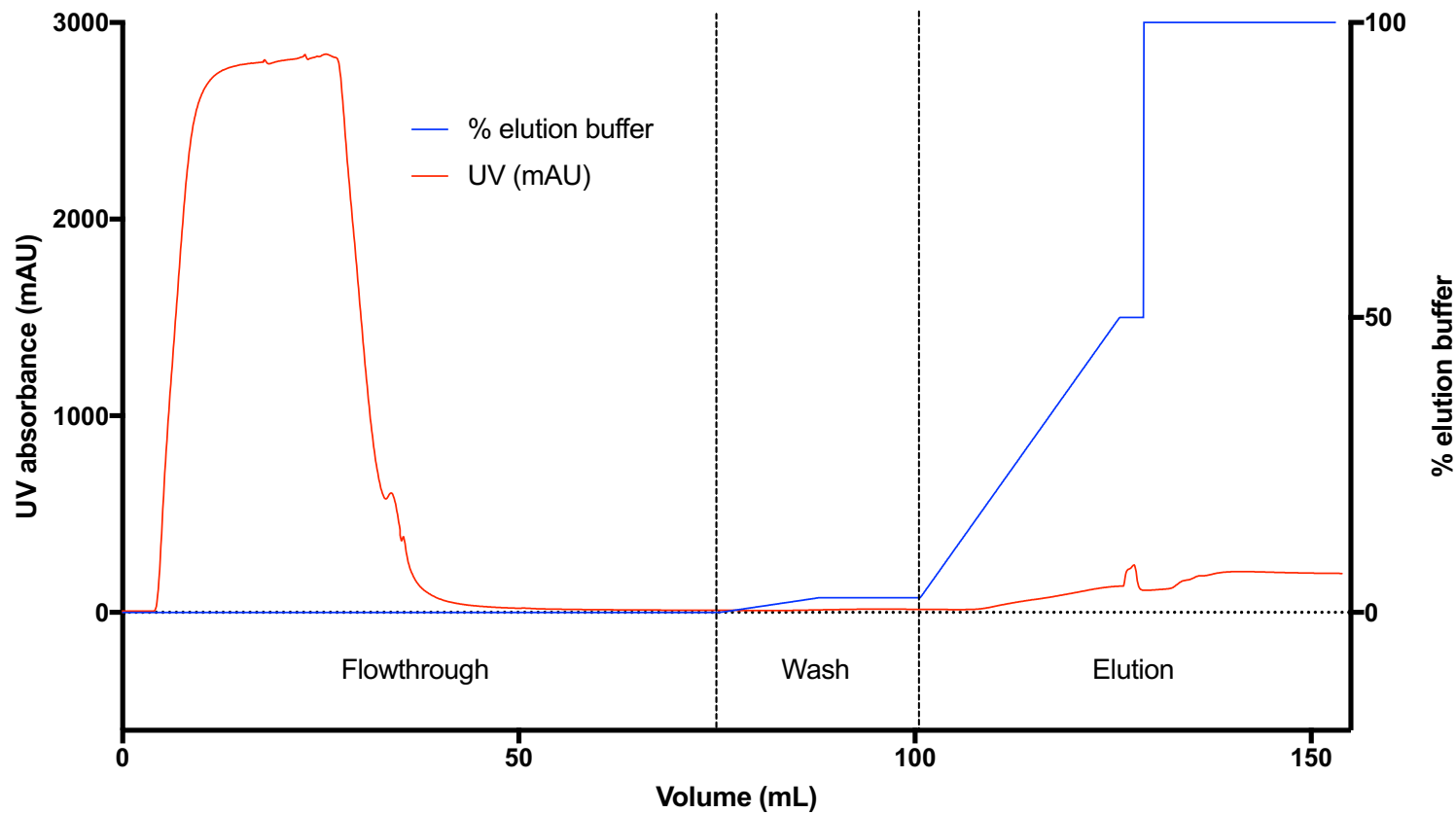


Figure 9.5. Chromatogram produced during the purification of LidR-VP64 (*Corynebacterium glutamicum*). The x-axis shows the volume of buffer progressively passed through the nickel-NTA column. The left y-axis shows the ultraviolet (UV) absorbance, given in milli-arbitrary units (mAU), which is indicative of protein content. The right y-axis shows the percentage of the elution buffer (containing 500 mM imidazole) present in the buffer being passed through the nickel-NTA column. The flowthrough, wash and elution stages of the purification run are demarcated in the figure by the vertical dotted lines.

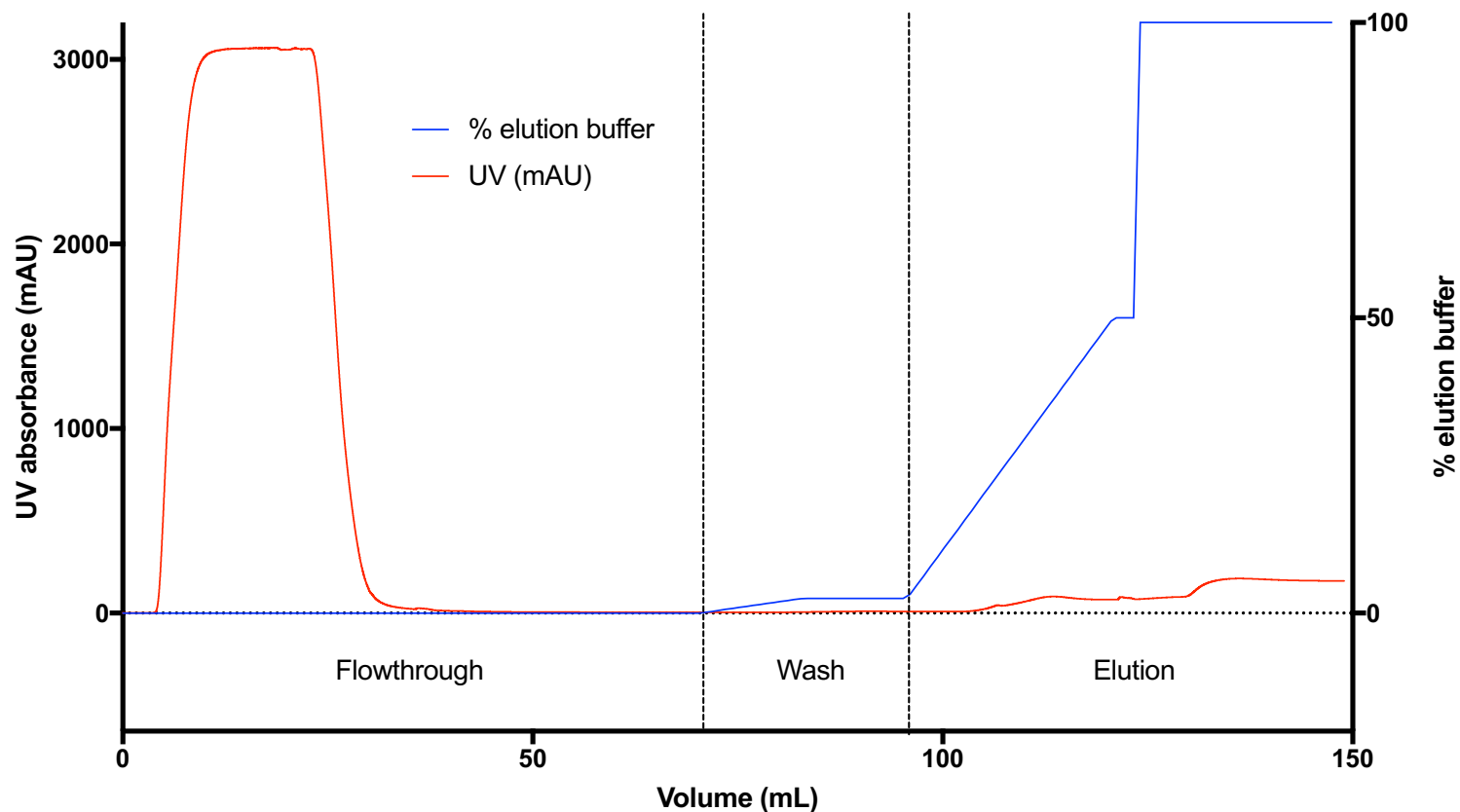


Figure 9.6. Chromatogram produced during the purification of VP64-LidR (*Corynebacterium glutamicum*). The x-axis shows the volume of buffer progressively passed through the nickel-NTA column. The left y-axis shows the ultraviolet (UV) absorbance, given in milli-arbitrary units (mAU), which is indicative of protein content. The right y-axis shows the percentage of the elution buffer (containing 500 mM imidazole) present in the buffer being passed through the nickel-NTA column. The flowthrough, wash and elution stages of the purification run are demarcated in the figure by the vertical dotted lines.

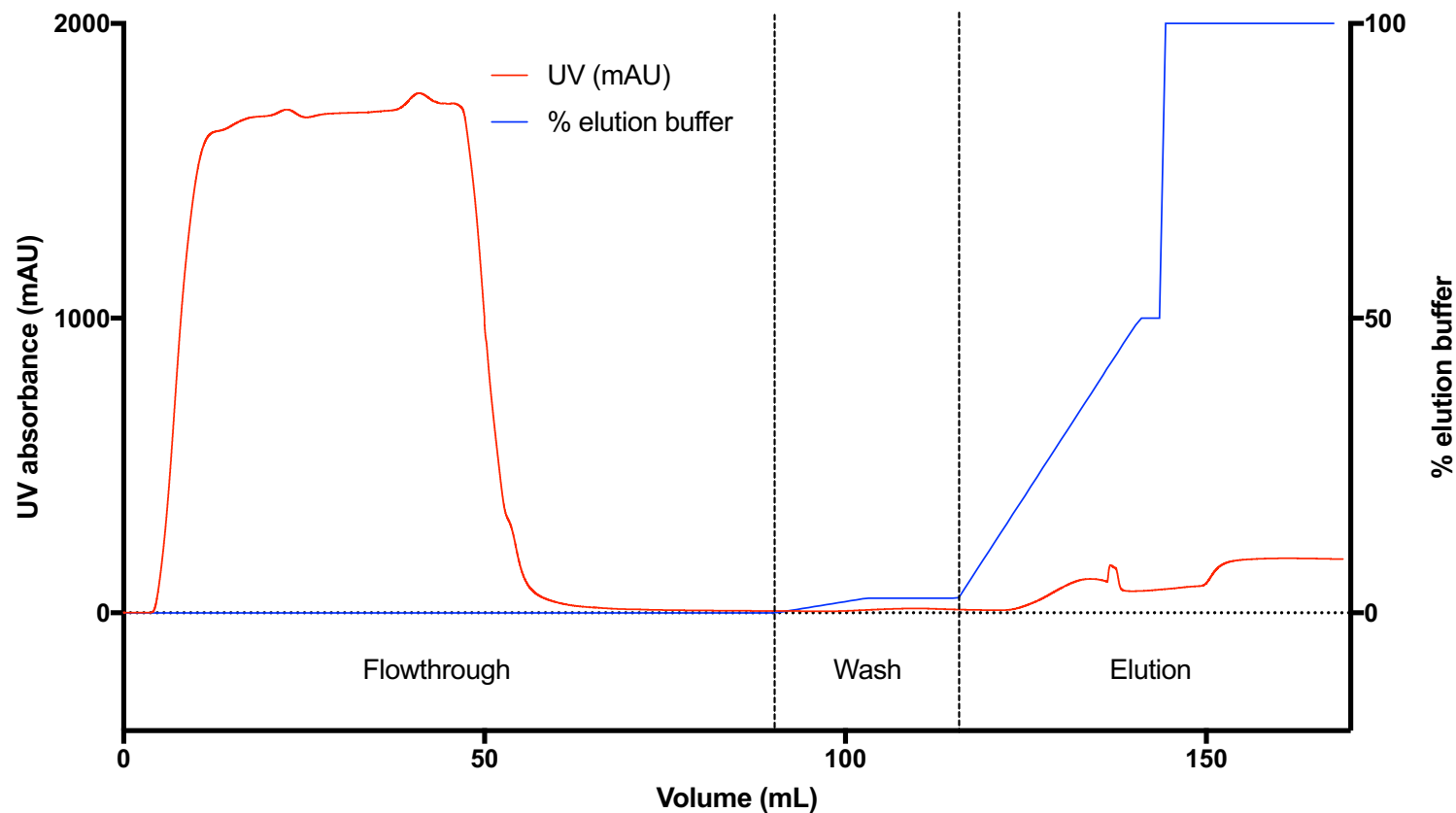


Figure 9.7. Chromatogram produced during the purification of LldR-15aa-VP64 (*Corynebacterium glutamicum*). The x-axis shows the volume of buffer progressively passed through the nickel-NTA column. The left y-axis shows the ultraviolet (UV) absorbance, given in milli-arbitrary units (mUA), which is indicative of protein content. The right y-axis shows the percentage of the elution buffer (containing 500 mM imidazole) present in the buffer being passed through the nickel-NTA column. The flowthrough, wash and elution stages of the purification run are demarcated in the figure by the vertical dotted lines.

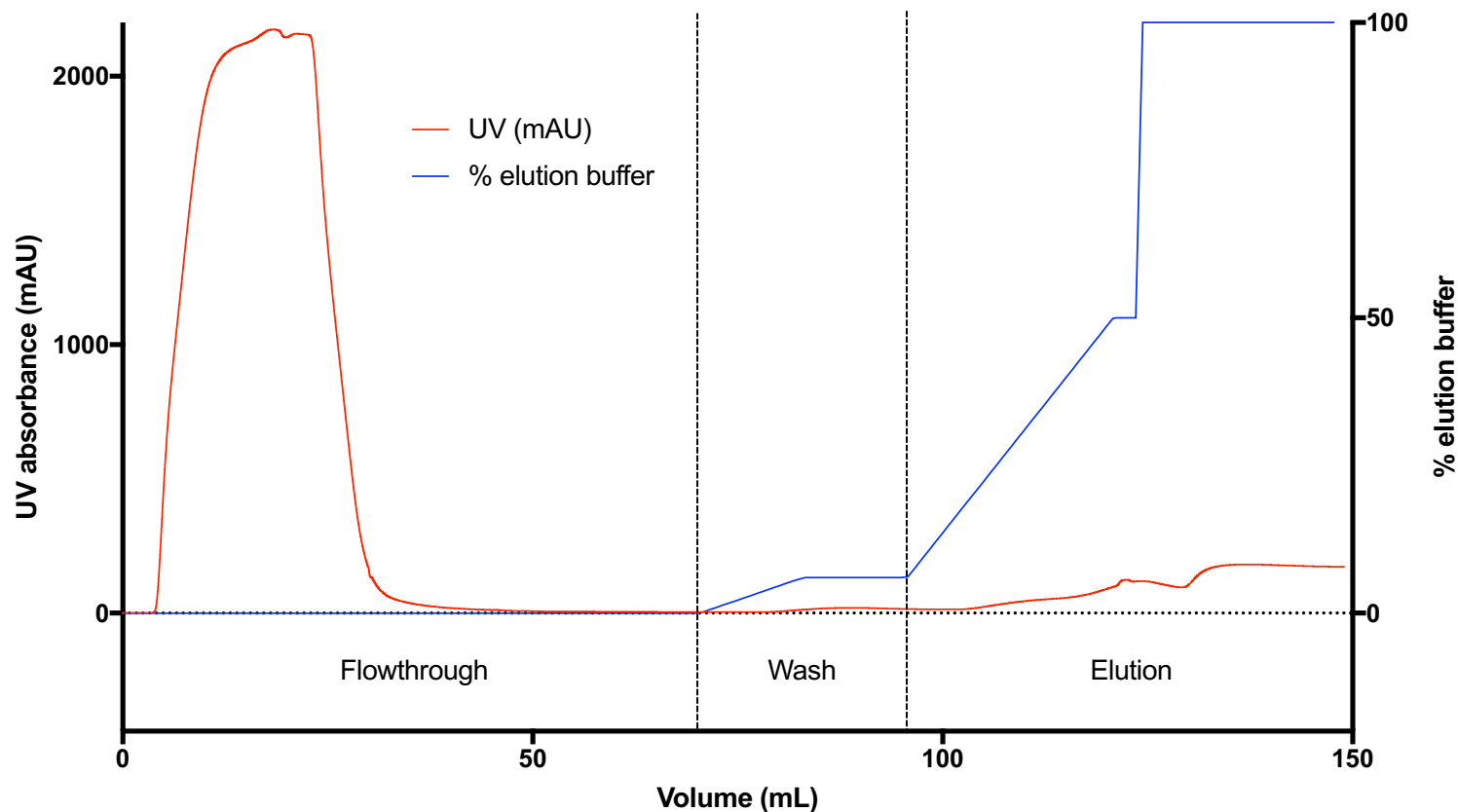


Figure 9.8. Chromatogram produced during the purification of LldR (*Pseudomonas aeruginosa*). The x-axis shows the volume of buffer progressively passed through the nickel-NTA column. The left y-axis shows the ultraviolet (UV) absorbance, given in milli-arbitrary units (mAU), which is indicative of protein content. The right y-axis shows the percentage of the elution buffer (containing 500 mM imidazole) present in the buffer being passed through the nickel-NTA column. The flowthrough, wash and elution stages of the purification run are demarcated in the figure by the vertical dotted lines.

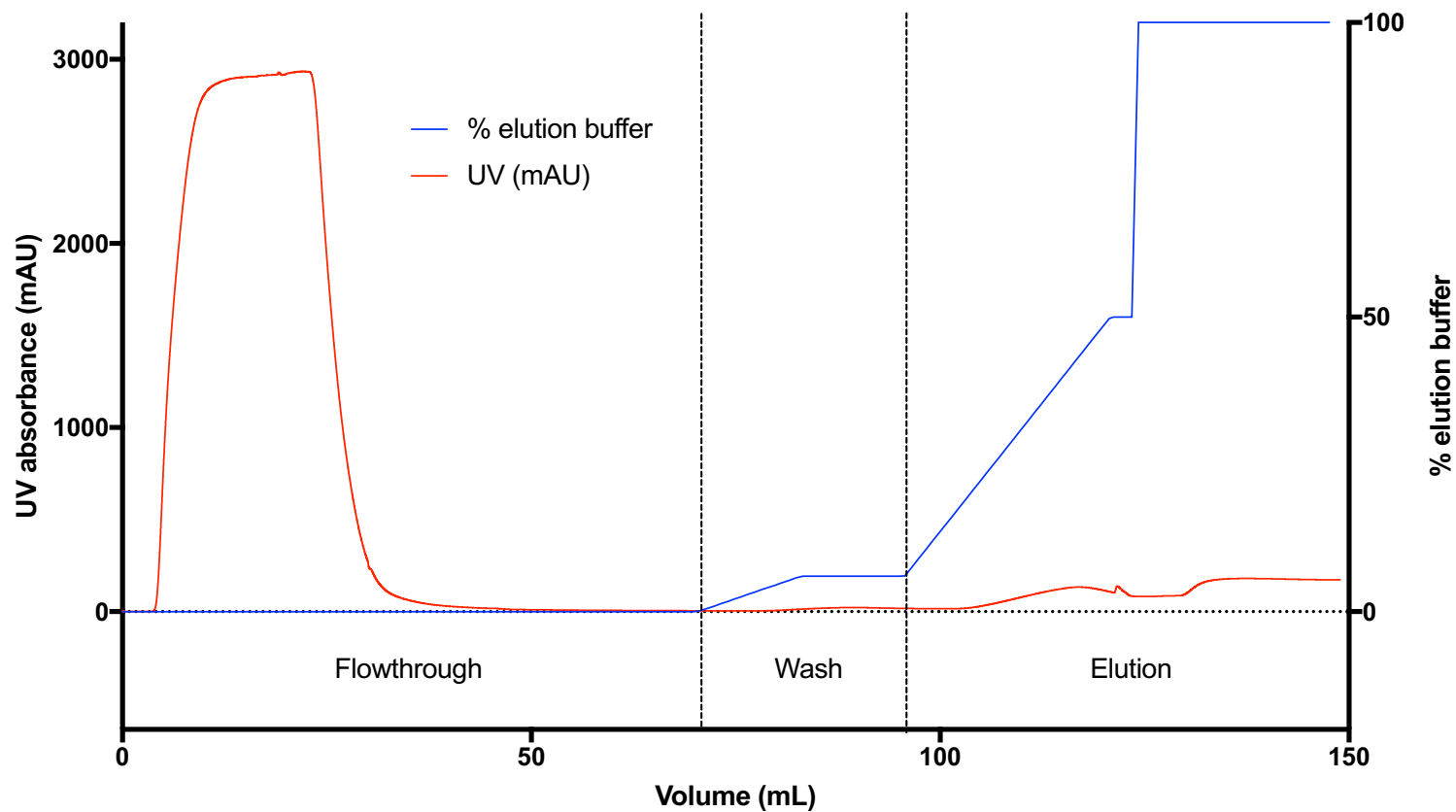


Figure 9.9. Chromatogram produced during the purification of LldR-VP64 (*Pseudomonas aeruginosa*). The x-axis shows the volume of buffer progressively passed through the nickel-NTA column. The left y-axis shows the ultraviolet (UV) absorbance, given in milli-arbitrary units (mAU), which is indicative of protein content. The right y-axis shows the percentage of the elution buffer (containing 500 mM imidazole) present in the buffer being passed through the nickel-NTA column. The flowthrough, wash and elution stages of the purification run are demarcated in the figure by the vertical dotted lines.

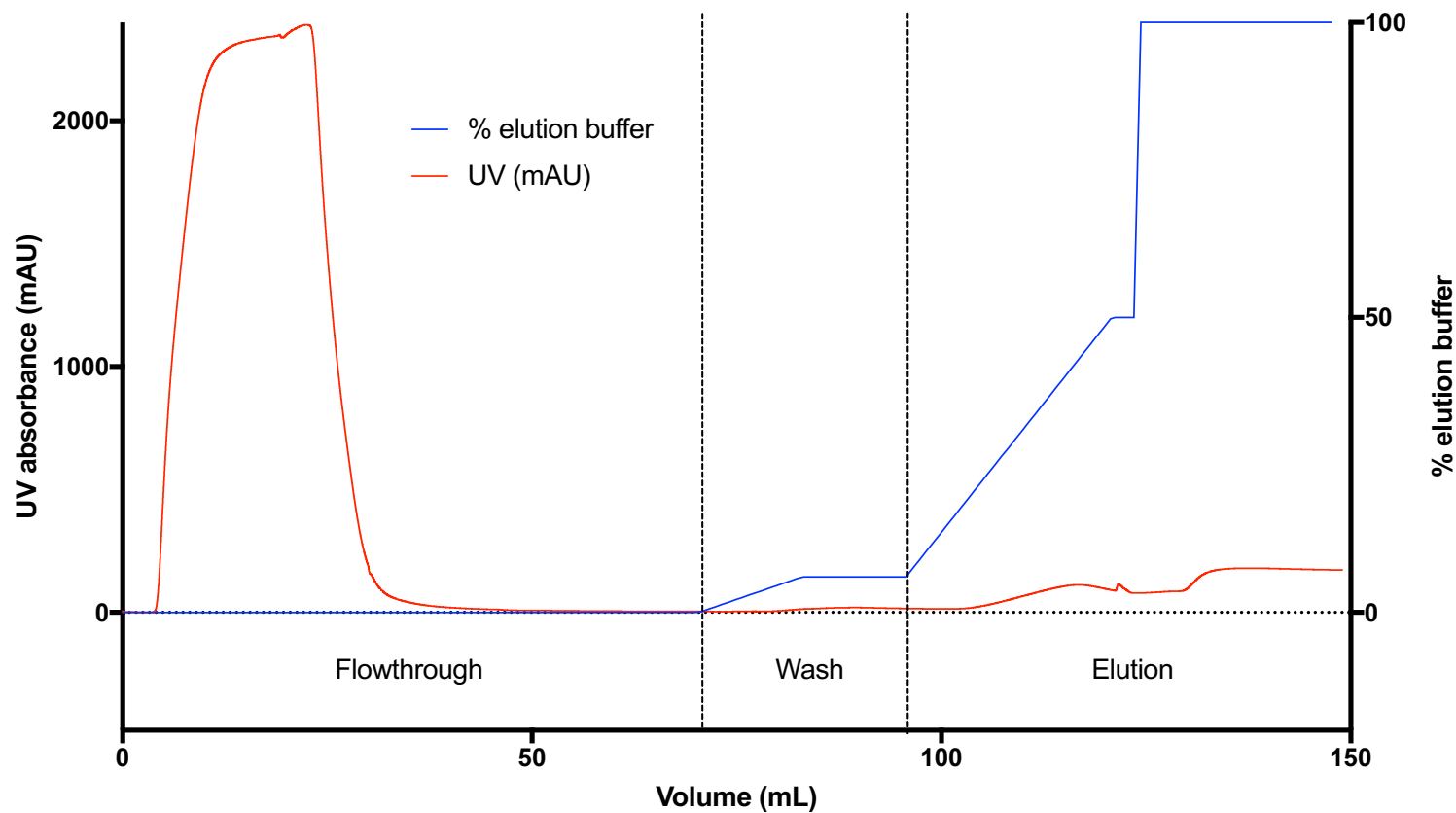


Figure 9.10. Chromatogram produced during the purification of LldR-15aa-VP64 (*Pseudomonas aeruginosa*). The x-axis shows the volume of buffer progressively passed through the nickel-NTA column. The left y-axis shows the ultraviolet (UV) absorbance, given in milli-arbitrary units (mUA), which is indicative of protein content. The right y-axis shows the percentage of the elution buffer (containing 500 mM imidazole) present in the buffer being passed through the nickel-NTA column. The flowthrough, wash and elution stages of the purification run are demarcated in the figure by the vertical dotted lines.

9.6 Evaluation of purification fraction content by SDS-PAGE

In all figures in this section, F refers to the flowthrough fractions, W refers to the wash fractions, and E refers to the elution fractions.

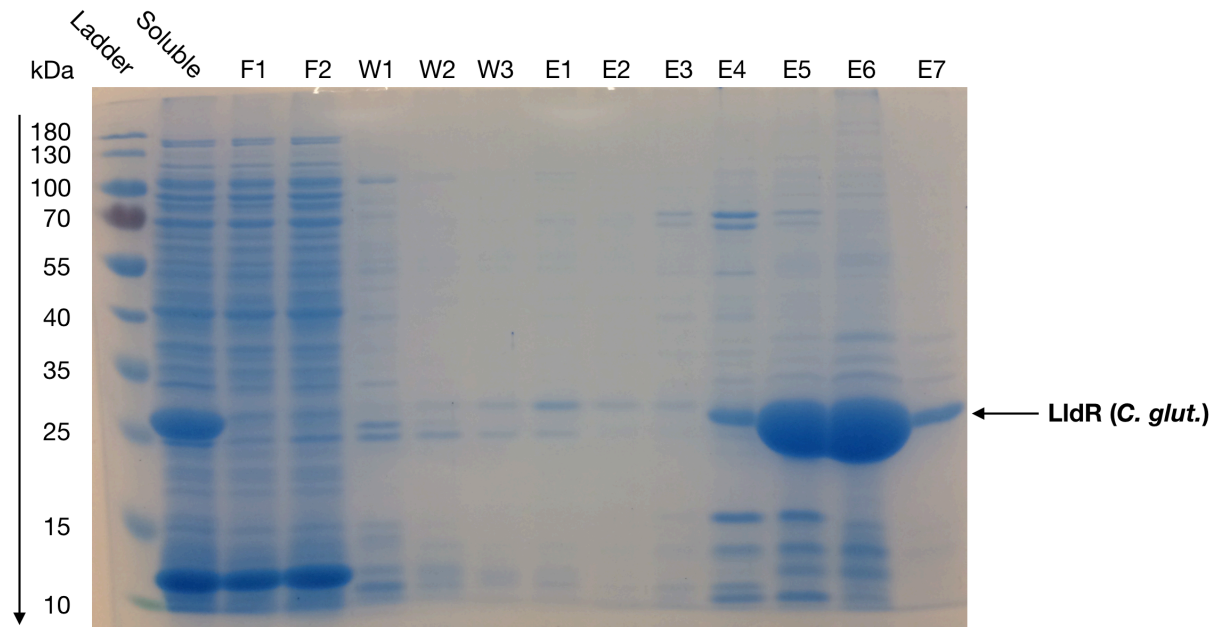


Figure 9.11. SDS-PAGE analysis of the fractions produced during the purification of LldR (*Corynebacterium glutamicum*).

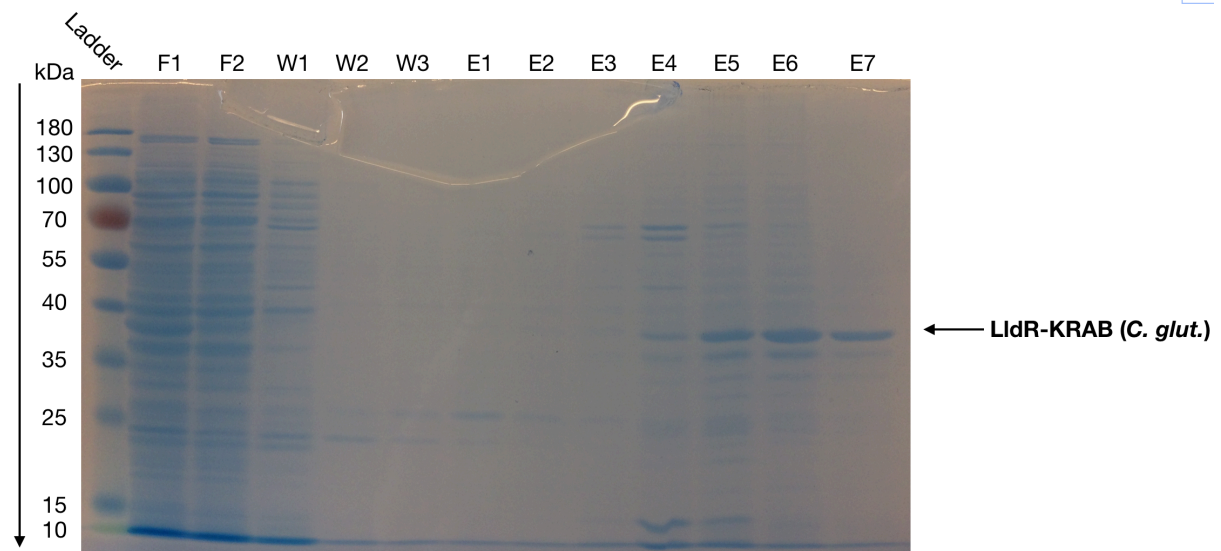


Figure 9.12. SDS-PAGE analysis of the fractions produced during the purification of LldR-KRAB (*Corynebacterium glutamicum*).

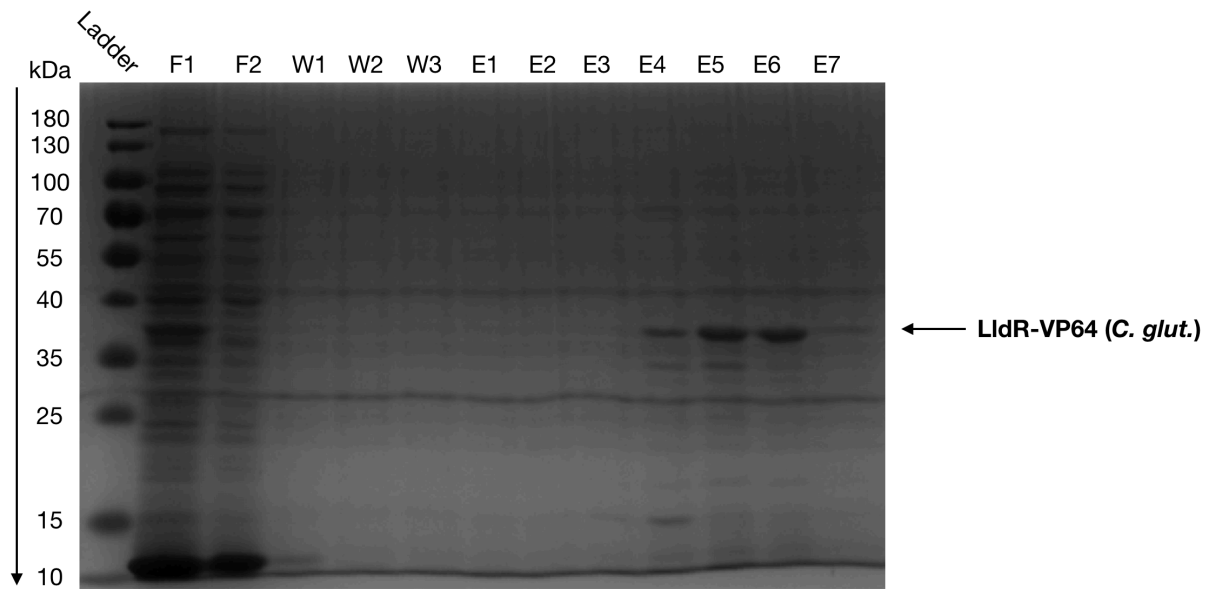


Figure 9.13. SDS-PAGE analysis of the fractions produced during the purification of LldR-VP64 (*Corynebacterium glutamicum*).

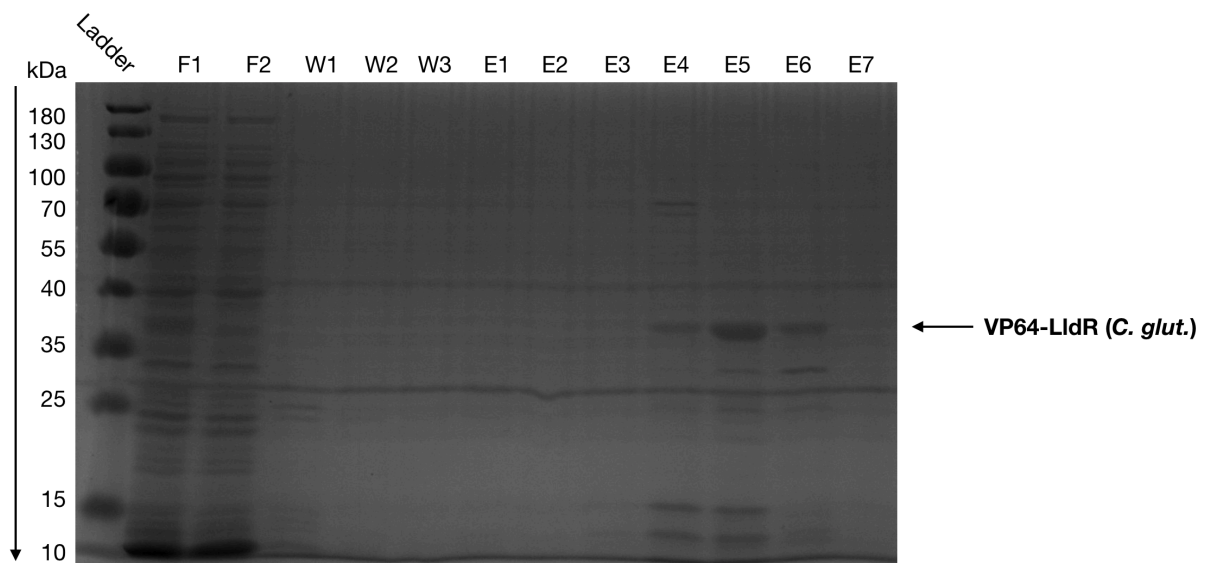


Figure 9.14. SDS-PAGE analysis of the fractions produced during the purification of VP64-LldR (*Corynebacterium glutamicum*).

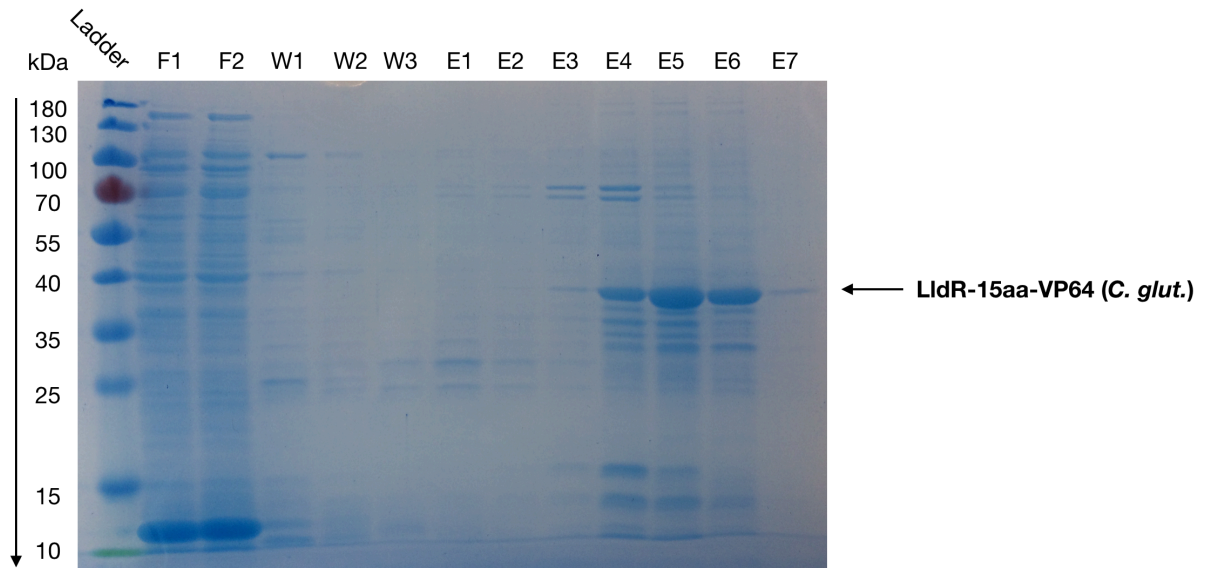


Figure 9.15. SDS-PAGE analysis of the fractions produced during the purification of LldR-15aa-VP64 (*Corynebacterium glutamicum*).

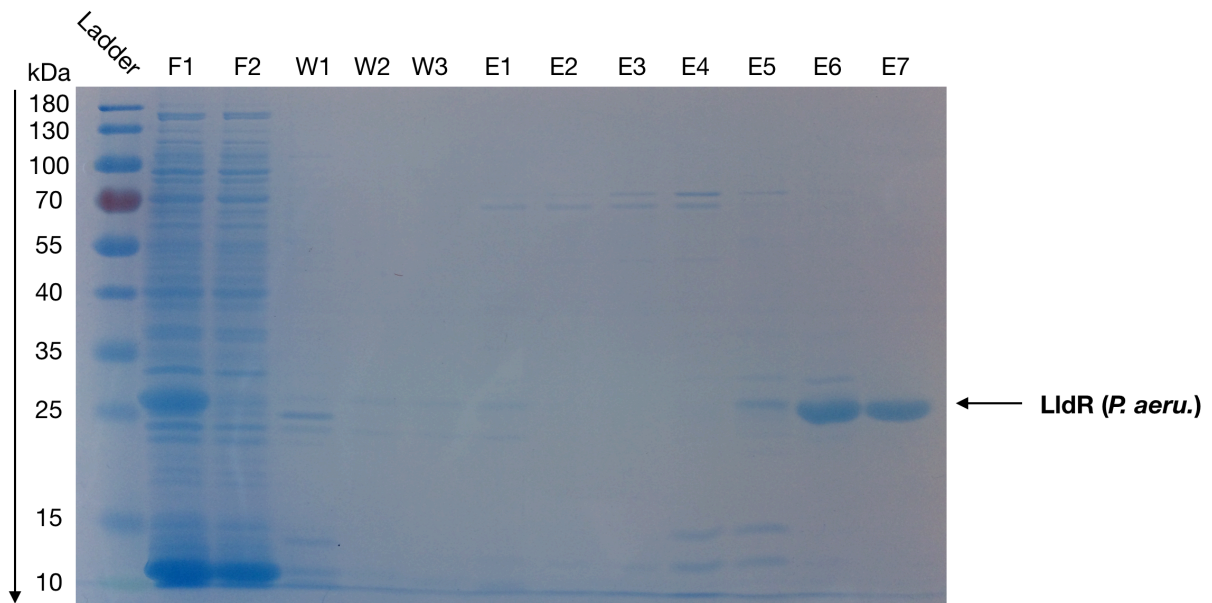


Figure 9.16. SDS-PAGE analysis of the fractions produced during the purification of LldR (*Pseudomonas aeruginosa*).

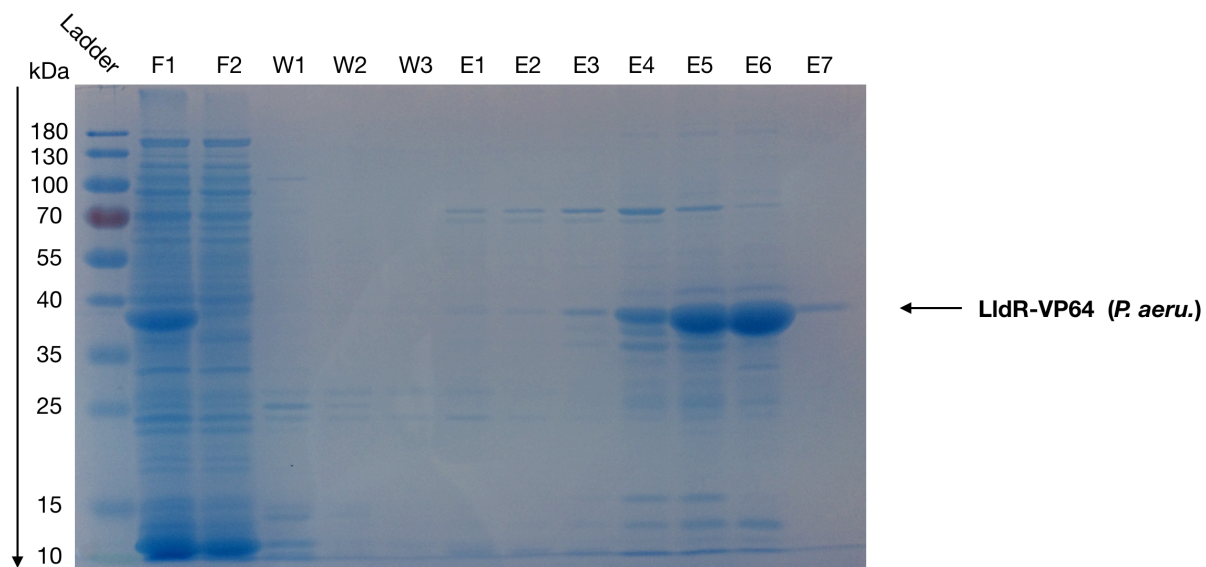


Figure 9.17. SDS-PAGE analysis of the fractions produced during the purification of LldR-VP64 (*Pseudomonas aeruginosa*).

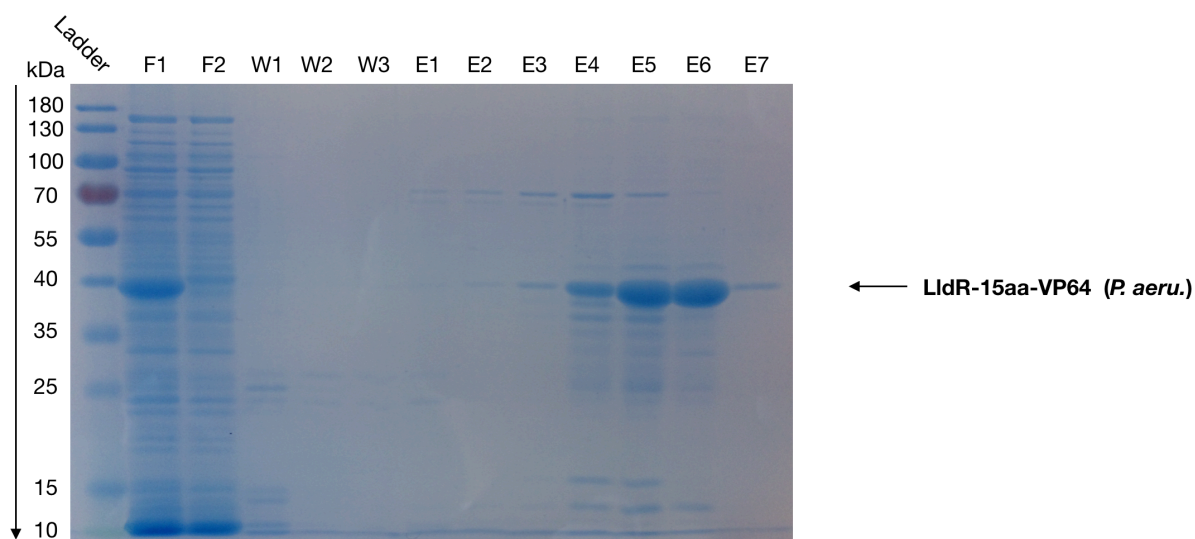


Figure 9.18. SDS-PAGE analysis of the fractions produced during the purification of LldR-15aa-VP64 (*Pseudomonas aeruginosa*).

9.7 Oligonucleotides used in EMSA experiments

Table 9.5. Double-stranded oligonucleotides used in EMSA experiments as part of *in vitro* optimisation. Text in bold shows the operator repeat sequences.

Oligonucleotide name	Forward sequence	Notes
cgl2917	TTGT GGTCTGACCATGA	Found in the genome of <i>Corynebacterium glutamicum</i> .
1xOperator	CTGT GGTCTGACCATGC	Used in the initial <i>in vivo</i> experiments performed in Chapter 3, investigated further in Chapter 4.
2xOperatorA	GCAT GGTCAGACCACAGTCGAGCATGGTCAGACCACAG	Two operator repeats spaced by 10 bases.
2xOperatorB	GCAT GGTCAGACCACAGTCAGCGATCTAGTCGAGCATGGTCAGACCACAG	Two operator repeats spaced by 22 bases.

9.8 EMSA experiments of LldR and variants with cgl2917

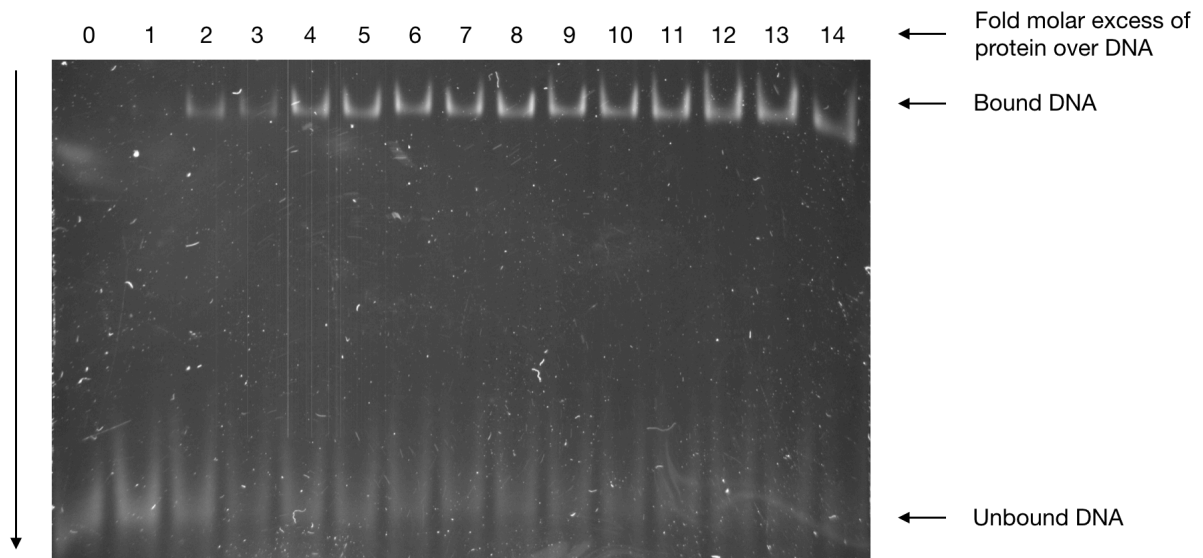


Figure 9.19. EMSA showing cgl2917 incubated with increasing concentrations of LldR (*Corynebacterium glutamicum*).

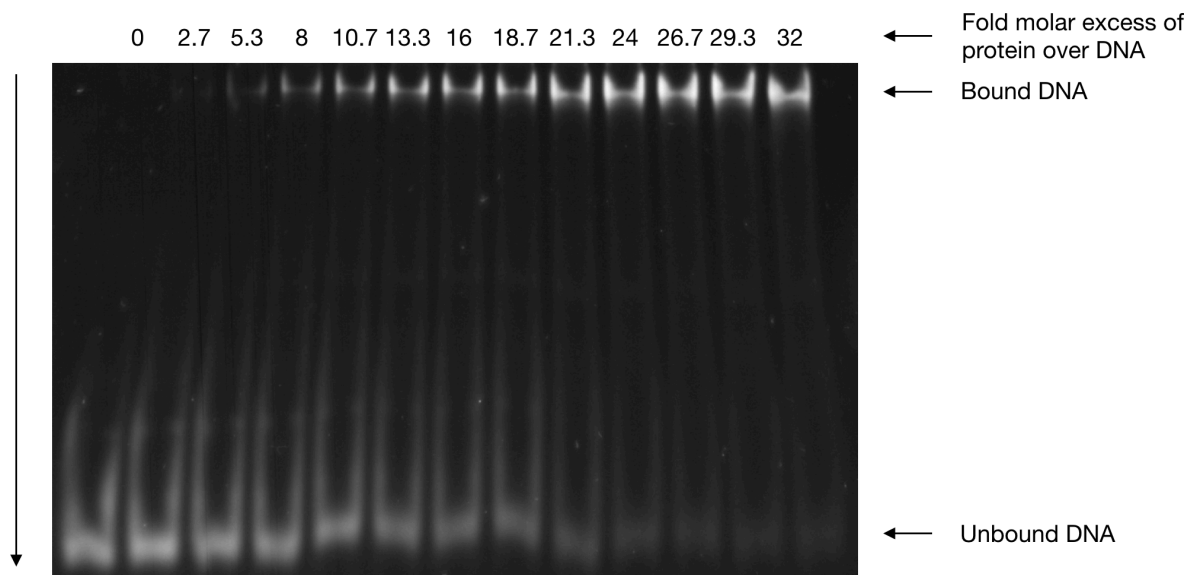


Figure 9.20. EMSA showing cgl2917 incubated with increasing concentrations of LldR-VP64 (*Corynebacterium glutamicum*).

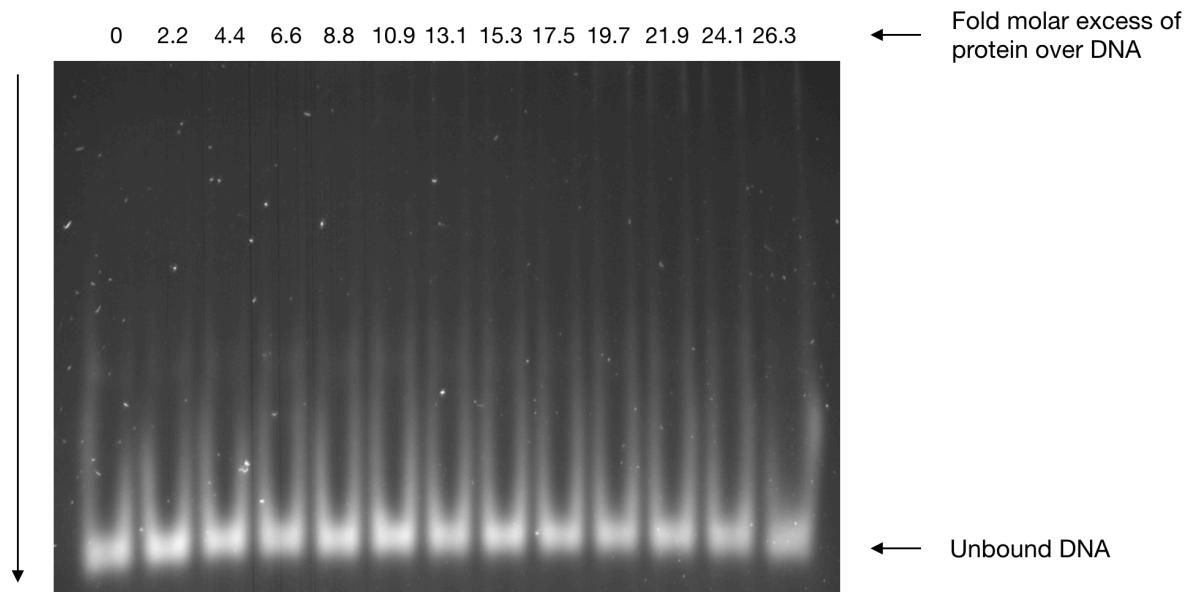


Figure 9.21. EMSA showing *cgl2917* incubated with increasing concentrations of VP64-LldR (*Corynebacterium glutamicum*).

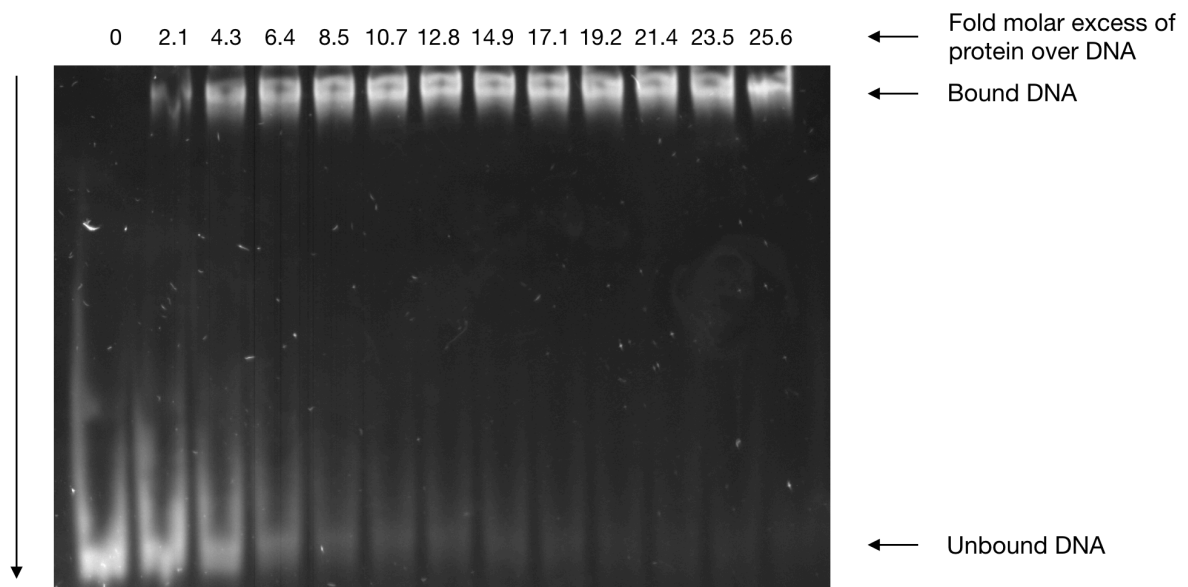


Figure 9.22. EMSA showing *cgl2917* incubated with increasing concentrations of LldR-15aa-VP64 (*Corynebacterium glutamicum*).

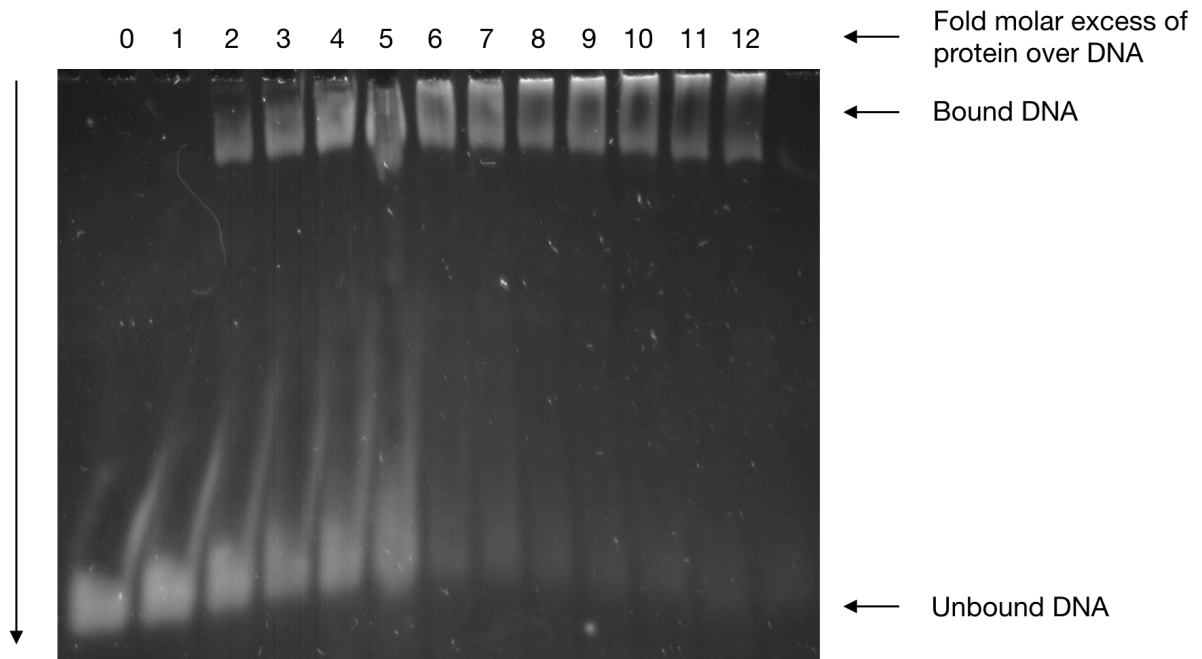


Figure 9.23. EMSA showing *cgl2917* incubated with increasing concentrations of LldR (*Pseudomonas aeruginosa*).

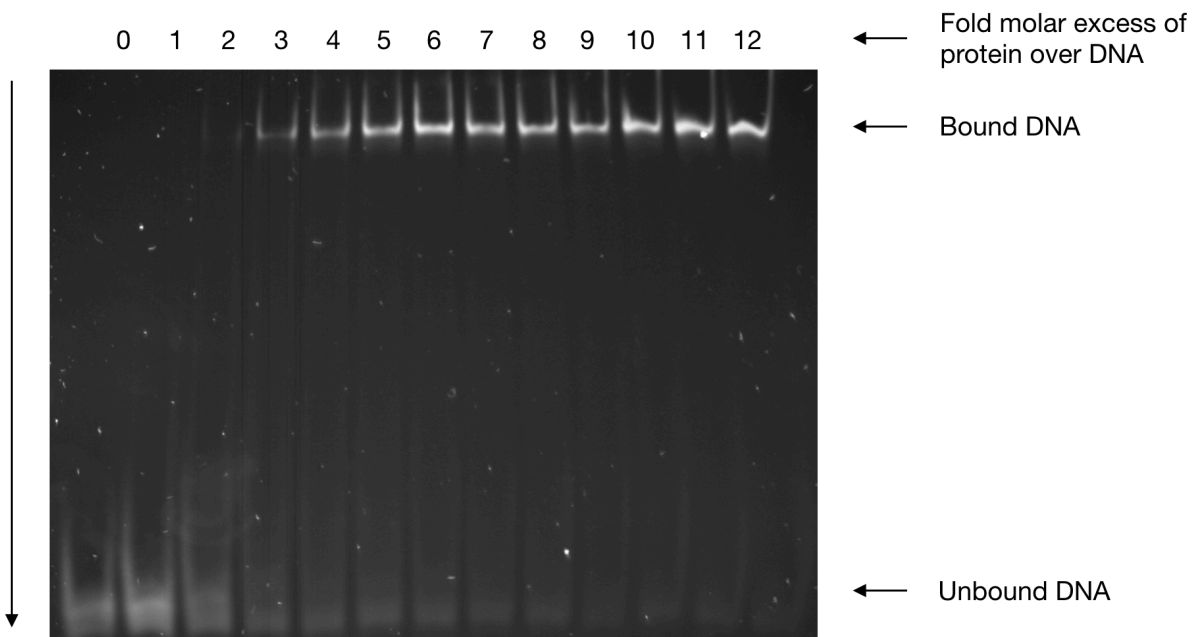


Figure 9.24. EMSA showing *cgl2917* incubated with increasing concentrations of LldR-VP64 (*Pseudomonas aeruginosa*).

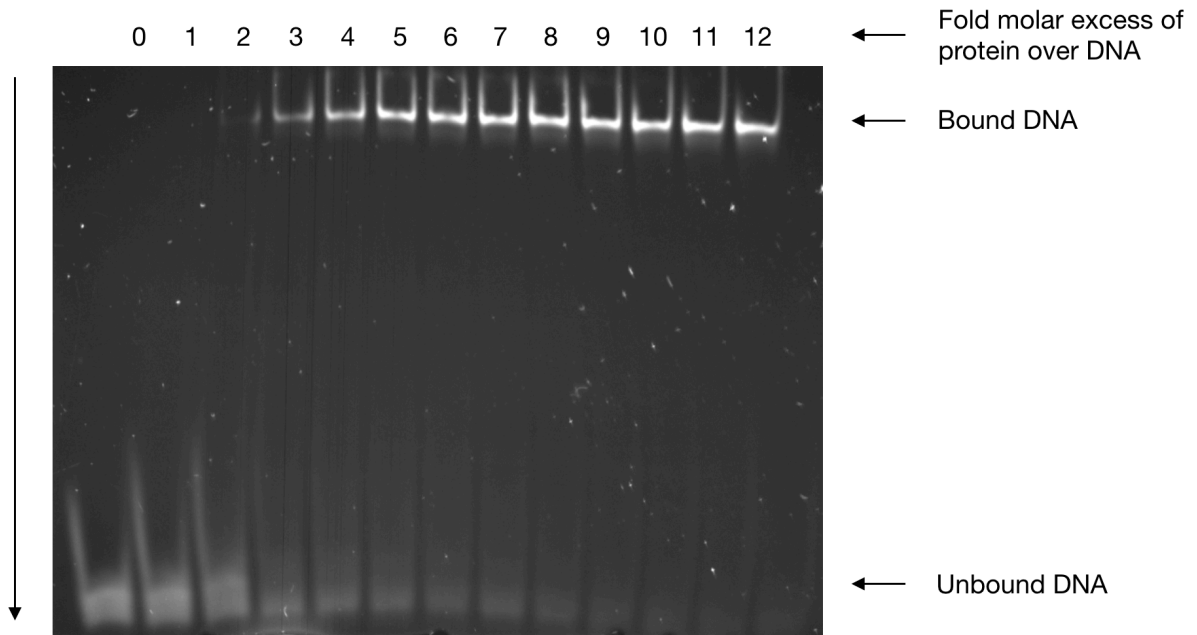


Figure 9.25. EMSA showing *cgl2917* incubated with increasing concentrations of LldR-15aa-VP64 (*Pseudomonas aeruginosa*).

9.9 Constructs used in chapter 5

Table 9.6. List of constructs designed for *in vivo* testing of an optimized lactate-inducible transgene expression system. Constructs 1 to 4 encode LldR only, constructs 5 to 18 encode lactate response elements with LldR operators (1x/2x indicating the number of operators present), and constructs 19 to 46 encode both LldR and lactate response elements. *C. g.* refers to *Corynebacterium glutamicum*, and *P. a.* refers to *Pseudomonas aeruginosa*.

Construct #	Construct name	Induction configuration	Single or double construct?	Successfully constructed?
1	CMV - L (<i>C. g.</i>)	De-repression	Single	Yes
2	CMV - L (<i>C. g.</i>) - linker - VP64	Transactivation	Single	Yes
3	CMV - L (<i>P. a.</i>)	De-repression	Single	Yes
4	CMV - L (<i>P. a.</i>) - VP64	Transactivation	Single	Yes
5	CMV - 1x - mAzamiGreen	De-repression	Single	Yes
6	CMV - 2x - mAzamiGreen	De-repression	Single	Yes
7	minCMV - 1x - mAzamiGreen	Transactivation	Single	Yes
8	minCMV - 2x - mAzamiGreen	Transactivation	Single	Yes
9	1x - minCMV - mAzamiGreen	Transactivation	Single	Yes

10	2x - minCMV - mAzamiGreen	Transactivation	Single	Yes
11	1x - minCMV - 1x - mAzamiGreen	Transactivation	Single	Yes
12	2x - minCMV - 2x - mAzamiGreen	Transactivation	Single	Yes
13	YB_TATA - 1x - mAzamiGreen	Transactivation	Single	Yes
14	YB_TATA - 2x - mAzamiGreen	Transactivation	Single	Yes
15	1x - YB_TATA - mAzamiGreen	Transactivation	Single	Yes
16	2x - YB_TATA - mAzamiGreen	Transactivation	Single	Yes
17	1x - YB_TATA - 1x - mAzamiGreen	Transactivation	Single	No
18	2x - YB_TATA - 2x - mAzamiGreen	Transactivation	Single	No
19	CMV - L (<i>C. g.</i>) - CMV - 1x - mAzamiGreen	De-repression	Double	No
20	CMV - L (<i>C. g.</i>) - CMV - 2x - mAzamiGreen	De-repression	Double	No
21	CMV - L (<i>P. a.</i>) - CMV - 1x - mAzamiGreen	De-repression	Double	Yes
22	CMV - L (<i>P. a.</i>) - CMV - 2x - mAzamiGreen	De-repression	Double	No
23	CMV - L (<i>C. g.</i>) - linker - VP64 - minCMV - 1x - mAzamiGreen	Transactivation	Double	Yes
24	CMV - L (<i>C. g.</i>) - linker - VP64 - minCMV - 2x - mAzamiGreen	Transactivation	Double	Yes
25	CMV - L (<i>C. g.</i>) - linker - VP64 - 1x - minCMV - mAzamiGreen	Transactivation	Double	Yes
26	CMV - L (<i>C. g.</i>) - linker - VP64 - 2x - minCMV - mAzamiGreen	Transactivation	Double	Yes

27	CMV - L (<i>C. g.</i>) - linker - VP64 - 1x - minCMV - 1x - mAzamiGreen	Transactivation	Double	Yes
28	CMV - L (<i>C. g.</i>) - linker - VP64 - 2x - minCMV - 2x - mAzamiGreen	Transactivation	Double	Yes
29	CMV - L (<i>C. g.</i>) - linker - VP64 - YB_TATA - 1x - mAzamiGreen	Transactivation	Double	No
30	CMV - L (<i>C. g.</i>) - linker - VP64 - YB_TATA - 2x - mAzamiGreen	Transactivation	Double	Yes
31	CMV - L (<i>C. g.</i>) - linker - VP64 - 1x - YB_TATA - mAzamiGreen	Transactivation	Double	Yes
32	CMV - L (<i>C. g.</i>) - linker - VP64 - 2x - YB_TATA - mAzamiGreen	Transactivation	Double	Yes
33	CMV - L (<i>C. g.</i>) - linker - VP64 - 1x - YB_TATA - 1x - mAzamiGreen	Transactivation	Double	No
34	CMV - L (<i>C. g.</i>) - linker - VP64 - 2x - YB_TATA - 2x - mAzamiGreen	Transactivation	Double	No
35	CMV - L (<i>P. a.</i>) - VP64 - minCMV - 1x - mAzamiGreen	Transactivation	Double	Yes
36	CMV - L (<i>P. a.</i>) - VP64 - minCMV - 2x - mAzamiGreen	Transactivation	Double	Yes
37	CMV - L (<i>P. a.</i>) - VP64 - 1x - minCMV - mAzamiGreen	Transactivation	Double	Yes
38	CMV - L (<i>P. a.</i>) - VP64 - 2x - minCMV - mAzamiGreen	Transactivation	Double	Yes
39	CMV - L (<i>P. a.</i>) - VP64 - 1x - minCMV - 1x - mAzamiGreen	Transactivation	Double	Yes
40	CMV - L (<i>P. a.</i>) - VP64 - 2x - minCMV - 2x - mAzamiGreen	Transactivation	Double	Yes
41	CMV - L (<i>P. a.</i>) - VP64 - YB_TATA - 1x - mAzamiGreen	Transactivation	Double	No
42	CMV - L (<i>P. a.</i>) - VP64 - YB_TATA - 2x - mAzamiGreen	Transactivation	Double	Yes
43	CMV - L (<i>P. a.</i>) - VP64 - 1x - YB_TATA - mAzamiGreen	Transactivation	Double	Yes

44	CMV - L (<i>P. a.</i>) - VP64 - 2x - YB_TATA - mAzamiGreen	Transactivation	Double	Yes
45	CMV - L (<i>P. a.</i>) - VP64 - 1x - YB_TATA - 1x - mAzamiGreen	Transactivation	Double	No
46	CMV - L (<i>P. a.</i>) - VP64 - 2x - YB_TATA - 2x - mAzamiGreen	Transactivation	Double	No

9.10 Summary of results from the transfection experiments in section 5.

Table 9.8. Detailed summary of results from the transfection experiments conducted with optimized *in vivo* constructs. The p-value is calculated through the use of a Student's T-test (one-tailed, two-sampled, equal variance) to determine whether the uninduced and induced populations were significantly different. *C. g.* refers to the *Corynebacterium glutamicum*, and *P. a.* refers to *Pseudomonas aeruginosa*.

NICE #	Exp. #	Construct numbers	Construct name(s)	Average geometric mean of GFP (AU)		FC induction (-lactate/+lactate)	p-value	Configuration type?	Do intended and apparent induction types match?
				-lactate	+lactate				
NICE1	1	1+5	L (<i>C. g.</i>) + CMV-1x	301377	221802	1.36	0.0712	Derepression	No
NICE1	2	1+6	L (<i>C. g.</i>) + CMV-2x	284785	206855	1.38	0.0757	Derepression	No
NICE1	3	3+5	L (<i>P. a.</i>) + CMV-1x	9009	9583	0.94	0.0277	Derepression	Yes
NICE1	4	3+6	L (<i>P. a.</i>) + CMV-2x	79259	69890	1.13	0.3383	Derepression	No

NICE1	5	2+7	L15V (<i>C. g.</i>) + minCMV- 1x	29517	14617	2.02	0.0045	Transactivation	Yes
NICE1	6	2+8	L15V (<i>C. g.</i>) + minCMV- 2x	38326	11062	3.46	0.0041	Transactivation	Yes
NICE1	7	2+9	L15V (<i>C. g.</i>) + 1x- minCMV	22832	9174	2.49	0.0084	Transactivation	Yes
NICE1	8	2+10	L15V (<i>C. g.</i>) + 2x- minCMV	16027	8658	1.85	0.0393	Transactivation	Yes
NICE1	9	2+11	L15V (<i>C. g.</i>) + 1x- minCMV-1x	17295	15393	1.12	0.0007	Transactivation	Yes
NICE1	10	2+12	L15V (<i>C. g.</i>) + 2x- minCMV-2x	23953	21764	1.10	0.0311	Transactivation	Yes
NICE1	12	2+14	L15V (<i>C. g.</i>) +YB-2x	15757	15057	1.05	0.2841	Transactivation	Yes

NICE1	13	2+15	L15V (<i>C. g.</i>) +1x-YB	12422	11285	1.10	0.1229	Transactivation	Yes
NICE1	14	2+16	L15V (<i>C. g.</i>) +2x-YB	12970	8742	1.48	0.0045	Transactivation	Yes
NICE1	17	4+7	LV (<i>P. a.</i>) + minCMV-1x	19163	17733	1.08	0.3418	Transactivation	Yes
NICE1	18	4+8	LV (<i>P. a.</i>) + minCMV-2x	35629	23779	1.50	0.0007	Transactivation	Yes
NICE1	19	4+9	LV (<i>P. a.</i>) + 1x-minCMV	25390	26625	0.95	0.3422	Transactivation	No
NICE1	20	4+10	LV (<i>P. a.</i>) + 2x-minCMV	24306	25297	0.96	0.3304	Transactivation	No
NICE1	21	4+11	LV (<i>P. a.</i>) + 1x-minCMV- 1x	23768	25457	0.93	0.4015	Transactivation	No
NICE1	22	4+12	LV (<i>P. a.</i>) + 2x-minCMV- 2x	44188	21257	2.08	0.0382	Transactivation	Yes

NICE1	24	4+14	LV (<i>P. a.</i>) + YB-2x	23123	8635	2.68	0.0087	Transactivation	Yes
NICE2	25	4+15	LV (<i>P. a.</i>) + 1x-YB	18589	16953	1.10	0.0650	Transactivation	Yes
NICE3	25			16394	14585	1.12	0.0037	Transactivation	Yes
NICE4	25			10792	9286	1.16	0.0115	Transactivation	Yes
NICE2	26	4+16	LV (<i>P. a.</i>) + 2x-YB	18864	17955	1.05	0.1771	Transactivation	Yes
NICE3	26			19481	16707	1.17	0.0106	Transactivation	Yes
NICE4	26			13459	10485	1.28	0.0324	Transactivation	Yes
NICE2	33	23	L15V (<i>C. g.</i>)- minCMV-1x	14786	10833	1.36	0.0039	Transactivation	Yes
NICE3	33			15167	10588	1.43	0.0198	Transactivation	Yes
NICE4	33			12716	10353	1.23	0.0143	Transactivation	Yes
NICE2	34	24	L15V (<i>C. g.</i>)- minCMV-2x	14525	12187	1.19	0.2616	Transactivation	Yes
NICE3	34			13884	11087	1.25	0.0633	Transactivation	Yes
NICE4	34			12370	11131	1.11	0.0276	Transactivation	Yes

NICE2	35	25	L15V (<i>C. g.</i>)- 1x-minCMV	15738	16804	0.94	0.1203	Transactivation	No
NICE3	35			9709	10067	0.96	0.1001	Transactivation	No
NICE4	35			9489	8501	1.12	0.0466	Transactivation	Yes
NICE2	36	26	L15V (<i>C. g.</i>)- 2x-minCMV	15391	16305	0.94	0.1624	Transactivation	No
NICE3	36			12931	13265	0.97	0.2087	Transactivation	No
NICE4	36			10257	8582	1.20	0.0088	Transactivation	Yes
NICE2	37	27	L15V (<i>C. g.</i>)- 1x-minCMV- 1x	14785	14635	1.01	0.4094	Transactivation	Yes
NICE3	37			12534	12436	1.01	0.3854	Transactivation	Yes
NICE4	37			10942	9138	1.20	0.0130	Transactivation	Yes
NICE2	38	28	L15V (<i>C. g.</i>)- 2x-minCMV- 2x	19455	17609	1.10	0.0148	Transactivation	Yes
NICE3	38			13193	15480	0.85	0.0154	Transactivation	No
NICE4	38			10086	9764	1.03	0.2445	Transactivation	Yes

NICE2	41	31	L15V (C. g.)- 1x-YB	7604	7170	1.06	0.0374	Transactivation	Yes
NICE3	41			7524	6913	1.09	0.0592	Transactivation	Yes
NICE4	41			6056	5703	1.06	0.0833	Transactivation	Yes
NICE2	42	32	L15V (C. g.)- 2x-YB	8406	7993	1.05	0.2468	Transactivation	Yes
NICE3	42			7380	7408	1.00	0.4785	Transactivation	No
NICE4	42			6229	5413	1.15	0.0000	Transactivation	Yes
NICE2	45	35	LV (P. a.)- minCMV-1x	15868	14450	1.10	0.0691	Transactivation	Yes
NICE3	45			11875	12043	0.99	0.4533	Transactivation	No
NICE4	45			8468	7381	1.15	0.0438	Transactivation	Yes
NICE2	46	36	LV (P. a.)- minCMV-2x	14140	13673	1.03	0.1627	Transactivation	Yes
NICE3	46			12008	12074	0.99	0.4745	Transactivation	No
NICE4	46			9114	8548	1.07	0.0813	Transactivation	Yes
NICE2	47	37	LV (P. a.)-1x- minCMV	28289	20260	1.40	0.0365	Transactivation	Yes

NICE3	47			22700	22735	1.00	0.4930	Transactivation	No
NICE4	47			13210	9463	1.40	0.0253	Transactivation	Yes
NICE2	48	38	LV (<i>P. a.</i>)-2x- minCMV	50455	22984	2.20	0.0112	Transactivation	Yes
NICE3	48			33884	22182	1.53	0.0669	Transactivation	Yes
NICE4	48			21902	15074	1.45	0.1127	Transactivation	Yes
NICE2	49	39	LV (<i>P. a.</i>)-1x- minCMV-1x	34905	36061	0.97	0.4770	Transactivation	No
NICE3	49			25439	12807	1.99	0.0237	Transactivation	Yes
NICE4	49			9217	15924	0.58	0.1153	Transactivation	No
NICE2	53	43	LV (<i>P. a.</i>)-1x- YB	7813	7683	1.02	0.1589	Transactivation	Yes
NICE3	53			7229	6684	1.08	0.0882	Transactivation	Yes
NICE4	53			5645	5458	1.03	0.2051	Transactivation	Yes

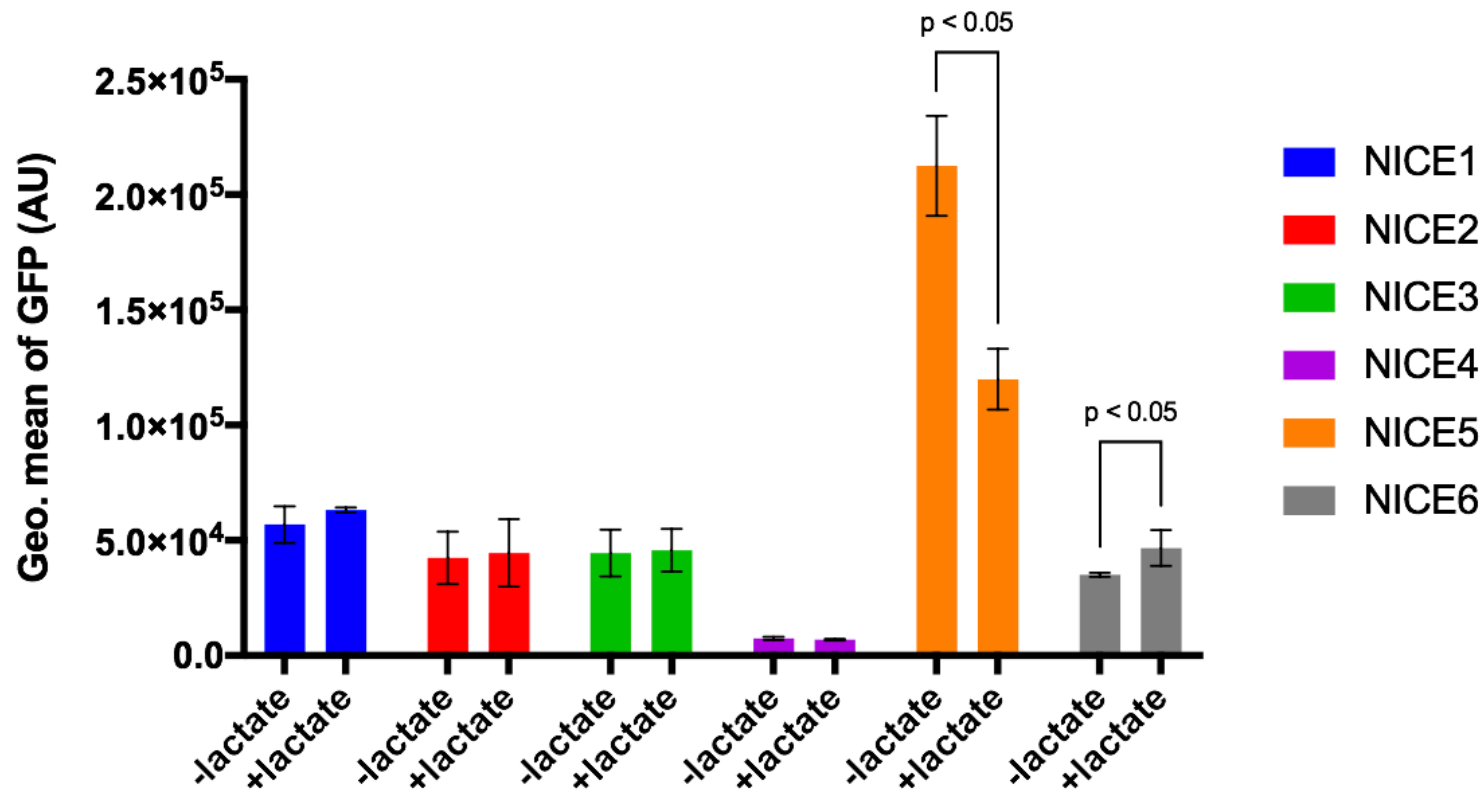


Figure 9.26. The impact of lactate on the GFP signal in mAzamiGreen-expressing controls. In 4 cases out of 6, no significant change in the GFP signal was seen upon the addition of 20 mM lactate. In the case of NICE5, a significant decrease in expression was seen, and in the case of NICE6, a significant increase was seen. Error bars show the standard deviation of three biological replicates. A Student's T-test (one-tailed, two-sampled, equal variance) was used to determine the p values for each experiment. AU is an abbreviation for arbitrary units.

9.11 Constructs used in section 6

Table 9.9. Constructs used to test the ability of Cas9-VPR to upregulate genes in CHO cells.

Plasmid name	Notes	Origin
hCas9-VPR	Expresses the endonucleolytically active Cas9 protein fused to the powerful VPR tripartite transactivator.	Chavez et al., 2015
pGL3-U6-sgRNA-sfGFP-dropout	The U6 promoter is upstream of two Bsal cleavage sites, which flank a sfGFP sequence. When transformed into <i>E. coli</i> , colonies will visibly express sfGFP. If the plasmid is first digested with Bsal and a complementary guide sequence is ligated in, sfGFP is no longer expressed and the plasmid is likely to have successfully integrated the guide DNA of interest.	This study
minCMV-mAzamiGreen	A minimal promoter expressing a green fluorescent protein without any additional cis-regulatory elements.	This study
minCMV-targeting guide RNA 1	A plasmid expressing a 14-nucleotide guide RNA targeting site 1 of the minCMV-mAzamiGreen plasmid.	This study
minCMV-targeting guide RNA 2	A plasmid expressing a 14-nucleotide guide RNA targeting site 2 of the minCMV-mAzamiGreen plasmid.	This study
minCMV-targeting guide RNA 3	A plasmid expressing a 14-nucleotide guide RNA targeting site 3 of the minCMV-mAzamiGreen plasmid.	This study

minCMV-targeting guide RNA 4	A plasmid expressing a 14-nucleotide guide RNA targeting site 4 of the minCMV-mAzamiGreen plasmid.	This study
minCMV-targeting guide RNA 5	A plasmid expressing a 14-nucleotide guide RNA targeting site 5 of the minCMV-mAzamiGreen plasmid.	This study
MCT1-targeting guide RNA 4-14	A plasmid expressing a 14-nucleotide guide RNA targeting site 4 of the endogenous MCT1 gene.	This study
MCT1-targeting guide RNA 5-14	A plasmid expressing a 14-nucleotide guide RNA targeting site 5 of the endogenous MCT1 gene.	This study
MCT1-targeting guide RNA 5-20	A plasmid expressing a 20-nucleotide guide RNA targeting site 5 of the endogenous MCT1 gene.	This study
MCT1-targeting guide RNA 6-14	A plasmid expressing a 14-nucleotide guide RNA targeting site 6 of the endogenous MCT1 gene.	This study
MCT1-targeting guide RNA 6-16	A plasmid expressing a 16-nucleotide guide RNA targeting site 6 of the endogenous MCT1 gene.	This study
MCT1-targeting guide RNA 6-20	A plasmid expressing a 20-nucleotide guide RNA targeting site 6 of the endogenous MCT1 gene.	This study

9.12 Table of primer pairs used in RT-qPCR

Table 9.10. Primer pairs used to quantify relative transcript levels in CHO.

Primer pair name	Forward sequence	Reverse sequence	Amplification efficiency	Origin
ACTB-2	TTGAACACGGCATTGTCACC	AGCTCGTTGTAGAAGGTGTGG	90.4%	This study
MCT1-2	TGGCTGTCATGTATGCTGGAG	AGCTGCAATCAAGCCACAAC	97.3%	This study

9.13 All DNA sequences used

>LldR (*Corynebacterium glutamicum*)

```
ATGAGCGTTAAGGCCACGAGAGTGTGATGGACTGGGTGACAGAAGAGCTTAGAAGCGGACG
GCTGAAAATCGGGGATCATCTTCCATCTGAGAGAGCCCTGAGTGAGACTCTGGGGGTATCTA
GAAGTAGTCTGCGCGAGGCATTGCGCGTTCTGGAGGCTCTGGGCACCATTTCTACAGCTACA
GGCTCCGGTCCCAGGTCTGGCACAAATTATTACCGCTGCCCCAGGGCAGGCTCTGTCCCTGTC
AGTCACCCTGCAGCTTGTGACAAATCAGGTGGGACACCATGATATCTACGAAACCCGCCAGC
TCTTGGAAGGCTGGGCTGCTCTCCACTCTTCCGCAGAAAGAGGAGACTGTGACGTGGCTGAG
GCTCTGCTGGAGAAAATGGACGACCCCTCACTCCCTCTGGAGGATTTTCTGCGCTTCGACGC
TGAATTTTCATGTGGTCATTAGTAAGGGAGCTGAGAACCCTCTGATTTCTACTTTGATGGAAG
CACTCAGATTGAGCGTAGCCGACCATAACAGTTGCCCGTGCTCGCGCCCTTCCAGATTGGCGG
GCCACTAGTGCTCGTCTCCAGAAGGAGCATCGGGCCATTCTGGCAGCACTCCGCGCAGGGGA
ATCCACCGTAGCAGCCACACTGATTAAAGAGCATATCGAGGGTTACTATGAGGAGACAGCCG
CAGCTGAGGCA
```

>LldR (*Pseudomonas aeruginosa*)

```
ATGGAATTTGGTCAGGTCAGGCAGCGCCGTCTGTGCGGATGACATCGTTGCGCAACTGGAGGC
GATGATCCTGGAGGGCAGCTGAAGTCCGGCGAGCGGCTGCCCGCCGAGCGCGTGCTTGCCG
AGCAGTTCGGGGTTTTCCCGGCCGTCGCTGCGCGAGGCGATCCAGAAACTGGTGGCGAAGGGG
CTGCTGGTCAGCCGCCAGGGTGGCGGCAACTATGTGACCGAATCGCTGGGCGCGACTTTCAG
CGATCCGCTGCTGCACCTGCTGGAGGGTAACCCGGAGGCCAGCGCGACCTGCTGGAGTTTC
GCCACACGCTGGAAGGGTCTGTGCCTACTACGCGGCGCTGCGTGCGACCTCCCTCGATCAC
CAGCGCCTGACCGAGGCCTTCGAGGGCGCTCCAGGCCTGCTATGCGCGCAACGACCAGGTCAG
CGCGGAAGAGGGTGCCGCCGACGCGCGCTTCCACCTGGCGATCGCCGAGGCCAGCCACAACA
CCGTGCTGCTGCACACCATCAAGGGCCTGTTGACTTGCTGCGGCGCAACGTGGTGACCAAT
ATCGGCGGGATGTACGCGCAGCGCACGGAAACCCGCGCGCAACTGATGGAGCAGCACCAGCG
CCTATACGACGCGATCATCAGCGGTCAGGCGGAGCTGGCCCGGGAGGTGTCCAACCAGCACA
TCCACTATGTGCAGGAGGTCTTGGCGGAGGTCCAGGAAGAGGCGCGCAGGATGAAGCGCTCG
CAGCGCCGGCGCAGCGTGCAGGAAGAC
```

>CMV

```
tagttattaatagtaatcaattacgggggtcattagttcatagcccatatatggagttccgcg
ttacataacttacggtaaatggcccgcctggctgaccgcccacgacccccgcccattgacg
tcaataatgacgtatggtcccatagtaacgccaatagggactttccattgacgtcaatgggt
ggagtatttacggtaaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgc
cccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgacctta
tgggactttcctacttggcagtacatctacgtattagtcacgctattaccatggtgatgcg
gttttggcagtacatcaatgggcggtggatagcggtttgactcacggggatttccaagtctcc
acccattgacgtcaatgggagtttgtttggcaccaaaatcaacgggactttccaaaatgt
cgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatat
aagcagagctggtttagtgaaccgctcagatccgctagggatctcagctg
```

>minCMV

```
tatgtcgaggtggcgtgtacggtgggaggcctatataagcagagctcgtttagtgaaccgctc
agatcgcctggagaattcagatctcagctg
```

>YB_TATA

```
TCTAGAGGGTATATAATGGGGGCCA
```

>VP64

GAGGCCAGCGGTTCCGGACGGGCTGACGCATTTGGACGATTTTGATCTGGATATGCTGGGAAG
TGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACC
TCGACATGCTCGGCAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTA ACTCTAGA
AGTTCCGGATCT

>KRAB

ATGGACGCAAAATCTCTCACCGCCTGGTCCAGAACTCTCGTGACATTTAAAGATGTCTTCGT
TGACTTCACACGGGAAGAATGGAAGCTTCTGGATACAGCACAGCAGATCGTGTACCGGAACG
TTATGTTGAAAATTATAAAAATCTTGTGAGTCTGGGCTATCAGCTGACAAAGCCAGATGTG
ATTTTGCCTGGAGAAGGGCGAGGAGCCATGGCTTGTGAGAGGGAGATTCATCAGGAAAC
TCACCCTGACAGTGAGACTGCA

>NLS

CCCCCAAGAAAAAGCGGAAAGTGTGTTAA

>P2A

gctagcggcagcggcgccacaaacttctctctgctaaagcaagcaggtgatggtgaagaaaa
ccccgggcctgcatcg

>mCherry

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGT
GCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCC
CCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCC
TGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGA
CATTCCCGACTACTTGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACT
TCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATC
TACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGAC
CATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGA
TCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTAC
AAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCTACAACGTCAACATCAAGTTGGACATCAC
CTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCA
CCGGCGGCATGGACGAGCTGTACAAG

>mAzamiGreen

atggtgagcgtgatcaagcccagatgaagatcaagctgtgcatgaggggacccgtgaacgg
ccacaacttcgtgatcgagggcgagggcaagggcaaccctacgagggcaccagatcctgg
acctgaacgtgaccgagggcgccccctgcccttcgcctacgacatcctgaccaccgtgttc
cagtacggcaacagggccttcaccaagtaccccgccgacatccaggactacttcaagcagac
cttccccgagggctaccactgggagaggagcatgacctacgaggaccagggcatctgcaccg
ccaccagcaacatcagcatgaggggagactgcttctctacgacatcagggttcgacggcacc
aacttccccccaacggccccgtgatgcagaagaagaccctgaagtgggagcccagcaccga
gaagatgtacgtggaggacggcgtgctgaagggcgacgtgaacatgaggctgctgctggagg
gcgggcgccactacaggtgagacttcaagaccacctacaaggccaagaaggaggtgaggtg
cccgacgcccacaagatcgaccacaggatcgagatcctgaagcagcacaaggactacaacia
ggtgaagctgtacgagaacgccgtggccaggtactccatgctgcccagccaggcc

>SV40pA

tgaatgaatgaccagaggatcataatcagccataaccacattttagtagaggttttacttgcttt
aaaaaacctcccacacctccccctgaacctgaacataaaatgaatgcaattgttgttgtta

acttgtttattgacgcttataatggttacaaataaagcaatagcatcacaatttcacaaat
aaagcatttttttctactgccccgagcttctctgctcactgact

>bGpA

tgaatgaatgactagactgagaacttcaggggtgagtttggggacccttgattggttctttctt
tttcgctattgtaaaattcatgttatatggagggggcaaagttttcaggggtgtggttagaa
tggaagatgtcccttgatcaccatggaccctcatgataattttggtttctttcactttcta
ctctggtgacaaccattgtctcctcttattttcttttcattttctgtaactttttcggttaa
ctttagcttgcaatttgtaacgaatttttaaattcacttttggtttatttgtcagattgtaagt
actttctctaactctctttttttcaaggcaatcaggggtatattatattgtacttcagcacag
ttttagagaacaattggtataattaaatgataaggtagaatattttctgcatataaattctgg
ctggcgtggaaatattcttattggtagaacaactacaccctgggtcatcatcctgcctttct
ctttatggttacaatgatatacactggttgagatgaggataaaatactctgagtccaaaccg
ggccccctctgctaaccatgttcatgccttcttctctttctctacagctcctgggcaacgtgct
ggttggtgctgctcatcattttggcaaagaattcactcctcaggtgcaggctgcctatc
agaaggtgggtggctgggtggtggccaatgccctggctcacaataaccactgagatctttttccc
tctgcaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaag
gaaatttttttcttgcaatagtgtgttggaaattttttggtctctcactcgggaaggacat
atgggagggcaaatcattttaaacaatcagaatgagatatttggttagagtttggaacatat
gccccatgctggctgccatgaacaagggtggctataaagaggtcatcagtatatgaaaca
gccccctgctgtccattccttattccatagaaaagccttgacttgagggttagattttttta
tattttggtttggttatttttttctttaacatccctaaaattttccttacatgttttacta
gccagatttttctcctcctcctgactactcccagtcatagctgtccctcttctcttatgaac
tcgactgc

>bGHpA

tgaatgaatgactgtgccttctagttgccagccatctggtggttgccccctccccgtgcctt
ccttgaccctggaaggtgccactcccactgtcctttcctaataaaaatgaggaaattgcatcg
cattgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaaggggga
ggattgggaagacaatagcagggcatgctggggatgcggtgggctctatg

>6xHisTag

CACCACCACCACCACCAC

>hCas9-VPR

ATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCCGGCTGGGCCGTCAT
TACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACA
GCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTTCGACTCCGGGGAGACGGCCGAAGCCACG
CGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCA
GGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGT
CCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGAC
GAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAG
TACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTGGG
GACTTCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATC
CAACTGGTTCAGACTTACAATCAGCTTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGA
CGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCAGGCGGCTCGAAAACCTCATCGCAC
AGCTCCCTGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTG
ACCCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGA
CACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTT
TTTTGGCGGCAAAGAACCTGTCAGACGCCATCTGCTGAGTGATATTCTGCGAGTGAACACG
GAGATCACAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGA

CTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCT
TCGATCAGTCTAAAAATGGCTACGCCGATAACATTGACGGCGGAGCAAGCCAGGAGGAATTT
TACAAATTTATTAAGCCATCTTGGAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCT
TAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCACCAGA
TTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAA
GATAACAGGGAAAAGATTGAGAAAATCCTCACATTTTCGGATAACCTACTATGTAGGCCCCCT
CGCCCGGGGAAATTCAGATTTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCT
GGAACCTCGAGGAAGTCGTGGATAAAGGGGCCTCTGCCAGTCCTTCATCGAAAGGATGACT
AACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCCCTAACACTCTCTGCTGTACGAGTA
CTTACAGTTTATAACGAGCTCACCAAGGTCAAATACGTACAGAAAGGGATGAGAAAGCCAG
CATTCCCTGTCTGGAGAGCAGAAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAA
GTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGA
AATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAA
TCATTAAGACAAGGACTTCCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTC
CTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACCTTACGCTCA
TCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGC
TGTC AAGAAAACCTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTT
CTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCAC
CTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTACGAGCACA
TCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCTGTG
GATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCG
AGAGAACCAAACCTACCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAG
AGGGTATAAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTT
CAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGA
ACTGGACATCAATCGGCTCTCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCA
AAGATGATTCTATTGATAATAAAGTGTGACAAGATCCGATAAAAATAGAGGGGAAGAGTGAT
AACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAATTTATTGGCGGCAGCTGCTGAACGC
CAAACCTGATCACACAACGGAAGTTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTG
AGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCAC
GTGGCCCAAATTTCTCGATTACGCATGAACACCAAGTACGATGAAAATGACAACTGATTTCG
AGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTTCAGAAAGGACTTTTCA
TTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTG
GTAGGCACTGCACTTATCAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTA
TAAAGTGATACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCG
CTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTTCAAGACCGAGATTACACTGGCCAAT
GGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGA
CAAGGGTAGGGATTTTCGCGACAGTCCGGAAGGTCTGTCCATGCCGCAGGTGAACATCGTTA
AAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGC
GACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCC
TACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCA
AAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCC
ATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAGACCTCATCATTAAAGCTTCC
CAAGTACTCTCTTTTGGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGC
TGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGC
CACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCTGTGGAACA
ACACAAACTACTCTTGTGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCC
TCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATC
AGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGC
CTTCAAGTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCTTG
ACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAG
CTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGTCGCCAGGGATCCGTCG

ACTTGACGCGTTaATATCAACAAGTTTGTACAAAAAAGCAGGCTACAAAGAGGCCAGCGGTT
CCGGACGGGCTGACGCATTGGACGATTTTGTATCTGGATATGCTGGGAAGTGACGCCCTCGAT
GATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGG
CAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTA ACTCTAGAAGTTCCGGATCTC
CGAAAAAGAAACGCAAAGTTGGTAGCCAGTACCTGCCCGACACCGACGACCGGCACCGGATC
GAGGAAAAGCGGAAGCGGACCTACGAGACATTC AAGAGCATCATGAAGAAGTCCCCCTTCAG
CGGCCCCACCGACCCTAGACCTCCACCTAGAAGAATCGCCGTGCCAGCAGATCCAGCGCCA
GCGTGCCAAAACCTGCCCCCCAGCCTTACCCCTTACCAGCAGCCTGAGCACCATCAACTAC
GACGAGTTCCTACCATGGTGTTCCTCCAGCGGCCAGATCTCTCAGGCCTCTGCTCTGGCTCC
AGCCCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCATGGTGTCTG
CACTGGCTCAGGCACCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCT
CCACCAGCCCCCTAAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCT
GCAGTTCGACGACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCA
CCGACCTGGCCAGCGTGGACAACAGCGAGTTCAGCAGCTGCTGAACCAGGGCATCCCTGTG
GCCCTCACACCACCGAGCCATGCTGATGGAATACCCGAGGCCATCACCCGGCTCGTGAC
AGGCGCTCAGAGGCCCTCCTGATCCAGCTCCTGCCCTCTGGGAGCACCAGGCCTGCCTAATG
GACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATGGATTTCTCAGCCTTGCTG
GGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAAGGGATGTTTTTGGCGAAGCCTGAGGCCGG
CTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAATCCGGCCAT
TTCATCCTCCAGGAAGTCCATGGGCCAACCGCCACTCCCCGCCAGCCTCGCACCAACACCA
ACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTCAGCCACTGGA
TCCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGA
GCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCT
GCAATCTGTGGCCAAATGGACCTTTCCTATCCGCCCCCAAGGGGCCATCTGGATGAGCTGAC
AACCACACTTGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAATTGA
ACGAGATTCTGGATACCTTCTTGAACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACA
GGACTGTCCATCTTCGACACATCTCTGTTT

>hU6 promoter

tttcccatgattccttcatatatttgcataatacagatacaaggctgtagagagataaattggaat
taatttgactgtaaacacaaagatattagtaaaaaatacgtgacgtagaaagtaataatttc
ttgggtagtttgcagttttaaattatgttttaaattggactatcatatgcttaccgtaact
tgaaagtatttcgatttcttggctttatataatcttgtggaaggacg

>*E. coli* sfGFP expression module

gaaagtgaaacgctgatttcatgcgtcattttgaaacattttgtaaactcttatttaataatgtg
tgcggaattcacatttaatttatgaatgttttcttaacatcgcggaactcaagaaacggc
aggttcggatcttagctactagagaaagaggagaaatactagatgcgtaaaggcgaagagct
gttcaactggtgctgcctcctattctggtggaactggatggatgtcaacggtcataagtttt
ccgtgcgtggcgaggggtgaagggtgacgcaactaatggtaaaactgacgctgaagttcatctgt
actactggtaaaactgccggttcttggcogactctggtaacgacgctgacttatggtgttca
gtgctttgctcgttatccggaccatatgaagcagcatgacttcttcaagtccgccatgccgg
aaggctatgtgcaggaacgcacgatttctttaaaggatgacggcagtcacaaaacgcgtgcg
gaagtgaaatttgaaggcgataccctggtaaacgcattgagctgaaaggcattgactttaa
agaggacggcaatatcctgggcccataagctggaatacaattttaacagccacaatgtttaca
tcaccgccgataaacaataaaatggcattaaagcgaatttttaaattcgccacaacgtggag
gatggcagcgtgcagctggctgatcactaccagcaaaacactccaatcggtgatggtcctgt
tctgctgccagacaatcactatctgagcagcaaacgcttctgtctaaagatccgaacgaga
aacgcgatcatatggttctgctggagttcgtaaacgcagcgggcatcacgcgatggtatggat
gaactgtacaaatgaccaggcatcaataaaacgaaaggctcagtcgaaagactgggccttt

cgttttatctgttgtttgtcgggtgaacgctctctactagagtcacactggctcaccttcggg
tgggcctttctgcgtttata

9.14 Reprint permissions

ELSEVIER LICENSE TERMS AND CONDITIONS

Apr 11, 2018

This Agreement between n/a -- Charles Motraghi ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4326020333345
License date	Apr 11, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Biotechnology
Licensed Content Title	Multivariate analysis of cell culture bioprocess data—Lactate consumption as process indicator
Licensed Content Author	Huong Le,Santosh Kabbur,Luciano Pollastrini,Ziran Sun,Keri Mills,Kevin Johnson,George Karypis,Wei-Shou Hu
Licensed Content Date	Dec 31, 2012
Licensed Content Volume	162
Licensed Content Issue	2-3
Licensed Content Pages	14
Start Page	210
End Page	223
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic

License Number	4431350698818
License date	Sep 17, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Biotechnology
Licensed Content Title	Considerations on the lactate consumption by CHO cells in the presence of galactose
Licensed Content Author	Claudia Altamirano,Andres Illanes,Silvana Becerra,Jordi Joan Cairó,Francesc Gòdia
Licensed Content Date	Oct 1, 2006
Licensed Content Volume	125
Licensed Content Issue	4
Licensed Content Pages	10
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1d.
Title of your thesis/dissertation	Developing a lactate inducible transgene expression system in Chinese Hamster Ovary cells
Expected completion date	Sep 2018
Estimated size (number of pages)	140
Requestor Location	n/a 34 Pontefract Road Bicester, Oxon OX26 1AP United Kingdom Attn: n/a
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Apr 11, 2018

This Agreement between n/a -- Charles Motraghi ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4326020333345
License date	Apr 11, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Biotechnology
Licensed Content Title	Multivariate analysis of cell culture bioprocess data—Lactate consumption as process indicator
Licensed Content Author	Huong Le,Santosh Kabbur,Luciano Pollastrini,Ziran Sun,Keri Mills,Kevin Johnson,George Karypis,Wei-Shou Hu
Licensed Content Date	Dec 31, 2012
Licensed Content Volume	162
Licensed Content Issue	2-3
Licensed Content Pages	14
Start Page	210
End Page	223
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic



Title: Metabolic engineering of CHO cells to alter lactate metabolism during fed-batch cultures

Author: Cécile Toussaint, Olivier Henry, Yves Durocher

Publication: Journal of Biotechnology

Publisher: Elsevier

Date: 10 January 2016

Crown copyright © 2015 Published by Elsevier B.V.

Logged in as:
Charles Motraghi
n/a

LOGOUT

Creative Commons Attribution-NonCommercial-No Derivatives License (CC BY NC ND)

This article is published under the terms of the [Creative Commons Attribution-NonCommercial-No Derivatives License \(CC BY NC ND\)](#).

For non-commercial purposes you may copy and distribute the article, use portions or extracts from the article in other works, and text or data mine the article, provided you do not alter or modify the article without permission from Elsevier. You may also create adaptations of the article for your own personal use only, but not distribute these to others. You must give appropriate credit to the original work, together with a link to the formal publication through the relevant DOI, and a link to the Creative Commons user license above. If changes are permitted, you must indicate if any changes are made but not in any way that suggests the licensor endorses you or your use of the work.

Permission is not required for this non-commercial use. For commercial use please continue to request permission via Rightslink.

BACK

CLOSE WINDOW

Copyright © 2018 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customer care@copyright.com

<p>Citation: Sladitschek HL, Neveu PA (2015) MXS-Chaining: A Highly Efficient Cloning Platform for Imaging and Flow Cytometry Approaches in Mammalian Systems. PLoS ONE 10(4): e0124958. https://doi.org/10.1371/journal.pone.0124958</p>
<p>Academic Editor: Srikumar P. Chellappan, H. Lee Moffitt Cancer Center & Research Institute, UNITED STATES</p>
<p>Received: January 13, 2015; Accepted: March 19, 2015; Published: April 24, 2015</p>
<p>Copyright: © 2015 Sladitschek, Neveu. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited</p>
<p>Data Availability: All relevant data are within the paper.</p>
<p>Funding: The study was funded by the European Molecular Biology Laboratory. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.</p>
<p>Competing interests: The authors have declared that no competing interests exist.</p>

AMERICAN
SOCIETY FOR
MICROBIOLOGY

Title: Regulation of L-Lactate
Utilization by the FadR-Type
Regulator LldR of
Corynebacterium glutamicum

Author: Tobias Georgi, Verena
Engels, Volker F. Wendisch

Publication: Journal of Bacteriology

Publisher: American Society for
Microbiology

Date: Jan 17, 2008

Copyright © 2008, American Society for Microbiology

LOGIN

If you're a **copyright.com**
user, you can login to
RightsLink using your
copyright.com credentials.
Already a **RightsLink user** or
want to [learn more?](#)

Permissions Request

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

AMERICAN
SOCIETY FOR
MICROBIOLOGY

Title: Lactate Utilization Is Regulated by the FadR-Type Regulator LldR in *Pseudomonas aeruginosa*

Author: Chao Gao, Chunhui Hu, Zhaojuan Zheng, Cuiqing Ma, Tianyi Jiang, Peipei Dou, Wen Zhang, Bin Che, Yujiao Wang, Min Lv, Ping Xu

Publication: Journal of Bacteriology

Publisher: American Society for Microbiology

Date: Apr 24, 2012

Copyright © 2012, American Society for Microbiology

LOGIN

If you're a **copyright.com user**, you can login to RightsLink using your copyright.com credentials.

Already a **RightsLink user** or want to [learn more?](#)

Permissions Request

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.