Engineering "'plug and play"' biosensors in *Escherichia coli*



THOMAS FOLLIARD

Oriel College
University of Oxford
Hilary Term 2017

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Abstract

ENGINEERING PLUG AND PLAY BIOSENSORS IN *Escherichia coli*

Thomas Folliard

Oriel College, University of Oxford

Submitted for the degree of Doctor of Philosophy, Hilary Term 2017

The ability of biological components to function reliably in different genetic contexts is one of the underlying fundamental principles in synthetic biology. Application of this principle allows for the development of distinct parts or bio-bricks and for the possibility of genetic abstraction. However, the sensitivity of genetic elements to changes in context can cause issues for this abstracted view of genetic elements as parts. We look at how this context sensitivity effects DNA and RNA based biosensors and we further investigate the ability of feedback to reduce noise in a biological device.

Riboswitches are structural genetic regulatory RNA elements that directly couple the sensing of small molecules to gene expression. They are able to sense a wide range of small molecules and in response regulate gene expression. We investigated how changes in genetic context effected riboswitch function and how these changes affected expression from a riboswitch and adaptation of a riboswitch to a novel downstream application. To overcome these contextual difficulties with riboswitches, we have designed and introduced a novel genetic element called a ribo-attenuator. This genetic element allows for predictable tuning, insulation from contextual changes and a reduction in expression variation. Ribo-attenuators allow riboswitches to be treated as a truly modular and tunable component, and thus increase their reliability for a wide range of applications.

Many biological processes use a form of feedback to achieve a highly robust and reliable output. Accurate control of a biological process is essential for critical functions in biology, from the cell cycle to proteome regulation. We design a tunable synthetic feedback network using a class of viral proteins called Integrases, which can flip the orientation of DNA segments in a digital manner. Excisionase feedback *closes the loop* in this system, creating a novel switch that provides a tunable and modular architecture for applications throughout synthetic biology and bio-manufacturing that require a highly robust and orthogonally controlled output. This system is highly orthogonal and evolutionary robust, and demonstrates a strong capability for regulating and reducing the variability in expression genes being transcribed under its control.

Declaration

The work in this thesis was undertaken in the Department of Biochemistry at the University of Oxford. Work was performed between September 2013 and January 2017. All the work in this thesis is my own unless stated and has not been submitted for a degree at this or any other university.

Acknowledgements

I would like to thank my supervisors as detailed below for providing me with the opportunities and freedom to explore my research area.

Through my time in the lab a number of people have been of great help and support. I would especially like to thank Barbara Mertins, Thomas Prescott, Karl Brune, Axel Nystrom, George Wadhams, Nick Delalez, Travis Bayer and Ciaran Kelly. I would like to thank Dr Pete Worden, director of NASA Ames research center (retired) for his support and appointment of Professor Lynn Rothschild as my supervisor.

I would like to thank the UK Space Agency (UKSA) and the European Space Agency (ESA) for considerable funding towards the Space studies program in 2014.

Finally I would like to thank my three supervisors:

Professor Judith P Armitge, Department of Biochemistry, University of Oxford
Professor Antonis Papachristodoulou, Department of Engineering, University of Oxford
Professor Lynn J Rothschild. National Aeronautics and Space Administration Ames, California

Contents

Abstract **i**

List of Figures **ix**

List of Tables **xii**

0.1 Contributions xiii

0.2 Publication list xiii

contributions **xiii**

1 Introduction **1**

1.1 Types of biosensor 2

1.1.1 Chemical biosensors 2

1.1.2 Whole-cell Biosensors 3

1.1.3 Limitations of whole cell biosensors 6

1.2 Biosensor output 6

1.2.1 Promoters 6

1.2.2 Inducible promoters 8

1.2.3 Constitutive promoters 10

1.2.4 Transcription termination 10

1.3 Translation 11

1.3.1 Ribosome Binding Site 11

1.3.2 Biofab RBS library 11

1.4 Biological outputs 12

1.4.1 Fluorescent proteins . . . 12
1.4.2 Other protein outputs: Colicins . . . 14
1.5 Alternative control of gene expression . . . 14
1.5.1 Flipping promoters and genes through Integrases . . . 15
1.5.2 Integrase families and mechanism . . . 16
1.5.3 Excisionases . . . 17
1.5.4 Novel Applications . . . 18
1.5.5 BXB1 phage . . . 18
1.5.6 TP910-1 phage . . . 19
1.5.7 Riboswitches . . . 19
1.6 Engineering principles . . . 24
1.6.1 Feedback . . . 24
1.7 Aims . . . 27

2 Materials and Methods **28**
2.1 Antibiotics . . . 28
2.2 Gibson Assembly . . . 28
2.3 Competent cells . . . 29
2.4 Plasmid purification . . . 30
2.5 Transformation of plasmids into competent cells . . . 30
2.6 Data collection . . . 30
2.6.1 Flow cytometry . . . 30
2.6.2 Plate reader . . . 31
2.6.3 Fluorescence microscopy . . . 31

3 Amplifying heavy metal transcriptional biosensors **33**
3.1 Aims . . . 33
3.2 Background . . . 34
3.3 Re-engineering transcriptional biosensors . . . 34

3.4 Mechanisms of heavy metal detection and toxicity 35
3.4.1 Arsenite 35
3.4.2 Copper 36
3.4.3 Zinc 36
3.5 Aim 1- testing a simple integrase amplifier 37
3.6 Materials and methods 37
3.6.1 Plasmids 37
3.6.2 Growth and induction 38
3.7 Results 39
3.7.1 Generation of a simple zinc responsive biosensor 39
3.7.2 Reproducing reported Integrase function 43
3.7.3 Characterisation of activation thresholds 44
3.7.4 Aim 2: To determine the context specific sensitivity of these parts to different promoter and trans-activating pairings . . 46
3.7.5 Folding analysis of the cis-repressing region 48
3.7.6 Integrating a zinc responsive biosensor 49
3.7.7 Integrating a copper biosensors 51
3.7.8 Integrating an arsenite biosensor 52
3.7.9 Growth in the presence of heavy metals 53
3.8 Discussion 55

4 Ribo-attenuators: novel elements for reliable and modular riboswitch engineering 57

4.1 Aims 58
4.2 Introduction 59
4.3 Materials and methods 63
4.3.1 Plasmids 63
4.3.2 Growth and induction 65
4.3.3 Data collection 65

4.3.4 Fluorescence microscopy 65
4.3.5 Cell disruption 66
4.3.6 Functional assay of ColE9 66
4.3.7 Preparation of strep-tagged OmpT 67
4.3.8 Model Methods 67
4.3.9 One-Component System 68
4.3.10 Two-Component System 69
4.3.11 Analysis 70
4.3.12 One-Component System 70
4.3.13 Two-Component System 72
4.4 Results 77
4.4.1 Change in open reading frame demonstrate sensitivity of 5 regulatory elements to secondary structure. 77
4.4.2 Design of ribo-attenuators 79
4.4.3 Ribo-attenuators can shift addA riboswitch induction response. 81
4.4.4 Ribo-attenuators can shift and amplify btuB riboswitch induction response. 84
4.4.5 Ribo-attenuators reduce uncertainty of riboswitch function . 86
4.5 Discussion 89

5 A synthetic viral feedback loop results in robust expression 94
5.1 Aims 95
5.2 Introduction 96
5.3 Methods 99
5.3.1 Plasmids 99
5.3.2 Cell culture and experimental conditions 99
5.3.3 Measurement and data analysis 100
5.3.4 Mathematical modelling 100

5.3.5 Simulation Setup and Parameters . . . 103
5.4 Results . . . 104
5.4.1 Assembly of the integrase/excisionase system . . . 104
5.4.2 System modelling and simulation . . . 105
5.4.3 Introduction of excisionase reduces the system's expression level and response time . . . 107
5.4.4 Synthetic feedback reduced output variation . . . 108
5.4.5 Intrinsic noise is reduced by the synthetic feedback loop . . 108
5.4.6 Downstream process is easily tuned to a predictable threshold 113
5.5 Discussion . . . 114

6 Conclusions **119**
6.1 Further work . . . 125
6.1.1 Chapter 3 . . . 125
6.1.2 Chapter 4 . . . 125
6.1.3 Chapter 5 . . . 126
6.2 Appendix . . . 126

Bibliography **128**

List of Figures

1.1 Schematic of a typical whole cell biosensor 4
1.2 The arabinose biosensor in its wild-type layout and re-factored to express GFP 5
1.3 Typical σ^{70} promoter schematic 7
1.4 Structures of Tetracycline and Anhydrotetracycline 9
1.5 Typical transcriptional terminator schematic 10
1.6 Biofab library of promoters and ribosome binding sites. 12
1.7 SFGFP and mKate2 emission and excitation spectra 13
1.8 Diversity in the recombinase family. 15
1.9 Typical integration and excision of a phage genome 16
1.10 Integrase mediated inversion 17
1.11 Mechanism for integrase binding site catalysis 18
1.12 Novel applications for phage integrases 19
1.13 addA riboswitch function 21
1.14 btuB riboswitch function 21
1.15 Trans-activating riboswitch function 23
1.16 Closing the loop around a process with feedback 25
1.17 Thesis overview 27

2.1 Gibson Assembly technique 29

3.1 Threshold proposal 37
3.2 Trans-activating riboswitch assembly 39
3.3 Preliminary biosensor amplification 40

3.4 Single cell populations for a regular and an amplified biosensor . . 42
3.5 Experimental validation of cis-repressed TP901-1 integrase 43
3.6 Dual induction of TP901-1 integrase with the L-rhamnose and the pTET promoters 44
3.7 Cross section as a function of aTc induction 46
3.8 Dual induction of TP901-1 integrase with the pTET and pRHA promoters 47
3.9 Cross section as a function of L-rhamnose and aTc induction 48
3.10 Folding analysis of the 5 region of cis-repressing mRNA transcribed from the tetracycline and L-rhamnose promoters 49
3.11 Dual induction of TP901-1 integrase with the pTET and zinc promoters 50
3.12 Dual induction of TP901-1 integrase with the pTET and zinc promoters 51
3.13 Introduction of a copper biosensor to the ribo-regulators 52
3.14 Introduction of an arsenate biosensor to the ribo-regulators 53
3.15 Growth curves showing toxic effect of heavy metal presence 54

4.1 Typical activating riboswitch function 60
4.2 Ribo-attenuator context and schematic 63
4.3 Ribo-attenuator secondary structures 64
4.4 addA and btuB riboswitches exhibit contextual sensitivity 78
4.5 Effect of Ribo-attenuators repression 80
4.6 Additional ribo-attenuator screening data 81
4.7 Ribo-attenuators allow addA riboswitch rational re-engineering . . 82
4.8 Immunoblot of strep-tagged OmpT 83
4.9 sfGFP size analysis 83
4.10 Ribo-attenuators allow btuB riboswitch rational re-engineering . . 85
4.11 Full single cell analysis for all Ribo-attenuators 86

4.12 Ribo-attenuators reduce riboswitch noise 88
4.13 Single cell comparison between similar expressing ribo-attenuators and the fused constructs 89
4.14 Single cell distinction between induced and un-induced riboswitches 90

5.1 The schematic architecture of the open-loop and closed-loop expression systems developed in this chapter 106
5.2 Block diagram for the modelled feedback system. 107
5.3 Steady state responses after for the open- and closed-loop systems over a range of aTc inputs 109
5.4 Steady state is achieved more quickly and robustly in the presence of feedback 110
5.5 Time-course fluorescence data for the open- and closed-loop systems induced with 50 ng/mL aTc over 12 hours 111
5.6 Single-cell flow cytometry data for triplicates of both the open- and closed-loop systems at a range of inducer concentrations 112
5.7 Effect of an excisionase mutant or removal of the SSRA tag on the feedback system 113
5.8 Output response from the closed loop system is predictably tunable 114

6.1 Thesis chapters overview 122
6.2 Thesis conclusion overview 123
6.3 Internal competing hairpins could stabilise smaller hairpins 126

List of Tables

1.1 Examples of detectable ligands and their associated detection limit 3

2.1 Concentration of antibiotics used 28

3.1 Strains and plasmids 38

4.1 List of strains and plasmids used 75

4.2 List of ribo-attenuators 76

5.1 Simulation Parameter Values 104

6.1 Biosensor engineering approaches - advantages and disadvantages 124

6.2 List of primers 127

0.1 Contributions

I am grateful to the Collins lab, MIT for RR12 plasmid used in chapter 3 and the modeling support recieved from Dr Thomas Prescott and Mr Harrison Steel in chapters 4 and 5.

0.2 Publication list

This thesis contains material from two submitted journal articles currently in review as detailed below.

- ”Ribo-attenuators: novel elements for reliable and modular riboswitch engineering”’ was submitted to Nature Scientific Reports on the 8^{th} of March 2017 (in press).
- ”A synthetic viral feedback loop results in robust expression” was submitted to ACS Synthetic Biology on the 20^{th} of December 2016 (currently in press)

Chapter 1

Introduction

In 1978, Szybalksi and Skalka remarked "The work on restriction nucleases not only permits us easily to construct recombinant DNA molecules and to analyse individual genes, but also has led us into the new era of 'synthetic biology' where not only existing genes are described and analysed but also new gene arrangements can be constructed and evaluated" [155]. Despite this and many amazing advances in biotechnology the engineering of synthetic biological systems is still a difficult, expensive and unreliable process. There are many possible reasons for these difficulties. For example it may be that many biological systems are far too complex to reliably engineer or that we simply do not know enough about a given biological system to engineer it. Whatever the cause, moving towards a cheap reliably engineerable system is likely to yield a wide range of benefits.

A major step towards realising many of these applications is the ability to reliably sense a wide range of relevant inputs. A major category of biological components that facilitates this control are biosensors. Biosensors are a broad term defined by the The International Union of Pure and Applied Chemistry as, "a device which is capable of providing specific quantitative or semi-quantitative analytical information using a biological recognition element which is retained in direct spatial

contact with a transducer element" [158]. They have been applied in a myriad of uses, from the real time analysis of *in vivo* metabolic processes [117] to the quantification of glucose in a medical assay for diabetes [32]. The first biosensor was reported in 1975 [38].

1.1 Types of biosensor

Biosensors belong to a broad range of biological component having many types and subtypes, ranging from electro-chemical to molecular. The element in common between all biosensors is that the recognition component has a high specificity for the target in question. Some rely on immobilized enzymes for the monitoring of bio-reactor glucose levels while others, such as immunosensors rely on the high affinity of antibodies to detect the presence of a molecule of interest. Some biosensors utilise whole cells as detectors, adapting the natural ability of the biology to detect a wide range of molecules while others utilise a specific component such as RNA *in vitro* to fulfil this function. Each type has different advantages making it more applicable in specific situations. For example RNA based biosensors allow for a significantly quicker response time than whole cell biosensors which require a further element of transcription [102].

1.1.1 Chemical biosensors

A large class of biosensors are the chemical class. This class relies on the output of a chemical reaction to detect a molecule and generate a signal. A few key biomarkers, their relevance and detection limits are shown in Table 1.1.

Ligand	Relevance	Detection limit	Reference
Cancer antigen-125	Ovarian cancer	0.5 U mL^{-1}	[162]
Epidermal growth factor receptor	Cancer	1 U mL^{-1}	[178]
Insulin	Diabetes	174 ng mL^{-1}	[181]
P53	Cancer	100 ng mL^{-1}	[179]
Prostate specific antigen	Prostate cancer	$<$0.0009 ng mL^{-1}	[182]

Table 1.1: Examples of detectable ligands and their associated detection limit

1.1.2 Whole-cell Biosensors

Typically, as shown in Figure 1.1, a whole cell biosensor is a cellular biosensor made up of parts that fill a detection, a regulator and an expression role. The detection part, while in Figure 1.1 is shown as a membrane protein, can also be filled by different sensing parts such as transcription factors. The detection part utilises a response regulator to alter the transcription rate of a promoter. The output is usually a non-toxic protein that is easy to assay for example a fluorescent protein. Detector parts usually are comprised of a transcription factor element and optionally, if possible, a membrane transporter to aid the influx of the molecule of interest. Biosensors fall into one of two broad categories, digital or analogue. Digital biosensors respond in a step like fashion giving an "'all or none response"'. Typically, digital biosensors control the response to a life or death situation such as the pTET promoter which controls the expression of the TetA transporter which allows for the efflux of the antibiotic tetracycline. Analogue biosensors give a much more graded response and allow for a more finely tuned response.

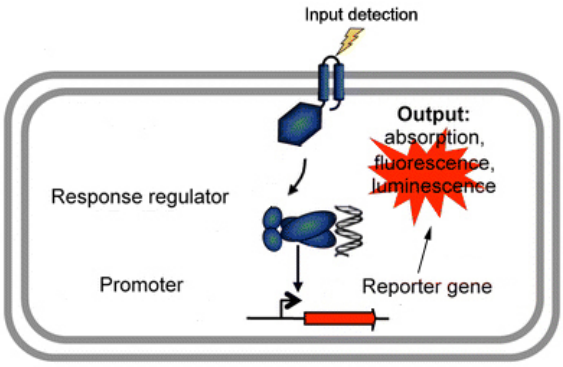


Figure 1.1: **Schematic of a typical whole cell biosensor**. A typical biosensor showing an input being sensed passing to a response regulator which regulates a promoter. Typically this promoter controls an output that is easy to assay such as a fluorescent protein. Adapted from [171]

An example of this is the L-arabinose biosensor shown in Figure 1.2. As seen in Figure 1.2A the biosensor is comprised of two divergent promoters, one controlling transcription of the AraC transcriptional regulator while the other pBAD controls transcription of the *araBAD* operon. The membrane transporter araE is a proton symporter with a low affinity for arabinose of 140-320 μM [34] which imports arabinose [95]. Transcription of *araE* and the *araBAD* operon is induced by arabinose [147]. In the wild type system the expression of the *araBAD* operon allows for arabinose to be catabolised to xylulose-5-phosphate then via the pentose phosphate pathway [141]. Refactoring the biosensor, placing a new gene of interest in place of the *araBAD* operon as shown in Figure 1.2B allows for inducible gene expression of a new gene of interest through a non-toxic and cheap inducer.

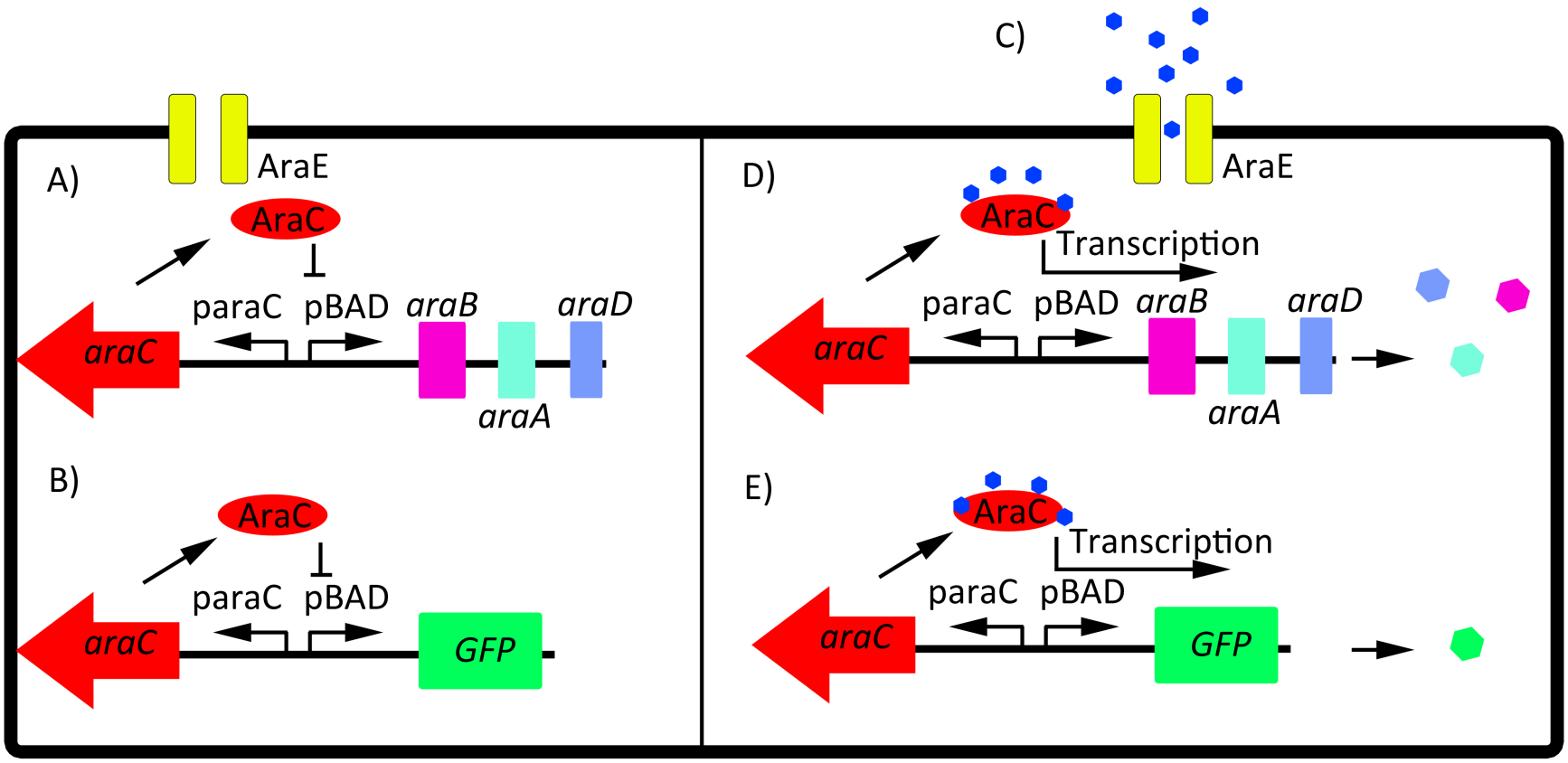


Figure 1.2: **The Arabinose biosensor in its wild-type layout and re-factored to express GFP.** A) in the absence of L-arabinose araC represses transcription from the pBAD promoter repressing expression of the BAD operon. B) in the case of the refactored biosensor, the transcription of GFP is repressed. C) L-arabinose enters the cell through araE, the arabinose transporter. D) AraC is bound by L-arabinose and up-regulates transcription from the pBAD promoter transcribing *araB*, *araA* and *araD* as a multi-cistronic operon. E) The refactored biosensor transcribes mRNA coding for GFP.

1.1.3 Limitations of whole cell biosensors

While there are many applications for whole cell biosensors there are some limiting aspects preventing their wider application. The presence of membrane transporters can change the sensitivity to a desired ligand by selectively transporting the ligand across the membrane resulting in variable sensitivity between different *Escherichia coli* strains and often the ligand needs to be present for the organism to grow [127]. Further to this, many naturally occurring biosensors such as the copper biosensor reported in *Ravikumar et al* [124] give an analogue response to the presence of a ligand, this limits their application where a digital response above a threshold is necessary.

1.2 Biosensor output

Transcription is the first stage of gene expression during which DNA is transcribed to RNA. The region of DNA transcribed is controlled by the presence of flanking genetic elements, promoters that initiate transcription and terminators that terminate transcription.

1.2.1 Promoters

A promoter is the genetic element that initiates the transcription of a gene or an operon, it can function in a constitutive fashion or in response to an input. A typical *E. coli* constitutive promoter is shown in Figure 1.3. Promoters recruit RNA polymerase mediated by the presence of a sigma factor. These sigma (σ) factors are small proteins needed for the initiation of transcription. The binding of an RNA polymerase to a promoter is mediated by one of these transcription initiation factors. There are several σ factors used by bacteria with the most com-

mon being σ^{70} which promotes transcription of most genes in growing cells [64]. Others control more specific responses such as σ^{32} which controls the heat shock genes such as DNA repair enzymes [106]. The sigma subunit of the transcription complex recognises consensus sequences in what is called the -10 and -35 box for σ^{70}. During transcription initiation the transcription complex unwinds the DNA forming an open promoter complex. The RNA polymerase then elongates the RNA template, releasing the sigma subunit.

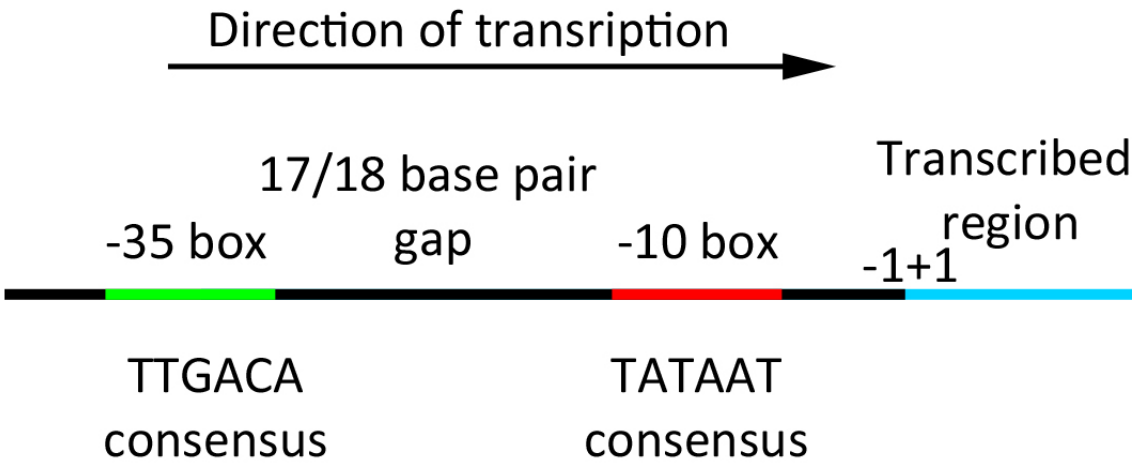


Figure 1.3: **Typical promoter σ^{70} schematic** The layout of a typical constitutive *E. coli* promoter showing the promoter region in black with -10 and -35 polymerase recognition sites with relevant consensus sequences. plus 1 and minus 1 transcriptional start sites are indicated along with the transcribed region. Made in photoshop.

Promoter strength is affected by several factors from the distance of the recognition elements to the up and down stream sequences [137]. For example the rRNA promoter for *rrnB P1* up stream sequence has a 30 to 70 fold increase on transcription levels [137]. These elements interact to affect the affinity of the promoter for the sigma factor and polymerase complex.

1.2.2 Inducible promoters

This class of promoter responds to the presence of a molecule by up regulating or down-regulating transcription.

Tetracycline promoter pTET

In its natural system the tetracycline promoter (pTET) drives expression of *tetA*, an efflux pump that removes the antibiotic tetracycline from the cell conferring resistance [113]. pTET is the promoter negatively regulated by the Tet-repressor, this 23 Kda protein functions as a homo-dimer to repress transcription. It consists of two DNA binding domains and a regulatory core [122]. The DNA binding domains consist of two Helix-Turn-Helix domains which bind a 15 base-pair palindromic sequence of the tetA operator in the absence of tetracycline [113]. This repression is alleviated through the addition of Anhydrotetracycline (aTc). Anhydrotetracycline is used as it gives a 35 fold higher induction response than tetracycline with 100 fold less antibiotic effect [36]. Recognition of aTc by the ribosome (in its antibiotic role) and by the tet-repressor appears to be different as a ketone group is substituted by a hydroxyl group in position 11 [36] as shown in Figure 1.4.

Arabinose promoter pBAD

In its natural system the arabinose promoter drives expression of the araBAD operon allowing for the catabolism of arabinose in *E. coli* in the presence of arabinose and absence of glucose [74]. The promoter is tightly regulated by the AraC transcription factor, a 32 Kda protein with a dual function as a repressor in the absense of arabinose and an activator in its presence. When AraC binds to Arabinose a conformation change is induced reducing the affinity of AraC for the

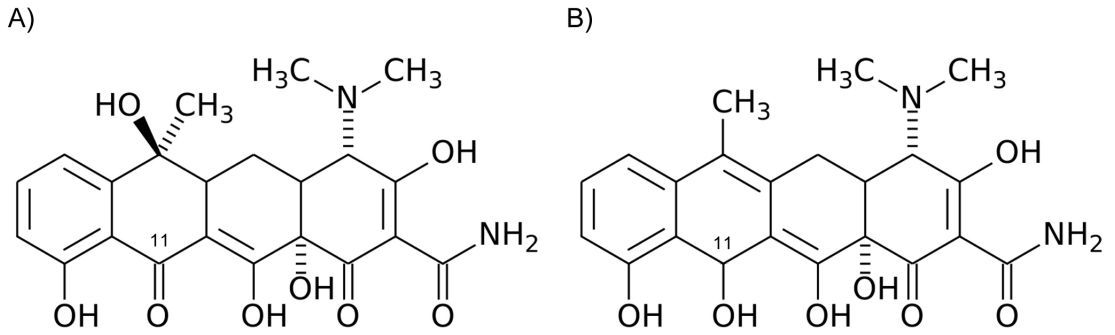


Figure 1.4: **structures of Tetracycline and Anhydrotetracycline** The difference in structure between A) Tetracycline and B) Anhydrotetracycline. The substitution of the ketone group at position 11 in Tetracycline with a hydroxyl group is thought to impair antibiotic effect while increasing affinity for the tet repressor by 35 fold [36] Made in photoshop.

pBAD promoter preventing repression [74]. However AraC bound with arabinose also functions as a transcriptional activator promoting transcription from the pBAD promoter by inducing a DNA conformation favourable for transcription initiation [86]. The pBAD promoter is characterised as an "'all-or-none"' promoter meaning it does not respond to its inducer in a linear fashion [78].

Rhamnose sensitive promoter

The Rhamnose promoter from *E. coli* is similar to the pBAD promoter in that in it allows the cell to catalyse an alternative carbon source. It is induced in the presence of L-rhamnose and in its natural system expresses the rhaBAD operon. Unlike arabinose promoter the rhamnose promoter responds to L-rhamnose in a graded fashion [159]. The pRHA promoter is used as a graded promoter because its graded response begins at concentrations where all the cells are not induced [123].

1.2.3 Constitutive promoters

Biofab promoter library

A key problem in the implementation of biological parts is predictability. The Biofab library [105] provides standard transcription and translation parts across over a thousand fold range between the highest and lowest responder in the form of defined constitutive promoters and bi-cistronic Ribosome Binding Sites. The promoter library was made by random mutagenesis to the -34 or -10 transcription boxes to alter the transcription initiation efficiency [105].

1.2.4 Transcription termination

The process of transcription is halted at genetic elements called terminators. Their structure and efficiency varies greatly but terminators commonly have a strong intrinsic hairpin followed by a poly-Uracil region as shown in Figure 1.5. This hairpin is thought to stall the polymerase and the poly-U region disassociates from the template due to its low number of hydrogen bonds. The length of the poly-U region directly effects termination efficiency [35].

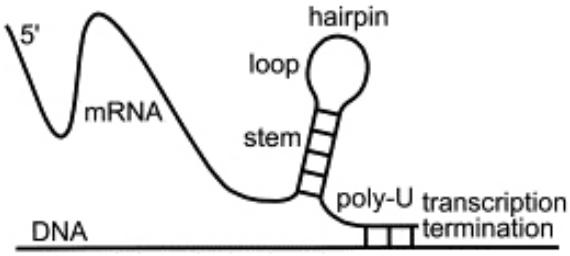


Figure 1.5: **Typical transcriptional terminator schematic** The layout of a typical transcriptional terminator showing the hairpin consisting of a stem and loop followed by a poly-U region. Taken from [44]

1.3 Translation

Translation is the process through which mRNA is translated to protein. This process has several stages. First the 30S subunit of the ribosome binds to the Ribosome Binding Site (RBS). This subunit then tracks up and down the mRNA until it encounters a start codon (AUG, GTG, TTG) [112]. After finding this start codon the 30s subunit is thought to pause, allowing for the recruitment of the 50s large ribosomal subunit.

1.3.1 Ribosome Binding Site

The Ribosome binding site (RBS) consists of a series of nucleotides containing a core of 6 nucleotides upstream of a start codon (AUG, GUG or UUG - translating at different efficiencies of 1:2:6 [112]) in *E. coli* . The core of the RBS is usually complementary to the anti-Shine-Dalgarno core in the 30s small ribosomal subunit. The strength of an RBS relies on several factors, how close the core is to the standard Shine-Dalgarno (AGGAGG) [151], if there is any competing secondary structure over the core or the start codon and which of the three start codons are used [105].

1.3.2 Biofab RBS library

The Bi-cistronic RBS library was made by random mutagenesis of the RBS core starting with the Shine-Dalgarno for the strongest variant [105]. These Bi-cistronic RBS' work through the utilisation of two RBS'. One is present at the 5′ end of the leader peptide and controls transcription of a short leader peptide of about 13 amino acids. This peptide contains the second RBS which controls translation of the gene of interest. The principle is that passage of the ribosome from RBS1

removes secondary structure around RBS 2 allowing for reliable expression of almost any down stream open reading frame. See Figure 1.6.

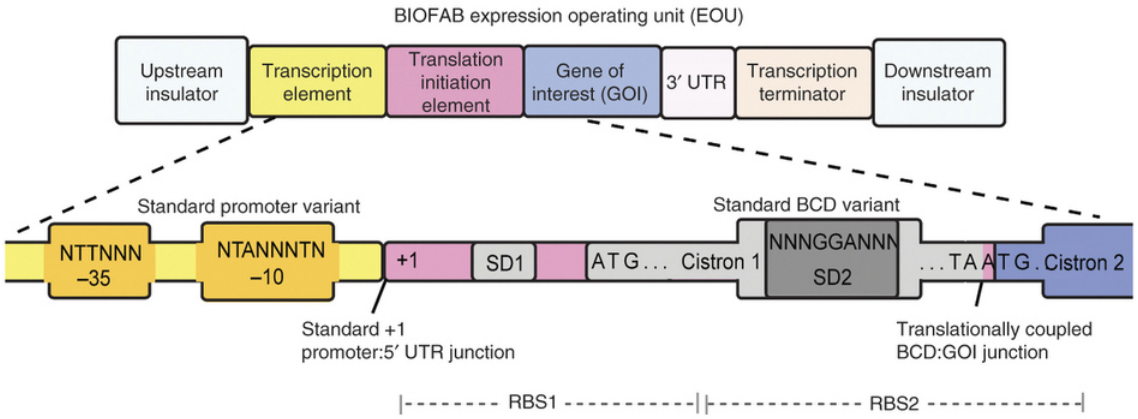


Figure 1.6: **Biofab library of promoters and Ribosome Binding Sites.** Schematic showing the Biofab expression operating unit with upstream insulator, transcription element, translation initiation element and gene of interest. The translation initiation element consists of two cistrons followed by a translationally coupled junction and the gene of interest. SD1 represents the first Shine-Dalgarno sequence while SD2 represents the second Shine-Dalgarno variant in the bi-cistron. Taken from [105]

1.4 Biological outputs

Every biological circuit has one or more outputs. These outputs can be functional in the sense of undertaking a role in the cell such as AraE which serves as an arabinose importer [34] or they can be informatory and serve to report the state of the cell or the circuit such as a fluorescent protein. In this work I expressed different proteins with engineered markers either as antibody tags or fluorescence.

1.4.1 Fluorescent proteins

Fluorescence is the term used to describe the absorption of light at a specific wavelength and its emission at another- it is very common especially among marine jellyfish species. A gene encoding a fluorescent protein was identified in the

1962 in *Aequorea victoria* [150]. This discovery made for major advances in a wide range of fields such as developmental biology and microscopy as they provide a relatively direct link between the expression of a gene and a visible local output although these applications did not become apparent until the cloning of the GFP gene had been reported in 1992 [121]. In this thesis we use Superfolder GFP as a reporter, this reliably folding variant has a peak excitation at 480nm and a peak emission at 520nm, see Figure 1.7 [118].

Conventional GFP will often misfold when fused to difficult proteins [161] therefore Superfolder GFP was developed to allow researchers to express a GFP that reliably folded in *E. coli* was created through multiple rounds of random mutagenesis followed by Fluorescence Activated Cell Sorting (FACS) [118].

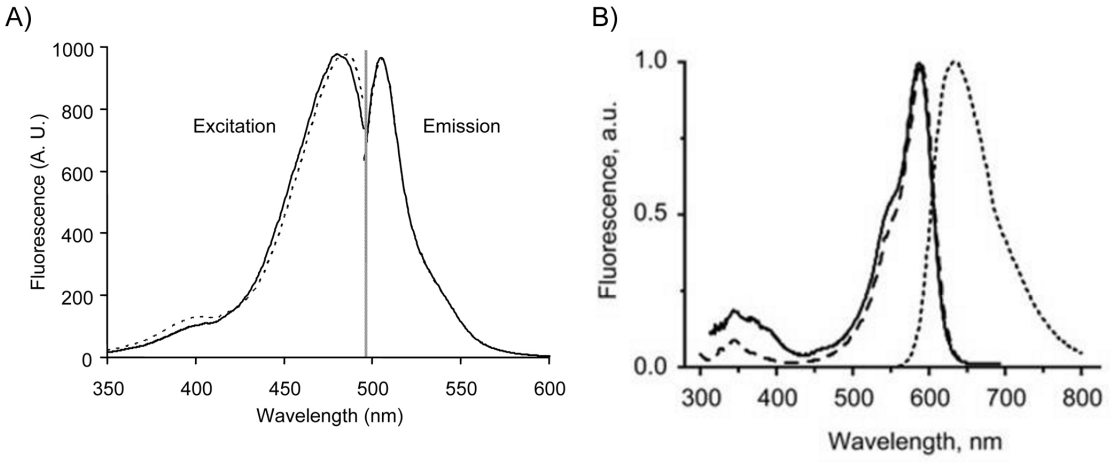


Figure 1.7: **SFGFP and MKate2 emission and excitation spectra. A) Fluorescence excitation and emission spectra for eGFP (dotted line) and superfolder GFP (solid line). B) Fluorescence excitation and emission spectra for mKate2. Solid line indicates normalised mKate2 absorption, ( ) indicates fluorescence excitation and (- - -) indicates emission spectra** Taken from [118, 146]

The RFP used in this thesis is MKate2 [145]. This far red fluorescent protein has been shown to have little spectral overlap with Superfolder GFP [12] with an excitation and emission of 588 and 620nm respectively. Measuring Optical Density at 600nm while using a red fluorescent protein can give inflated cell density due to absorbance by the RFP leading to an underestimation of per-cell fluorescence

by up to 10% [65]. As such OD measurements when an MKate2 was used were taken at OD_{700}.

1.4.2 Other protein outputs: Colicins

To overcome some problems of fluorescent measurement we used the expression of a colicin as a functional output. Colicins are secreted protein antibiotics that are part of the *E. coli* stress response to kill other bacteria. They have generated much interest due to their high specificity for bacteria and action at low concentration [67] Depending on their mechanism some are lytic such as the ColE1 which is a pore forming agent cleaving peptidoglycan in the cell wall. Some are nucleases, instead killing the cell through nuclease action [67]. Depending on their import pathway colicins are classed into one of two groups. Group A pass from the periplasm into the cytoplasm via the Tol complex while group B pass through the Ton complex [116]. Usually colicins are induced via the SOS pathway in response to stress but the mechanism through which the colicin is released into the environment is poorly understood. In order to avoid cell suicide all colicins are produced with a small (10 kDa) immunity protein which binds the C terminal active domain rendering the colicin harmless [82].

ColE9 is a type B colicin that binds the btuB and OmpF receptor in the outer membrane then binds the Tol proteins to translocate from the periplasm into the cytoplasm where it has a DNAse activity [67].

1.5 Alternative control of gene expression

Many circuits function through the conventional direct transcription-translation of a reporter gene however there are more complicated examples of ways to con-

trol gene expression.

1.5.1 Flipping promoters and genes through Integrases

Integrases and their associated excisionases are a class of viral proteins that mediate integration and excision of the phage genome during the lysogenic and lytic phases of the phage life cycle [153, 15, 77, 71]. They were first discovered in 1981 with the first use of *cre* recombinase [154]. Integrases have been characterised into two distinct families, the tyrosine and serine recombinases which can be separated into several distinct smaller families, Figure 1.8. The two families of integrases are defined by the conserved amino acid in the catalytic domain of the protein in question, tyrosine or serine. Integrases can catalyse one of several reactions; integration, excision or inversion [13].

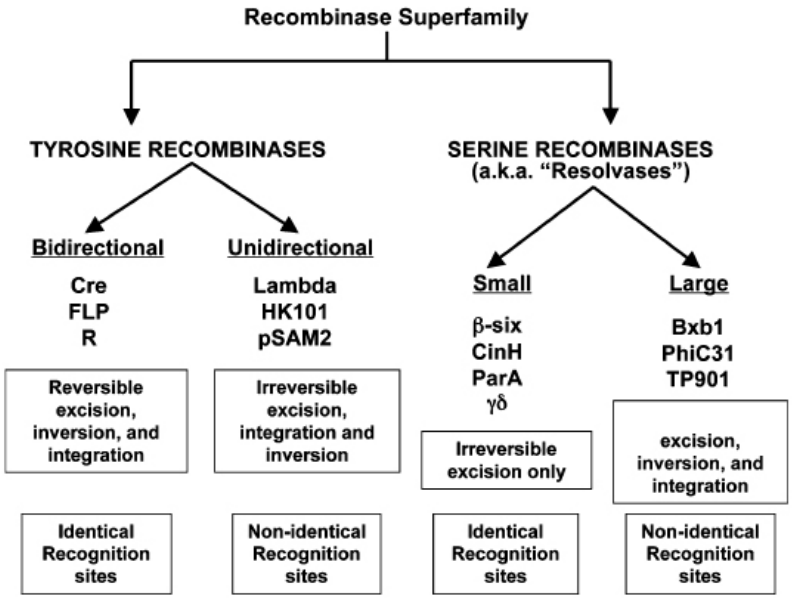


Figure 1.8: **Diversity in the recombinase family.** Taken from [165]

Integrases in their natural environment allow phage to integrate their genome into the host cell and enter the lysogenic phase [25]. They catalyse the insertion of DNA through the recognition of specific c.50 base pair long binding sites referred to as attB and attP (attachment Bacteria and attachment Phage) as shown

in Figure 1.9.

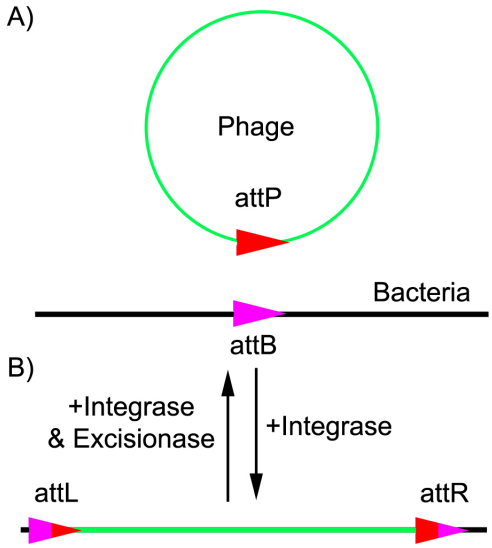


Figure 1.9: **Integration and excision of a phage into a bacterial genome** A) a phage genome containing attP (attachment Phage) integrates into the bacterial genome at the sequence recognised by the viral integrase called attB (attachment Bacteria) by a viral integrase. B) Expression of a Integrase and an Excisionase allow for excision of the phage genomic DNA.

1.5.2 Integrase families and mechanism

Integrases belonging to many phages such as TP901-1, Bxb1 and ϕC31 recognise orthogonal attB and attP recognition sites via a helix-turn-helix domain [103]. The orientation of the recognition sites determines whether the recognition sites are excised, inverted or integrated together. If both sites are present on different DNA strands the integrase will integrate the two strands together as seen in Figure 1.9, if the sites are present in the inverted orientation on the same DNA strand then the Integrase will invert the sequence between the two binding sites as shown in Figure 1.10, if the sites are on the same DNA strand and not inverted then an excision event will occur.

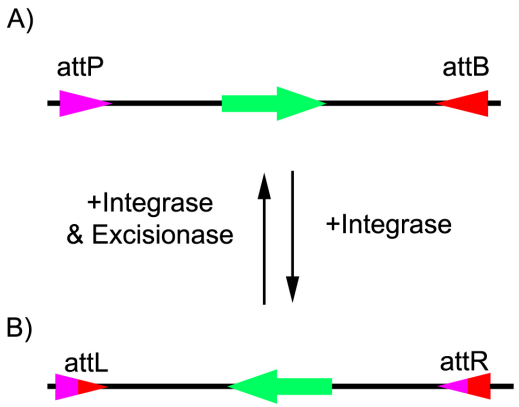


Figure 1.10: **Integrase mediated inversion** A) Inverted attP and attB sites on the same strand of DNA lead to inversion of the region between both sites. B) expression of integrase and its associated excisonase leads to re-inversion and a return to the previous orientation.

The process of integration is undertaken via strand cleavage, inversion and ligation to attL and attR (attachment Left and attachment Right) sites. This reaction is catalysed by a central conserved catalytic domain of the integrase [62]. The underlying process of this reaction was proposed by *Biswas et al* [10] is shown in Figure 1.11 suggesting that the formation of a Holliday junction as an intermediate step in lambda integrase. In the early stage the integrase has bound to both recognition sites and cleaved part of the DNA, the cleaved strands of DNA are not yet exchanged. The formation of a Holliday like junction within the integrase leads to the integrase domains changing conformation prior to a strand inversion exchange.

1.5.3 Excisionases

Excisionases, in conjunction with integrases, allow phages to enter the lytic phase of the life cycle [8]. Passage from lysogenic to lytic phase is usually controlled by a high gain bistable switch that is biased towards expression of integrase and excisionase in the lytic phase [73]. This is usually controlled by the cleavage or phos-

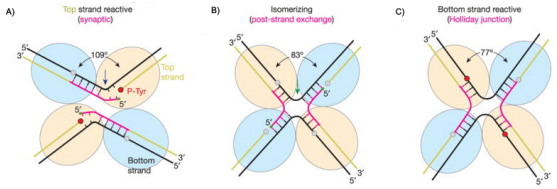


Figure 1.11: **Mechanism for integrase binding site catalysis** A) Conserved residues cleaves the DNA strand at a conserved tyrosine (p-Tyr) residue. B) A Holliday junction is firmed between the two DNA strands. C) Further conformational change uncatalysed unidirectional sites lead to excision of the intergenic region, excisionase and integrase expression allows for re-integration. Blue circles represent one integrase subunit, red DNA strands represent integrase recognition site. Taken from [165]

phorylation of a transcription factor signalling the start of the lytic phase [53]. The expression of an excisionase cofactor allows for the integrase-excisionase complex to recognise attL and attR sites and subsequently catalyse the reverse reaction back to attP and attB [140, 54].

1.5.4 Novel Applications

The properties of integrases and excisionases has led to a particular focus on their characterisation and application in the field of synthetic biology. Researchers in MIT and Stanford developed Boolean Integrase Logic (BiL) gates which confer the full range of logic at the genetic level (XOR, AND, OR, XNOR etc) by inverting or removing directional elements such as terminators or promoters [13, 152]. This has led to integrases being applied in a wide range of areas, such as highly sensitive medical assays in healthcare and memory storage [177, 32, 12, 47] as well as metabolic pathway assembly [30] as shown in Figure 1.12.

1.5.5 BXB1 phage

Mycobacteriophage Bxb1 is a temperate phage of *Mycobacterium smegmatis* [100]. The BXB1 genome is 50,550 bp in length coding for 86 genes. Regulation is con-

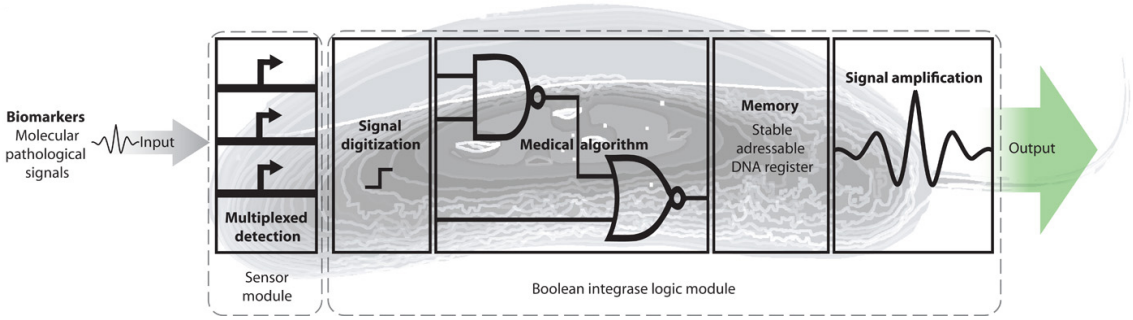


Figure 1.12: **Novel applications for phage integrases** Phage integrases can be employed with a sensor module to provide signal digitisation passed through a logic operation. the output of which is amplified before being read. Taken from [32]

ducted through 13 bp asymetric sequences located within the short intergenic regions that the phage repressor binds [79]. BXB1 integrase has been shown to work reliably under control of the pTET promoter and has orthogonality with TP901 integrase [13].

1.5.6 TP910-1 phage

Lactococcal phage TP901-1 is also a temperate phage with a genome of 37667 bp encoding 56 genes [18, 29]. TP901-1 integrase has been shown to function orthogonally from BXB1 phage and is functionally expressed under control of the pBAD promoter. [13].

1.5.7 Riboswitches

Riboswitches are a class of structural regulatory biosensors mostly present in the 5 UTR of messenger RNA. Since their first discovery in 2002 as RNA-based sensors of vitamin derivatives [52], riboswitches have become recognised as wide spread and important elements in the control of gene expression. They usually allow for the regulation of a downstream gene or operon in response to the bind-

ing of small molecules such as cellular metabolites or metal ions [63, 98, 142, 164]. The existence of riboswitches in all domains of life lends support to the theory that life first existed as RNA before protein [14]. This theory requires that functions critical to life performed by proteins could be performed by RNA alone. It is suggested that riboswitches represent the sensing and regulatory part of this hypothesis. [14].

Riboswitches that regulate translation do so through the allosteric effects of small molecules binding to their aptamer domain. This structural rearrangement allows for the formation or removal of a transcriptional terminator or more commonly changes the accessibility of a ribosome binding site (RBS) core [133, 125, 28]. Here we look at three examples of riboswitches, one repressive, one activating and one synthetic based around short non-coding RNAs and their interactions.

addA riboswitch

The Adenine riboswitch is found in the 5′ UTR of a gene encoding adenine deaminase of *Vibrio vulnificus*. This riboswitch is an activating riboswitch that in the absence of adenine sequesters away the part of the mRNA coding the RBS and the start codon preventing translation. Upon binding of a suitable ligand such as 2-aminopurine the RBS is freed and translation can occur [98] as shown in Figure 1.13.

btuB riboswitch

The btuB riboswitch is present in many species of bacteria. It is a repressive riboswitch and it controls expression of cobalamin (B_{12}) related genes. The btuB riboswitch is an example of a repressive riboswitch where it represses translation

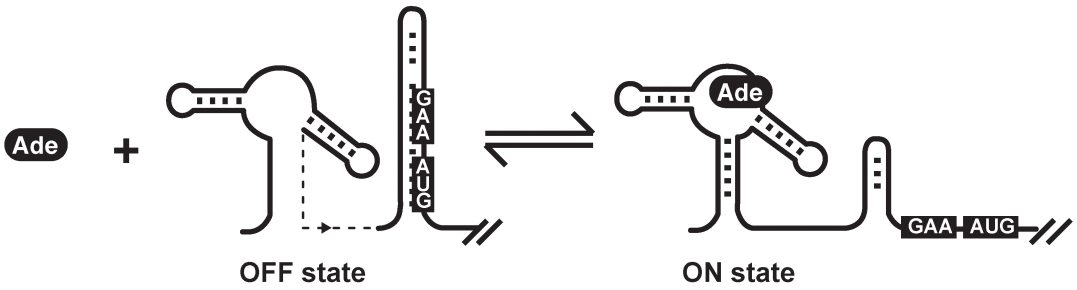


Figure 1.13: **addA riboswitch function** Schematic of addA riboswitch function showing the formation of a repressing hairpin in the absence of a ligand. A complimentary region of the repressor stem forms, preventing translation. Binding of the ligand to the binding site changes the conformation displacing the repressor stem and exposing the RBS allowing translation. Taken from [132]

in the presence of its ligand but promotes translation in the absence of the ligand. It achieves this through an internal kissing-loop that promotes secondary structure across the RBS and start codon in the presence of a ligand as shown in Figure 1.14. Binding of a ligand displaces this kissing loop allowing for the formation of strong internal hairpins exposing the RBS and start codon, this kissing loop has been shown to be pivotal for coordinating riboswitch conformational changes by providing a binding site for the ligand [93].

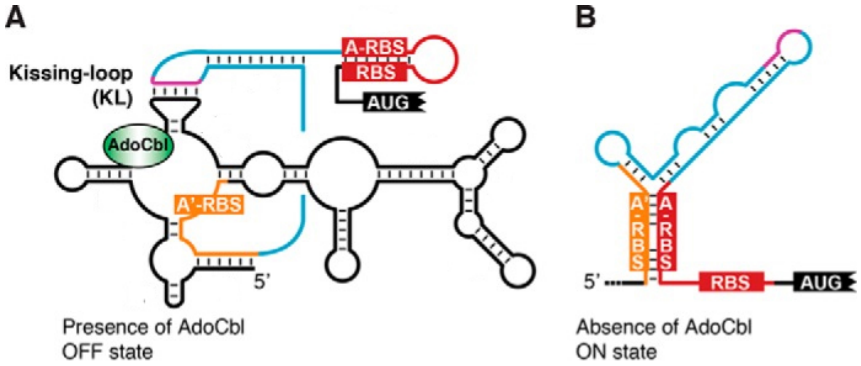


Figure 1.14: **btuB riboswitch function** Schematic of the btuB riboswitch function from *E. coli* showing the formation of a kissing-loop (blue) in the presence of a ligand inducing the presence of a silencing hairpin over the RBS (red) preventing translation. Absence of an inducer results in the formation of an internal hairpin competing out the previous hairpin silencing the RBS, allowing for ribosome recruitment. adapted from [93]

Trans-activating synthetic riboswitch

Trans-activating riboswitches represent a direct link between transcriptional and translation control. They function through the interaction of two non-coding RNAs, one a cis-repressing region is present in the 5′ UTR of the down stream gene. It forms a strong internal hairpin across the RBS- a common occurrence in natural riboswitches. However this internal hairpin is not removed by the binding of a small molecule, instead it is displaced by the interaction of a second trans-activating RNA which bears strong sequence complementarity to the hairpin. Binding of this trans-activating RNA competes with the internal hairpin of the cis-repressing UTR exposing the RBS, allowing for the translation of the downstream gene [70]. This riboswitch has the advantage that it can directly couple to transcriptional signals into a single translational signal bearing the unique advantage of the easy tuning of a transcriptional response from an independent signals shown in Figure 1.15.

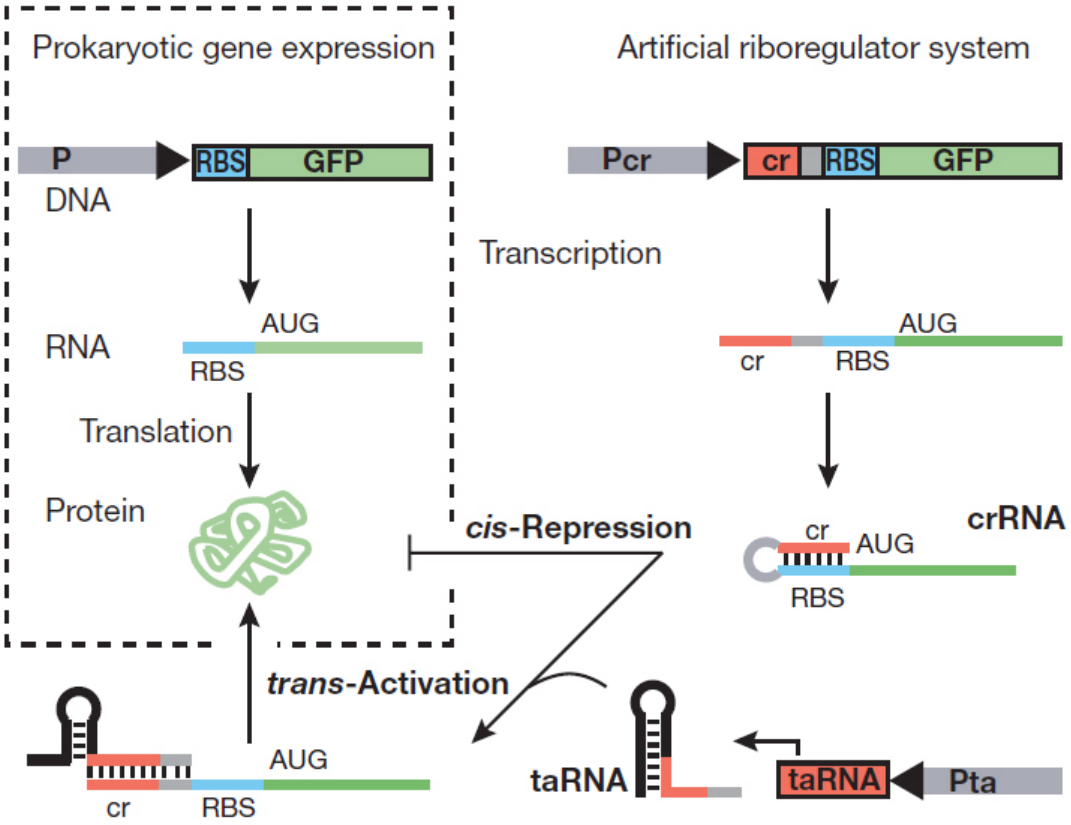


Figure 1.15: **Trans-activating riboswitch function** Schematic of the Riboregulator riboswitch function showing conventional prokaryotic gene expression in the dotted box producing a protein of interest. The artificial riboregulator system relies first on the cis-repressed region transcribed from promoter Pcr, the cis repression prevents translation of the protein, crRNA. The trans-activating RNA (taRNA) is transcribed by Pta. Both these RNA molecules form a complex trans-activating, making available the RBS allowing for translation of the gene of interest. Taken from [70]

Synthetic applications

While many riboswitches evolved to control the expression of enzymes in natural systems, synthetic riboswitches have been adapted from natural ones to respond to specific molecules such as the adaption of the addA riboswitch to respond to a non-natural ligand [134, 169, 167, 39, 91]. Synthetic riboswitches responding to novel molecules could allow for rapid and high throughput screening of mutant

libraries to find mid stages in bio-manufacturing or allow for new highly sensitive assays currently not available. These studies along with recent advances towards the in silico design of riboswitches [163, 5, 45] represent significant scientific interest in the application of riboswitches.

Riboswitch limitations

Despite much promise and many potential applications for riboswitches there remain few applied examples. This is due to a limited and often poor induction range, a difficulty adapting the controlled gene to novel purposes and high variation within and between populations. In chapter 4 we demonstrate these limitations in further detail and look at an approach to mediating these limitations.

1.6 Engineering principles

Synthetic Biology is often thought of as the application of engineering principles to biology. Engineering principles rely on the design of robust and reproducible parts or abstraction. However the limitations of these engineering principles often exist in synthetic biology due to our lack of understanding of the biological system as a whole.

1.6.1 Feedback

Aström *et al* [2] explain Feedback control as referring to ”a situation in which two (or more) dynamical systems are connected together such that each system influences the other and their dynamics are thus strongly coupled”. Taking an open system and inferring its output as the input allows the loop to be closed

around a process. As shown in Figure 1.16 in a conventional open loop system an input (r) passes into system one giving an output (u) which passes into system two which gives the final outout (y). A biological example of this is the expression of GFP from the tetracycline promoter pTET. In a closed loop system as seen in 1.16B the output (y) is fed back into system one as an input. In biological terms this can be represented by co-expressing GFP and the tetracycline repressor from pTET.

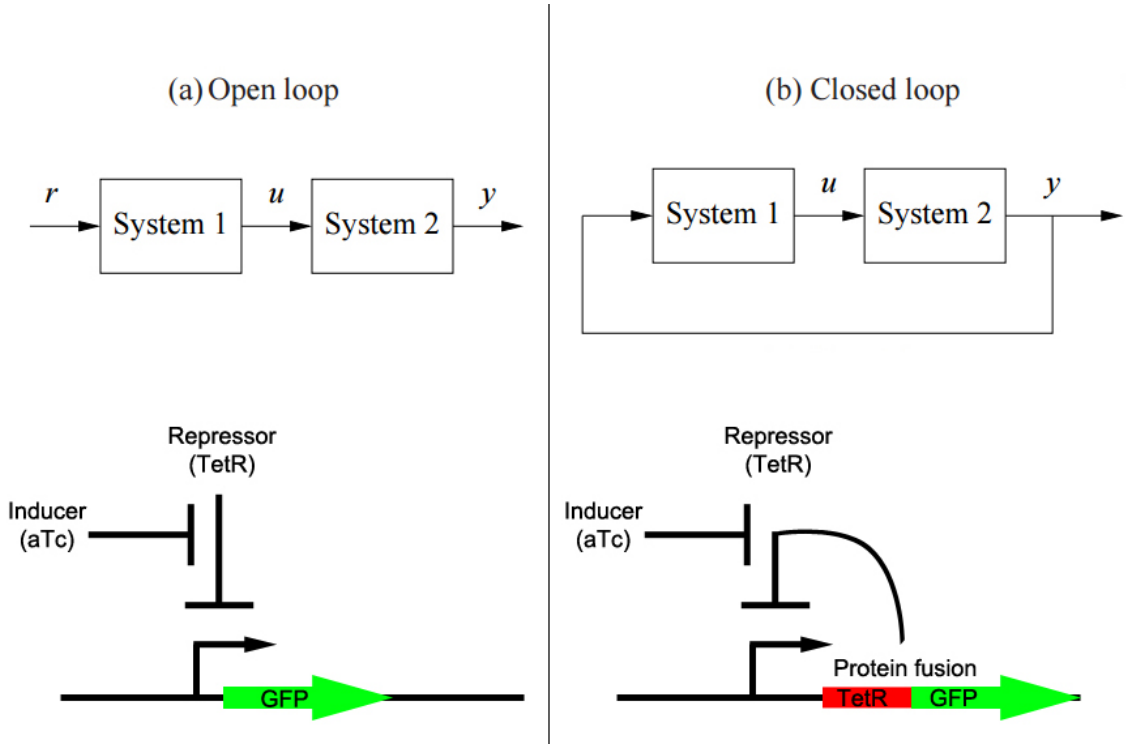


Figure 1.16: **Closing the loop around a process with feedback** A) An open loop system where there is no connection between the output of system 2 (y) and the input of system 1 (r) . B) A closed loop system where the output of system 2 (y) becomes the input of system 1. Adapted from [2]

Feedback has many interesting properties. Feedback can make the system more resilient to external influences, it can also create linear behaviour out of non-linear components. Generally feedback allows a system to be insensitive to variations in internal elements or external effects. Feedback also has several disadvantages in that it can create oscillations or unwanted noise in a systems [2].

Biological origins of feedback

The complexity common in biological systems often arises from the interactions of smaller systems with reaction to a multitude of external stimuli. As a result of this complexity and unpredictability, feedback is a widely observed biological phenomenon. It is implemented from sub-cellular systems in gene control for example in many cellular differentiation pathways [72] or signal transduction to multi-cellular systems in the form of immunological and cardiovascular systems [87]. Frequently, in biological systems, positive feedback is used to create a digital, switch like behaviour through gene auto-regulation and to create osculations such as in the circadian rhythm [19] or the cell cycle [41].

Applications

Feedback has wide reaching applications in synthetic systems with its ability to reduce variation between and within populations, its ability to provide a faster response time and its ability to easily fine tune accurate expression. Some of the theorised applications include; maintaining expression of a metabolic pathway to within a narrow range in a bioreactor where variation in environmental conditions can cause variation in expression [46], allowing for the faster detection of contaminants through biosensors [2] and the tuning of components of a enzymatic metabolic pathway to increase yield [180].

Feedback limitations

Despite these potential applications there have been very few applied uses of feedback in synthetic circuits. This is thought to be due to a low absolute expression rate, low modularity and high burden.

In Chapter 5 I look at these limitations in further detail.

1.7 Aims

In this thesis I look at biosensors where complexity and a lack of modularity leads to challenges in their application. As described in Figure 1.17 I first look at how traditional transcription factor based biosensors could benefit from threshold tuning, I then look at making riboswitches a tunable and more reliable device and I look at how one could reduce system noise around a biological process that required more steady and reliable expression.

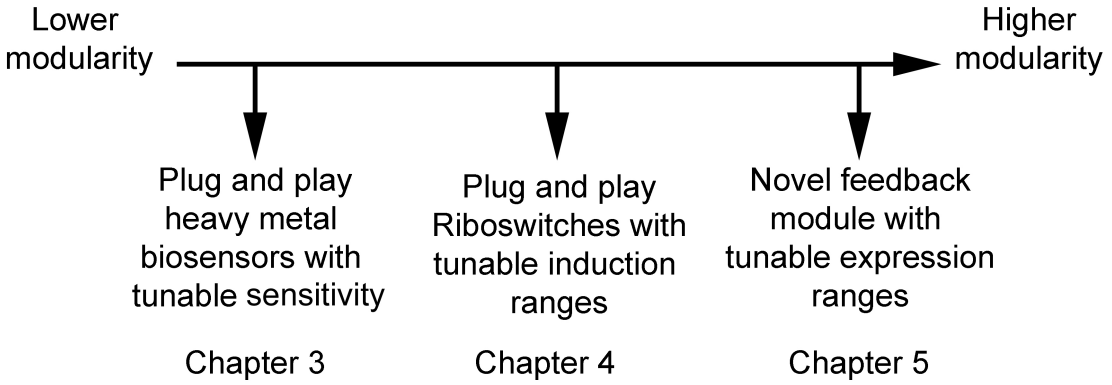


Figure 1.17: **Thesis overview** Higher modularity of biosensors is discussed in terms of heavy metal biosensors, riboswitches and feedback modules.

Chapter 2

Materials and Methods

2.1 Antibiotics

Plasmids were maintained by antibiotics shown in Table 2.1.

Table 2.1: Concentration of antibiotics used

Chloramphenicol	25 μg/mL
Ampicillin	100 μg/mL
Kanamycin	25 μg/mL

2.2 Gibson Assembly

All plasmids were assembled using Gibson Assembly [57, 56] employing a 20 base pair overlap. Gibson assembly does not use specific restriction sites; instead it relies on specific sequence overlaps as shown in Figure 2.1. This approach allows for the construction of small DNA molecules up to several hundred kilobases in a quick and reliable fashion. Gibson assembly relies on complementary overlaps [57, 55].

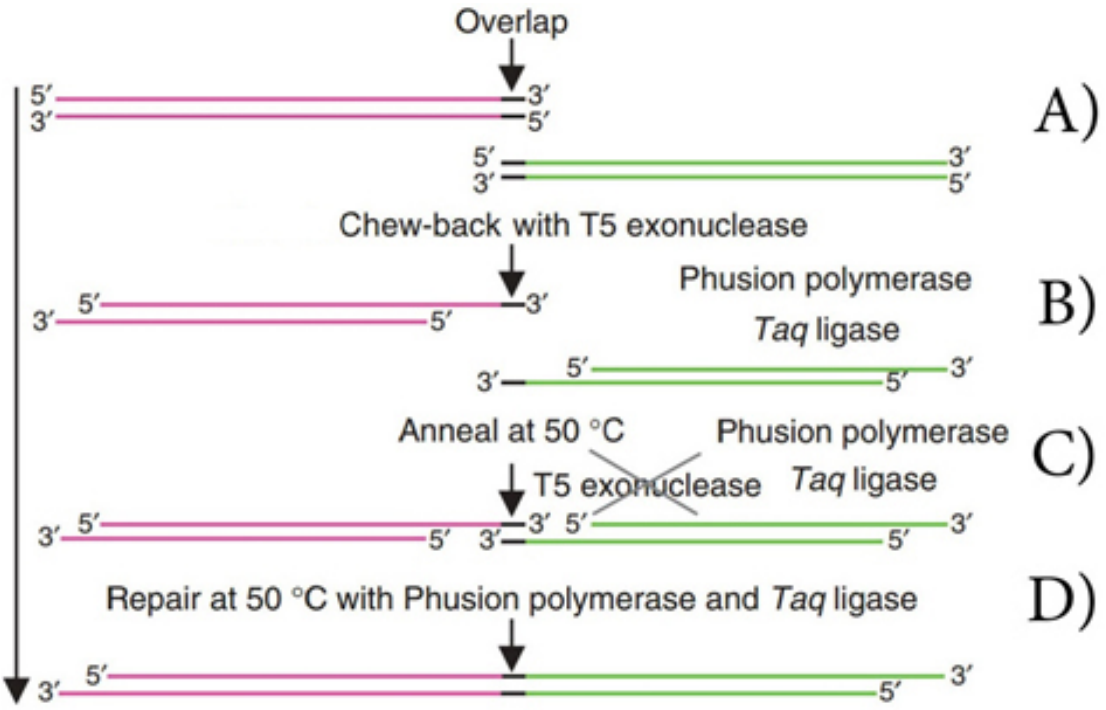


Figure 2.1: **Gibson Assembly technique** A) Overlaps exist between sequence parts, these overlaps occur in a sequential manner allowing ordered assembly. B) The 5 end is digested by an exonuclease, this exposes the overlaps. C) These overlaps anneal to each other at 50 ^{o}C, this temperature deactivates the exonuclease preventing too much 5 digestion. D) A polymerase and a ligase then polymerase and ligate the single stranded DNA. Taken from [57]

2.3 Competent cells

Competent cells were prepared in accordance with the Calcium Chloride method [3], cells were grown to an OD_{600} of 0.6 to 0.8 then spun down at 3500 G for 10 minutes before being resuspended in 0.1M calcium chloride and incubated for 30 minutes on ice. Cells were then re-pelleted as previously before being re-suspended in 0.1M Calcium Chloride with 25% V/V glycerol. Competent cells were stored in 100 μl aliquots at -80°C .

2.4 Plasmid purification

Overnight 5 ml cultures of *E. coli* containing the plasmid of interest maintained under appropriate antibiotic selection were grown overnight at 37°C . The plasmid was then extracted using the QIAGEN plasmid mini prep kit as per the manufacturer instructions. Plasmids were eluted in 50 μl of water and sequence confirmed.

2.5 Transformation of plasmids into competent cells

Plasmids derived from a miniprep were transformed by the 5 minute transformation protocol [120], 100ng of plasmid was incubated with 100 μl of comptetent cells for 5 minutes on ice before being spread on LB-Agar with appropriate antibiotics. Plasmids that were derived from Gibson assembly were incubated with 100μl of prepared competent cells on ice for 30 minutes. The cells were then heat shocked for 30 seconds at 42°C before being returned to ice for 2 minutes. Subsequently the cells were recovered in 1 ml of SOC media for 1 hour before being plated on appropriate antibiotics.

2.6 Data collection

2.6.1 Flow cytometry

Flow cytometry acquisition was performed at Department of Biochemistry, University of Oxford using a Attune flow cytometer (Lifetech Scientific, Basingstoke, UK) for each data point after induction with appropriate inducer or inducers, 200 μl of each culture was centrifuged at 3500 G, washed with 500 μl of Phosphate-

buffered saline (PBS) and resuspended in 500 μL of PBS before 50,000 events were measured. Cells were diluted 1:1000 in PBS. All acquisitions were performed using the same machine with the same settings (FSC 560V, SSC 420V; GFP laser 360V; RFP laser 400V). Flow cytometry data were analysed using the FlowJo software (Treestar inc, Ashland, USA).

2.6.2 Plate reader

Single measurements of culture density and florescence was determined using a ClarioStar platereader (BMG Labtech, Ortenberg, Germany). OD_{600} measurements were taken in clear well plates and GFP measurements in black well plates (both Greiner bioscience). Continuous GFP and OD_{600} measurements were taken in clear bottom, black well plates (Greiner bioscience) GFP was quantified by normalising by OD_{600}. Optical density was estimated by:

$$OD = -log_{10}\frac{I}{I_o}.$$

Where I_o is the intensity of light directed at the sample and I is the intensity of light that reaches the detector after passing through the sample [83]. It is assumed that the dominant source of light scattering is caused by cells [90] GFP measurements were made at an excitation of 480nm and an emission of 520nm. The gain was 1000.

2.6.3 Fluorescence microscopy

Thin, flat agarose pads (1% agarose in Milli-Q water, w/v) were generated on microscope slides. Two microliters of cells were added to the pad immobilising them on the surface of the agarose. Cells were then imaged with a Nikon Eclipse

Ti microscope, 100X phase contrast objective (Nikon), GFP filter set (Chroma), Andor iXON CCD camera. Images were acquired using NIS Elements software (Nikon).

Chapter 3

Amplifying heavy metal transcriptional biosensors

3.1 Aims

Whole-cell biosensors have seen wide application in synthetic biology, often driving expression of a fluorescence reporter in response to a concentration of a specific molecule. However a biosensors sensitivity to a ligand is often very difficult to adapt to a different desired range although expression of membrane transporters to enhance uptake or over expression of relevant transcription factors have shown some ability to modulate sensitivity. The aims of this Chapter are:

1. To test the ability of TP901-1 integrase to amplify the weak signal from heavy metal biosensors and to use the trans-activating riboswitch to modulate this sensitivity.
2. to determine the context specific sensitivity of these parts to different promoter and Trans-activating pairings.

3.2 Background

In both developing and developed countries, environmental pollution by heavy metals is a serious issue [156]. Typically, environmental levels of metal pollutants are quantified by methods such as mass spectrometry which require extensive sample preparation and are therefore time and resource intensive [107]. As a result there has been significant scientific interest in developing a rapid, high throughput and cheap way of monitoring environmental metal levels. One approach is to refactor bacterial metal responsive operons to control a reporter gene's expression in the presence of a metal of interest. For example a zinc and copper biosensor has been developed from the ZntA-ZntR operon [124].

3.3 Re-engineering transcriptional biosensors

Many naturally occurring biosensors give an analogue induction response which is often unsuitable for many applications where a more distinct and digital response is called for. One possible cause of this is the focus on purely transcriptionally controlled biosensors. Combining elements of transcriptional biosensors with their robust response with the tunability of RNA based biosensors along with the amplification of viral proteins could yield a widely applicable, tunable, plug and play approach to biosensing. Traditional transcription factor biosensors rely on the presence of a responsive transcription factor that, in response to the signal of interest, gives a transcription induction or repression response [97]. Such whole-cell biosensors are usually adapted from inducible operons that allow the cell to respond selectively to the presence of, for example a toxin. These whole-cell biosensors have drawn significant interest due to their sensitivity and cost, for example a whole-cell methanol biosensor developed by *Wen et al* [166] demonstrated a whole-cell biosensor able to detect low levels of methanol in a

comparable fashion to gas chromatography with a sensitivity of 0.05 to 2.5 mM. The ability to detect a wide range of environmental containments such as copper, urainium, aresnite, toxins and pathogens [104, 88, 66], their ability for high throughput detection, cheap storage and low cost production make whole cell biosensors attractive alternatives from other more conventional techniques such as mass spectrometry.

3.4 Mechanisms of heavy metal detection and toxicity

Over the course of this chapter I describe a class of metal sensitive biosensors. These biosensors all originate from the *E. coli* genome and are used to control the expression of survival genes for growth in the presence of toxic metal concentrations.

3.4.1 Arsenite

The mechanisms of arsenite toxicity are not well understood although it is thought that arsenic could interact with critical thiols in proteins and inhibit their activity [68]. The *arsRBC* operon in *E. coli* is an auto-regulated, arsenite responsive operon which allows for resistance to arsenical and antimonite metals [37, 23] by controlling the expression of an arsenite export system [23]. The first protein encoded in the operon -ArsR- is a repressor for the operon. It has a helix-turn-helix domain, a metal binding site and a dimerisation domain. It binds arsenite at an inducer binding site via three cystine residues [148].

3.4.2 Copper

Excess copper is toxic to aerobic bacteria [110], it is thought that this toxicity occurs when membrane bound copper catalyses the formation of peroxide free radicals [135]. Toxic levels that impair growth for *E. coli* have been shown to be between 4 and 6 mM [61]. *E. coli* has multiple mechanisms to mediate environmental copper toxicity from the translocating ATPase CopA to the multi-copper oxidase CueO [126]. The copper responsive promoter used in this chapter is the promoter of the CUS operon in *E. coli* . The proximal gene encodes CusC which is an outer membrane lipoprotein involved in the detoxification of copper and silver ions in *E. coli* . It forms a part of the CusCFBA copper/silver efflux system. pCUS is regulated by CusR (CU sensing regulator). This gene regulates genes relative to the copper and silver efflux response [104]. It recognises the consensus sequence GTAAA(N)(4-8)GTAAA [176].

3.4.3 Zinc

While zinc is required for the stability of many functional proteins [6] concentration in excess of 0.5mM have been found to inhibit growth as zinc ions block the initial step of respiration by reacting with active residues of bacterial dehydrogenase [76]. *E. coli* has 2 known zinc exporters, ZntA [4] and ZitB [60], to maintain intracellular concentration below toxic levels.

The zinc responsive promoter of ZntA in its natural context controls expression of a P-type ATPase -named due to a short attachment of inorganic phosphate at the aspartate residues at the time of activation- called ZNTA which is involved in the efflux of Pb(II), Cd(II) and Zn(II) [128, 9]. The promoter is positively regulated by the zinc transport Regulator (ZntR) which binds as a dimer to an inverted repeat sequence of 22 base pairs located between the -10 and -35 boxes [114, 17].

3.5 Aim 1- testing a simple integrase amplifier

In this chapter I use a previously characterised integrase amplifier [152] to adapt biosensors from an analog output to a very large, distinct and digital response. I look at varying the digital threshold using a previously described Trans-activating riboswitch as shown in Figure 3.1 with the intention on demonstrating a digital response from a biosensor above a variable threshold.

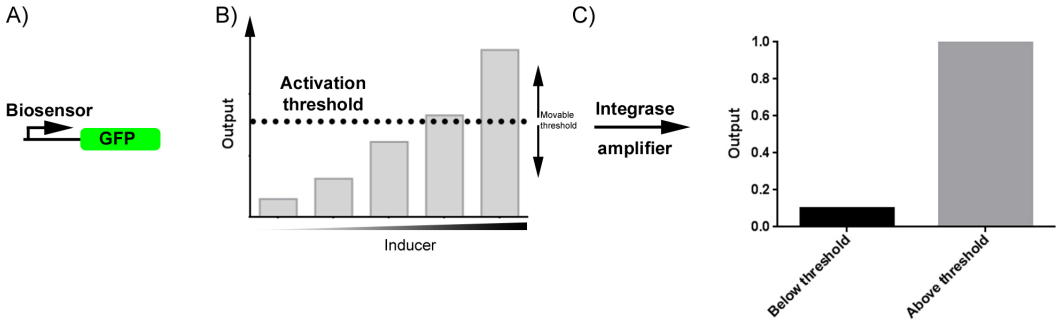


Figure 3.1: **Biosensor amplification above a variable threshold** A) A typical schematic for a biosensor expressing GFP that yields a linear response shown in B) A desirable characteristic is to produce a strong, distinct signal when expression crosses above a variable threshold. C) The output should be a digital OFF/ON response

3.6 Materials and methods

3.6.1 Plasmids

All cassettes that placed GFP under control of the biosensor directly were cloned into the pBADHisA plasmid [70] (ColE1 origin; 50-60 copies, Kanamycin resistance). All biosensors were amplified from the *E. coli* genome and assembled by Gibson assembly [57, 55] and then sequence confirmed. The integrase target plasmid pSB3A1 contained RFP flanked by TP901 attP and attB binding sites in the inverted direction to a P7 constitutive promoter (P15A origin, 15-20 copies).

A full list of plasmids, are available in Table 3.1 and the assembly strategy is described in Figure 3.2.

Table 3.1: Strains and plasmids

Plasmid	Description	Reference or Source
pSB3A1	Target plasmid, expresses RFP	This study
pSB50K2	pBAD:Trans-activating,pTET:Cis-repressing:TP901	This study
pSB50K3	pRHA:Trans-activating,pTET:Cis-repressing:RFP	This study
pSB50K4	pTET:Trans-activating, pRHA:Cis-repressing:RFP	This study
pSB50K5	pZNTA:Trans-activating, pTET:Cis-repressing:RFP	This study
pSB50K6	pARSR:Trans-activating, pTET:Cis-repressing:RFP	This study
pSB50K7	pCUSC:Trans-activating, pTET:Cis-repressing:RFP	This study
pSB50K8	pZNTa:RFP	This study
pSB50K9	pCUSC:RFP	This study
pSB50K10	pARSR:RFP	This study

3.6.2 Growth and induction

All experiments were performed in *E. coli* MG1655. For each experiment three freshly transformed colonies were inoculated in LB medium (10g/L tryptone, 5g/L yeast extract, 10g/L NaCl) with stated antibiotics and grown overnight. Then diluted 1:100 into 1 mL LB in a deep well plate (Greiner bio-one) with fresh antibiotics and stated concentrations of heavy metal inducer. Cultures were induced for 12 hours at 37 °C with shaking before data collection. Biological triplicates were chosen to ensure adequate analysis of the standard deviation around the mean. For the two plasmid system antibiotics used were Kanamycin (25μg/mL) and Ampicillin (100μg/mL), for the maintenance of the single plasmid system only Kanamycin (25μg/mL) was used.

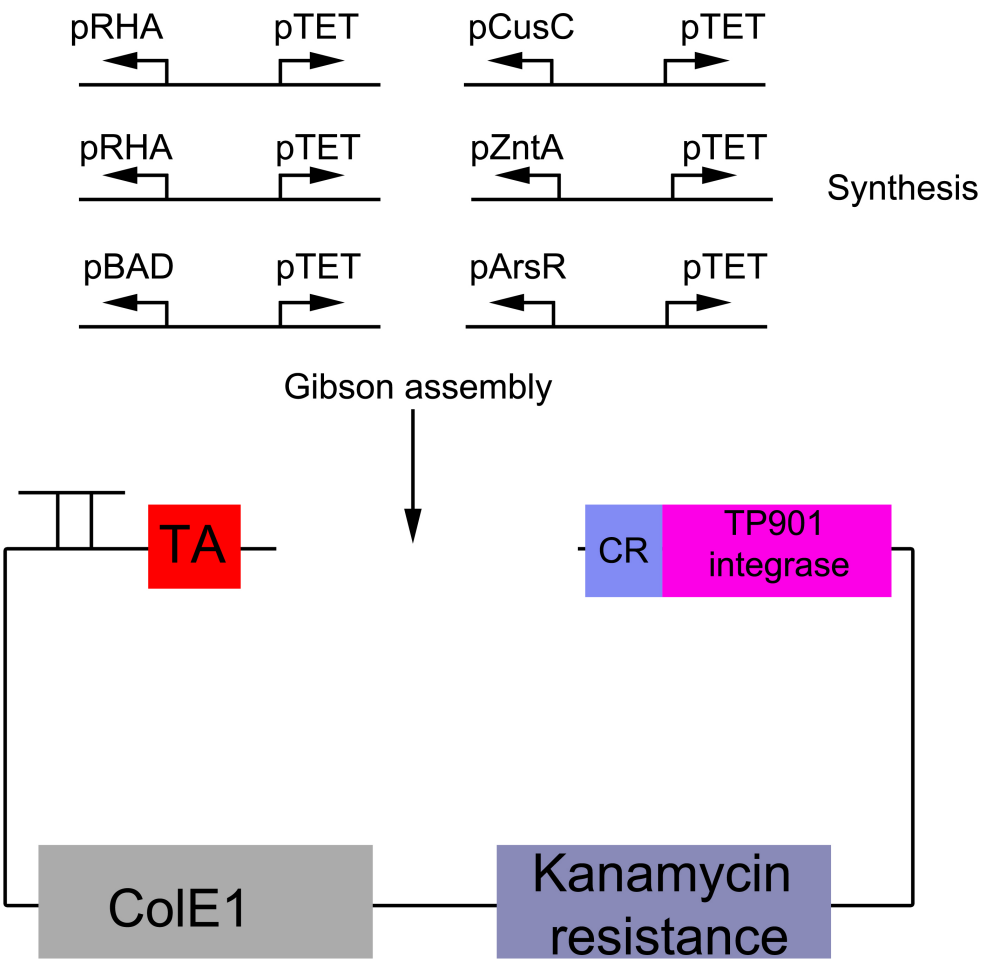


Figure 3.2: **Trans-activating riboswitch assembly** Gibson assembly of divergent promoter region controlling the trans-activating riboswitch. Promoter regions were synthesised by IDT and assembled into the pBADHisA backbone. TA represents the Trans-activating RNA and CR represents the Cis-repressed region controlling TP901 integrase translation.

3.7 Results

3.7.1 Generation of a simple zinc responsive biosensor

Two proposed approaches are shown in Figure 3.3A, the zinc responsive promoter [175] directly drives the expression of RFP. while in Figure 3.3C the RFP was replaced with TP901-1 integrase [15]. This integrase inverts the same fluorescent reporter bringing it under control of the P7 constitutive promoter from

the biofab collection [105]. The output of the simple zinc biosensor has a similar low basal expression in the absence of any inducer compared to the integrase amplified biosensor (Mean of 573 vs 482 RFP/OD) but the output is significantly higher in the presence of inducer when an integrase amplifier is employed yielding a greater than 10 fold increase in fluorescence.

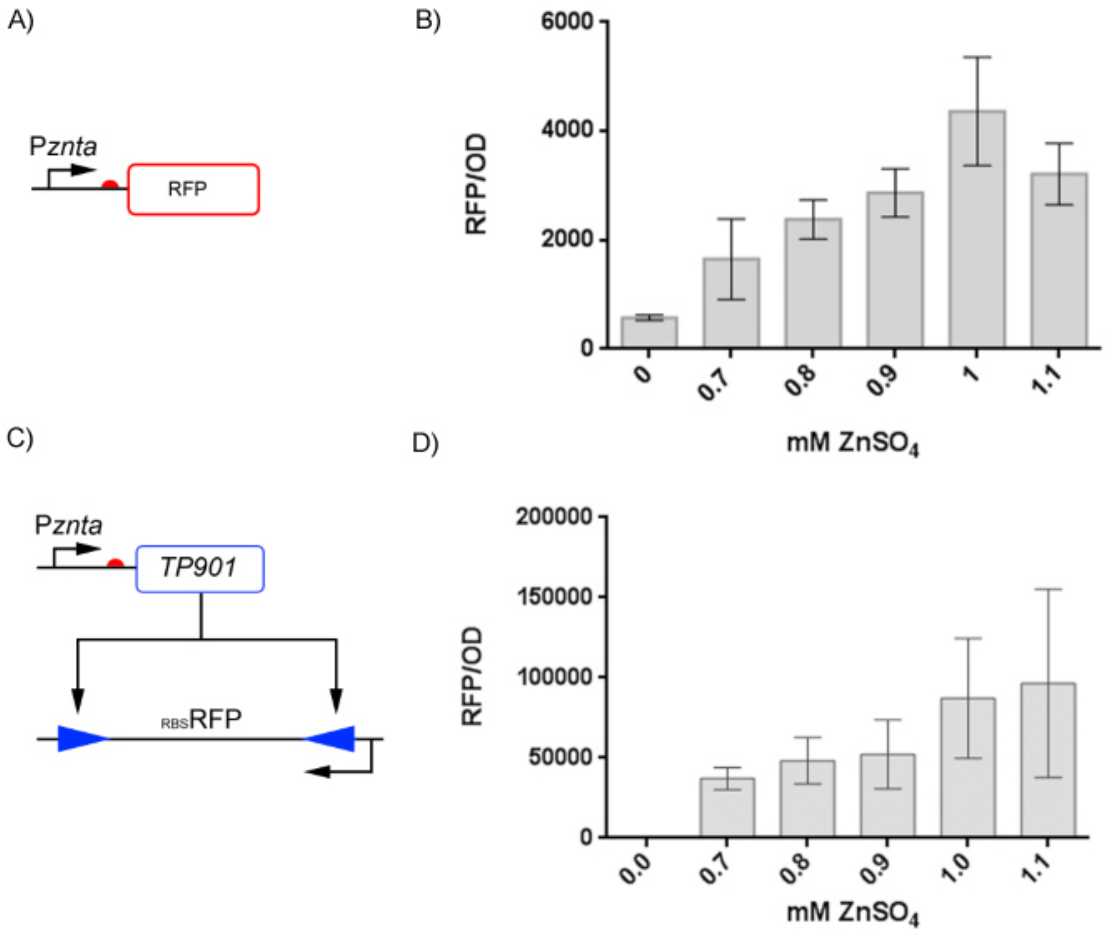


Figure 3.3: **Amplification of biosensor output** A)Schematic of zinc responsive biosensor expressing RFP that yields a response shown in B) to $ZnSO_4$. C) Integrase is introduced in place of the reporter yielding a large amplification over the simple biosensor in D). Basal expression in the absence of any inducer was not significantly different (Mean of 573 for (B) vs 482 for (D) RFP/OD)

Analysis of the single cell profile yielded a valuable insight into the population data. We see in Figure 3.4 A) that the simple biosensor yields a shifting population that is distinct from the negative control only when fully induced. Inducer concentrations between 0.7 and 1.0 mM yield an unpredictable response with a

clearer induction response to 0.8mM than 0.9mM. By comparison the amplified biosensor provides a very predictable shift into a distinct on state in a inducer dependent manner i.e. an increase in inducer concentration yielded more fluorescence. Ideally the inducer concentration that caused this shift would be tunable though another independent signal. These early data yielded significant promise and led to further characterisation.

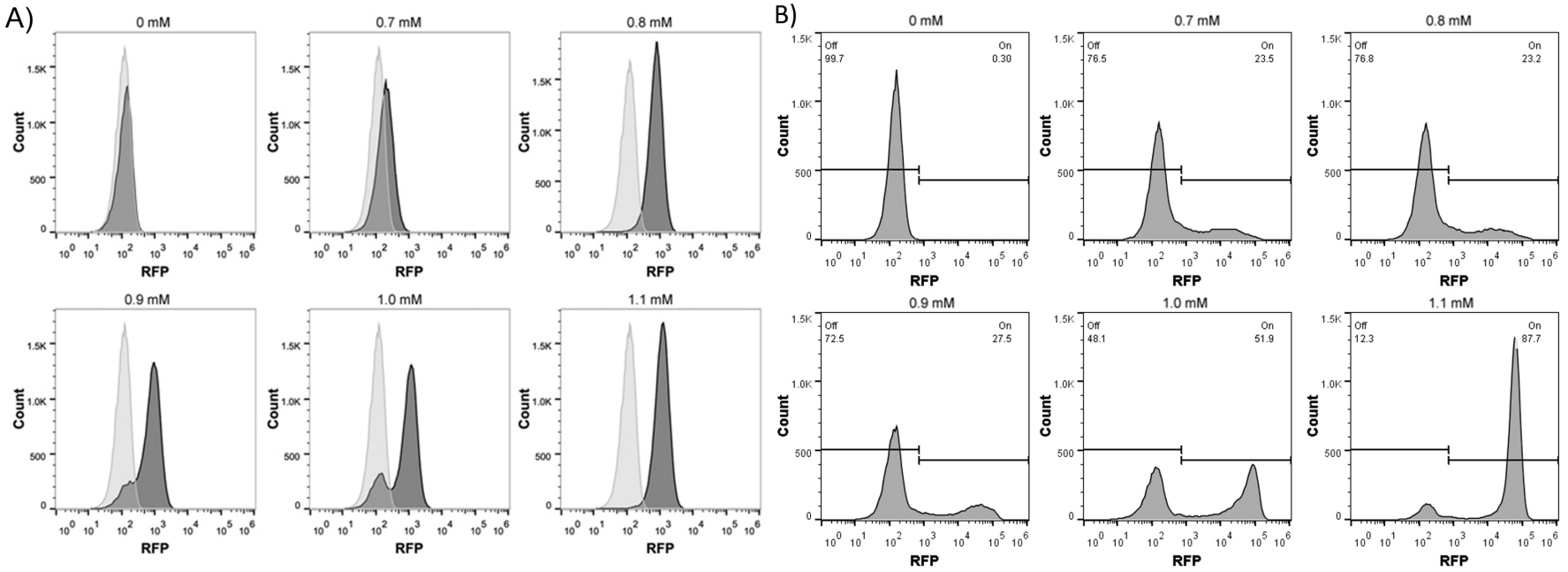


Figure 3.4: **Single cell populations for a regular and an amplified biosensor** A) Direct expression of RFP from pZNTA yields a distinct activated population above 1.1mM $ZnSO_4$, below this concentration population overlap makes it hard to discern an distinct response. Light grey populations indicate cells harbouring no plasmid. B) Amplified expression through TP901-1 integrase yields a very clear population that shifts to a distinct profile in response to greater $ZnSO_4$ concentrations. Gating was applied as described in section 2.6.1

3.7.2 Reproducing reported Integrase function

To move towards developing a plug and play approach to tunable biosensors a previously characterised integrase logic gate [152] centering around TP901-1 integrase was reproduced. This logic gate functions by placing TP901-1 integrase under the control of a riboswitch, repressing translation of the integrase by a cis-repressing region in the 5′ UTR. Translation occurs after binding of a short trans-activating RNA aptamer which exposes the RBS allowing for translation as shown in Figure 3.5A. Transcription of the cis-repressing and trans-activating domains is under control of the pTET and pBAD promoters respectively. When both promoters are induced TP901-1 integrase is translated and an AND logic function results, Figure 3.5B. A small increase in fluorescence when only the cis-repressed integrase is expressed indicates of the limitations of this approach. It would seem plausible that either there are some of the cis-repressed molecules or RNAs improperly fold allowing for unintentional translation of the integrase.

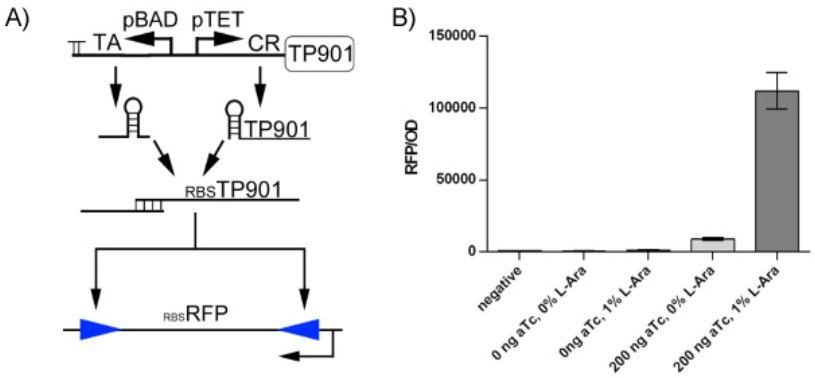


Figure 3.5: **Experimental validation of cis-repressed TP901-1 integrase**. A) Overview schematic of trans-activating/cis-repressing system showing dual expression of both components as required for DNA inversion. B) Experimental demonstration of AND gate functionality showing expression associated with expression from both promoters.

3.7.3 Characterisation of activation thresholds

To fully characterise the different activation thresholds, the pBAD promoter was replaced by the L-rhamnose inducible promoter pRHA to allow for a graded transcription of the trans-activating RNA. This construct, shown in Figure 3.6A, was then characterised over a wide range of aTc and L-rhamnose concentrations (Figure 3.6B). Expression of TP901 very rapidly ceases to be a linear function of inducer concentration suggesting that at low inducer concentrations TP901-1 integrase is quickly saturated. Furthermore the high basal expression in the absence of any L-rhamnose but with 150ng/ml aTc, is in line with the previous experiment shown in Figure 3.5B.

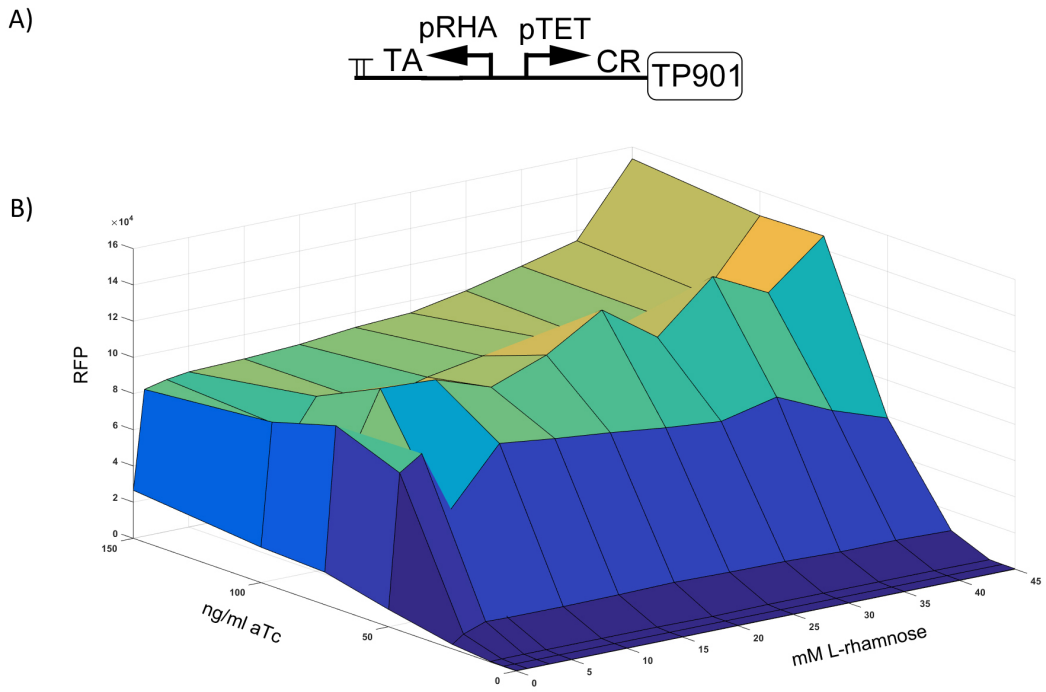


Figure 3.6: **Dual induction of TP901-1 integrase with the L-rhamnose and the pTET promoters**. A) Schematic of the promoter orientation. B) Full characterisation between 0 and 150ng/ml aTc induction for the cis-repressed TP901-1 integrase and trans-activating aptamer induction of between 0 and 45 mM L-rhamnose.

Analysis on the population and single cell level where one inducer was held static yielded an insight into the activation range. In the presence of 45mM L-rhamnose

the induction window for TP901-1 integrase was between 10 and 50 ng/ml aTc (Figure 3.7A) and single cell analysis revealed that a distinct but not discrete population resulted from full induction (Figure 3.7B).

When aTc was held at 150ng/ml and L-Rhamnose was varied a much sharper response was observed between 0 and 1 mM L-Rhamnose as seen in Figure 3.7C. This was accompanied by a much less discrete uninduced population as shown in Figure 3.7D.

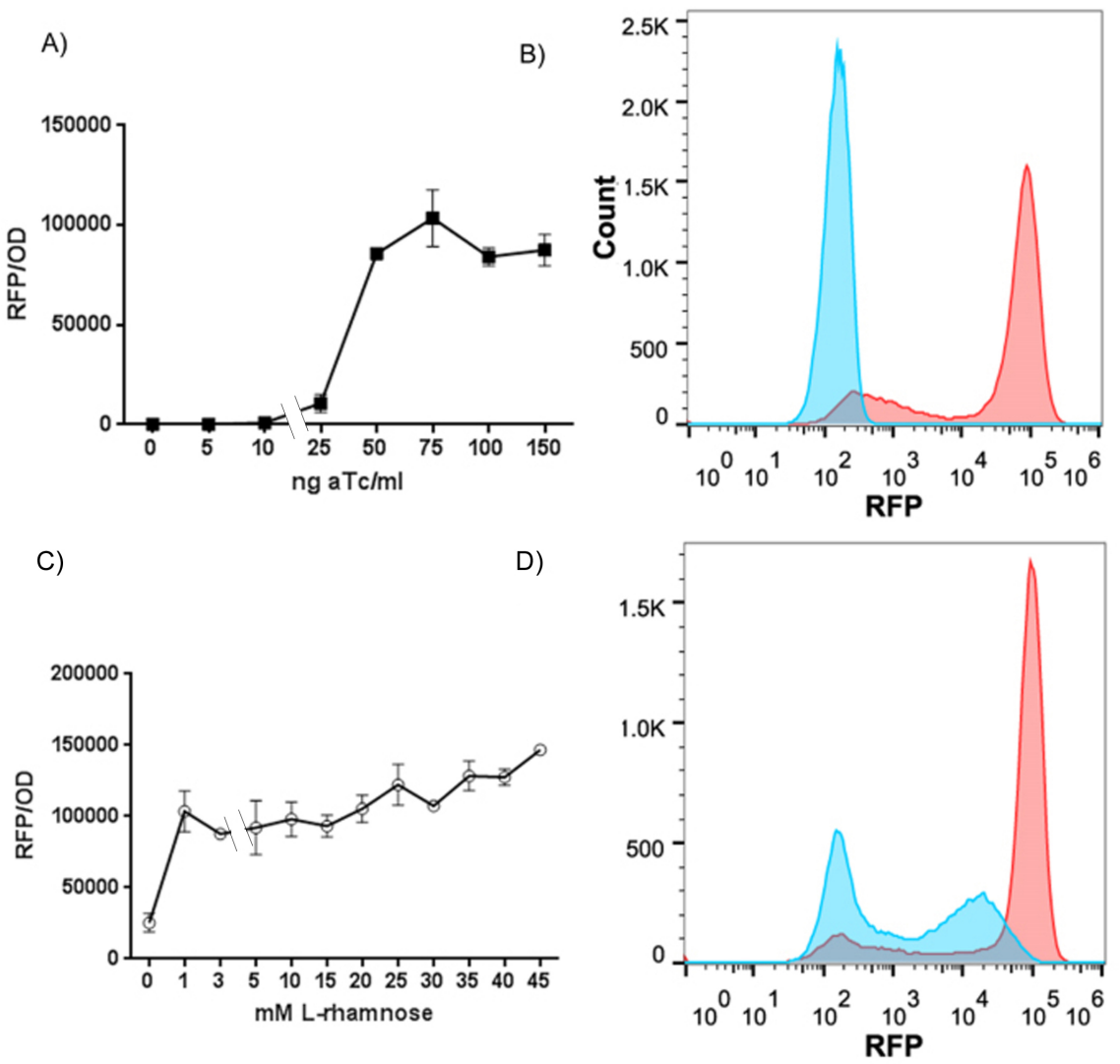


Figure 3.7: **Cross section as a function of aTc induction**. A) induction response in the presence of 45mM L-rhamnose and varying aTc concentrations. B) single cell analysis showing representative populations at 0 (blue) and 150 (red) ng/ml aTc. C) Induction response in the presence of 75 ng aTc and varying concentrations of L-rhamnose. D) Single cell analysis showing representative populations in the presence of 0mM (Blue) and 45 mM (Red) L-rhamnose.

3.7.4 Aim 2: To determine the context specific sensitivity of these parts to different promoter and trans-activating pairings

To further investigate sensitivity in promoter pairings, the orientation of the two inducible promoters was reversed resulting in the trans-activating RNA being ex-

pressed from the pTET promoter and the Cis-repressed TP901-1 integrase being expressed from the pRHA promoter as shown in Figure 3.8A. This construct was tested over the same range of aTc and L-Rhamnose concentrations and yielded the expression profile in Figure 3.8B. This construct gave a different activation profile to that shown in Figure 3.6B; only showing an induction profile as a function of L-rhamnose with no obvious effect from aTc induction.

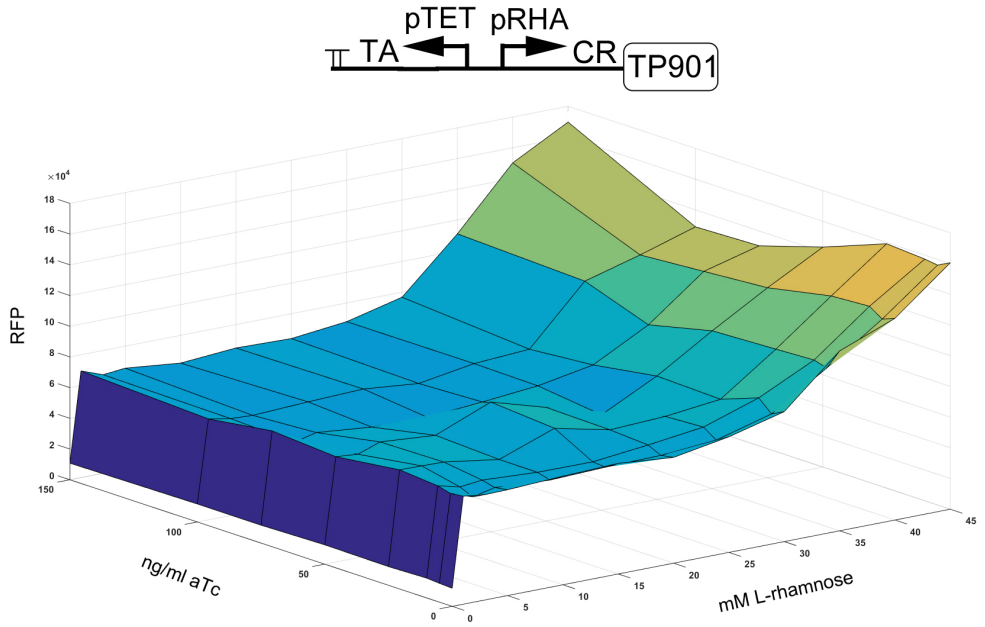


Figure 3.8: **Dual induction of TP901-1 integrase with the pTET and pRHA promoters** Full characterisation between 0 and 150ng/ml aTc induction for the trans-activating aptamer and induction of between 0 and 45 mM L-rhamnose for the cis-repressed TP901-1 integrase

The results when one inducer was held static showed that while L-rhamnose alone gave an expected induction response on a population (Figure 3.9A) and a single level (Figure 3.9B) there was no induction response from the addition of aTc (Figure 3.9C).

One possible explanation for this is a change in context for the ribo-regulators that control TP901-1 integrase translation. The act of swapping promoters caused the ribo-regulator to be transcribed with a few extra bases on the 5′ end. This appears to have caused the ribo-regulator hairpin to misfold and cease to regulate TP901-1

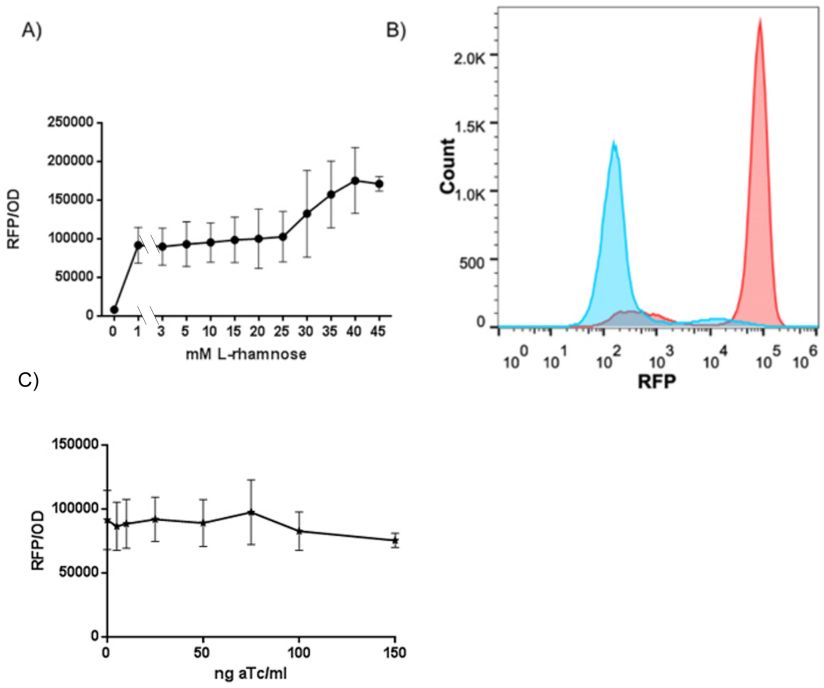


Figure 3.9: **Cross section as a function of L-Rhamnose and aTc induction**. A) induction response in the presence of 150 ng/ml aTc and varying L-rhamnose concentrations. B) single cell analysis showing representative populations at 0 (blue) and 45 (red) mM L-rhamnose. C) Induction response in the presence of 45mM L-rhamnose and varying aTc concentrations

Integrase expression as a function of the trans-activator binding. As such TP901-1 Integrase expression is only controlled by the Rhamnose promoter.

3.7.5 Folding analysis of the cis-repressing region

To investigate the source of this context sensitivity the cis-repressing region was modelled in RNAfold [99] to investigate any change in secondary structure. The folding analysis of the mRNA transcribed from both tetracycline and L-rhamnose promoters shown in Figure 3.10 showed no change in secondary structure and only slight reduction in hairpin stability. While previous results indicate that a change in hairpin stability has taken place, this is not supported by the *in vitro* modelling. The use of the integrase gave a step like, digital response compared

with the simple biosensor.

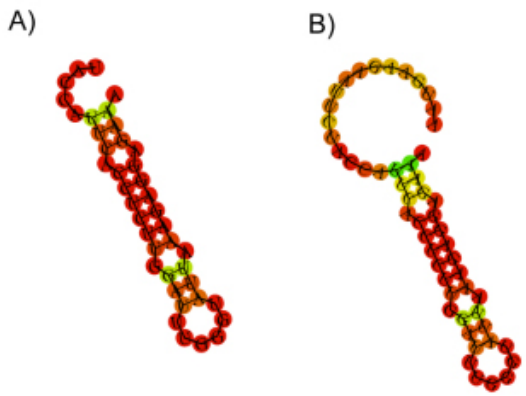


Figure 3.10: **Folding analysis of the 5 region of cis-repressing mRNA transcribed from the tetracycline and L-rhamnose promoters.** A)Expression from the tetracycline promoter. B) Expression from the L-rhamnose promoter

3.7.6 Integrating a zinc responsive biosensor

The first biosensor I investigated was a zinc responsive promoter region of Znta. The trans-activating RNA was placed under control of the Znta promoter. This gave a predictable RFP expression as a function of both aTc and zinc inducers as shown in Figure 3.11.

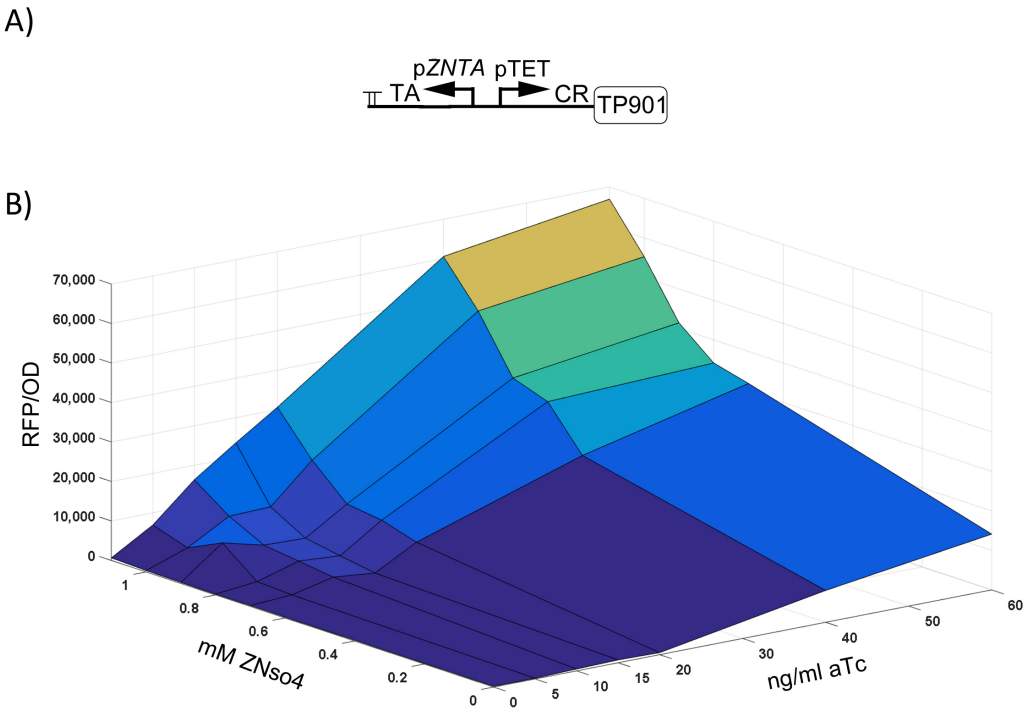


Figure 3.11: **Dual induction of TP901-1 integrase with the pTET and zinc promoters** A) schematic of the promoter orientation. B) Full characterisation between 0 and 60ng/ml aTc induction for the cis-repressed TP901-1 integrase and induction of between 0 and 1.1 mM ZNso4.

Further analysis showed that the biosensor sensitivity is tunable. In Figure 3.12A we can see the effect of varying aTc concentration in the presence of 1.1mM $ZnSO^4$. As aTc concentration increases, sensitivity to zinc also increases. In 3.12B we see the alternative effect of varying zinc concentration at 20 and 60 ng/ml aTc where basal expression was much higher, with greater aTc giving greater fluorescence. This effect was also seen through single cell analysis at 20 ng/ml aTc as shown in Figure 3.12.

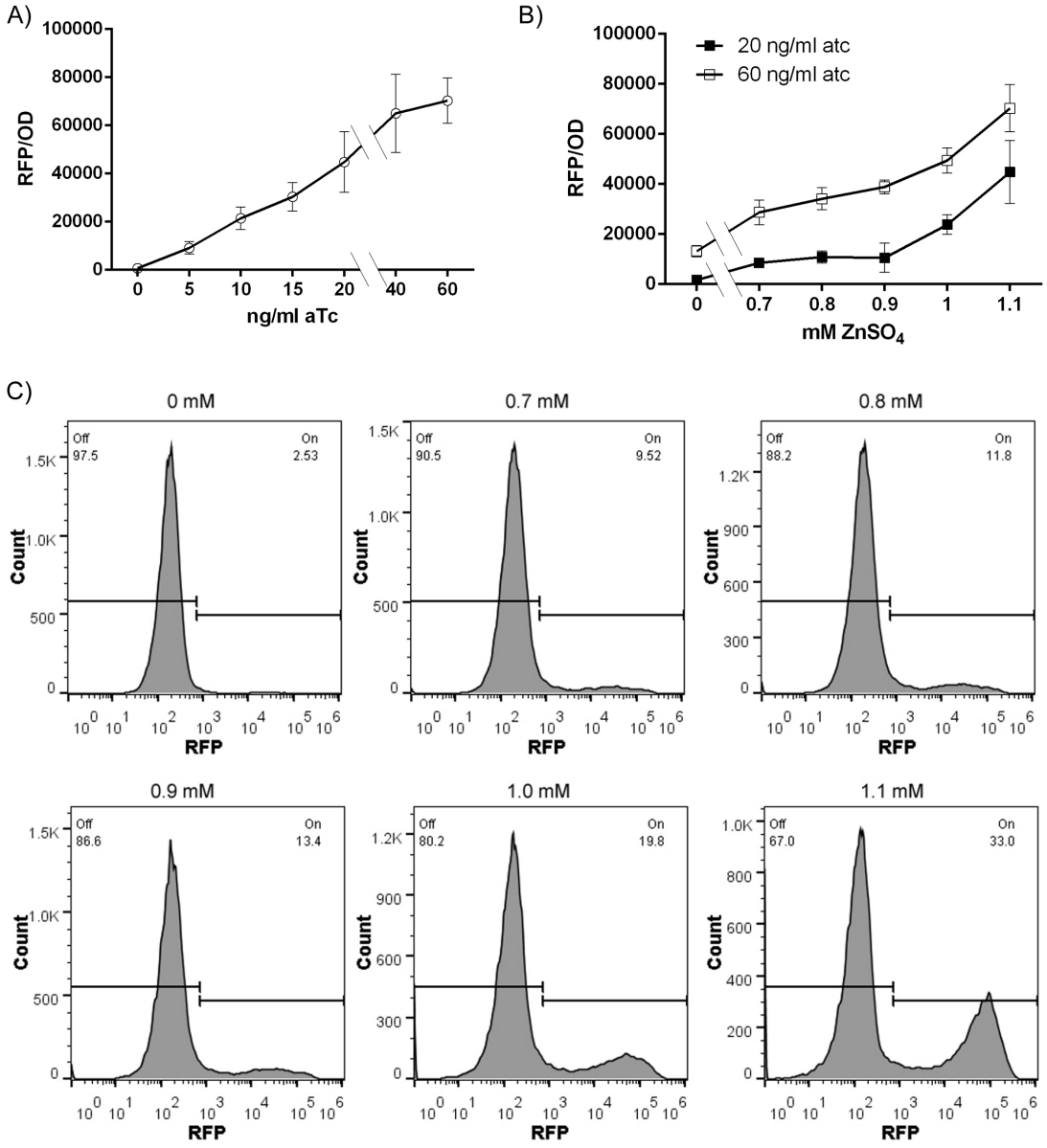


Figure 3.12: **Dual induction of TP901-1 integrase with the pTET and zinc promoters** A) Induction with 1.1mM $ZnSO^4$ over a range of aTc concentrations. B) Induction from 0 to 1.1 mM $ZnSO^4$ with either 20 ng/ml aTc or 60 ng/ml aTc. C) Single cell analysis in the presence of 20 ng/ml aTc and varying concentrations of $ZnSO^4$ inducer.

3.7.7 Integrating a copper biosensors

To further investigate this system a copper biosensor was introduced in the same fashion as the zinc biosensor. The biosensor worked well when expressing RFP

directly however when the trans-activating RNA was placed under its control it failed to express functional trans-activating RNA as shown in Figure 3.13.

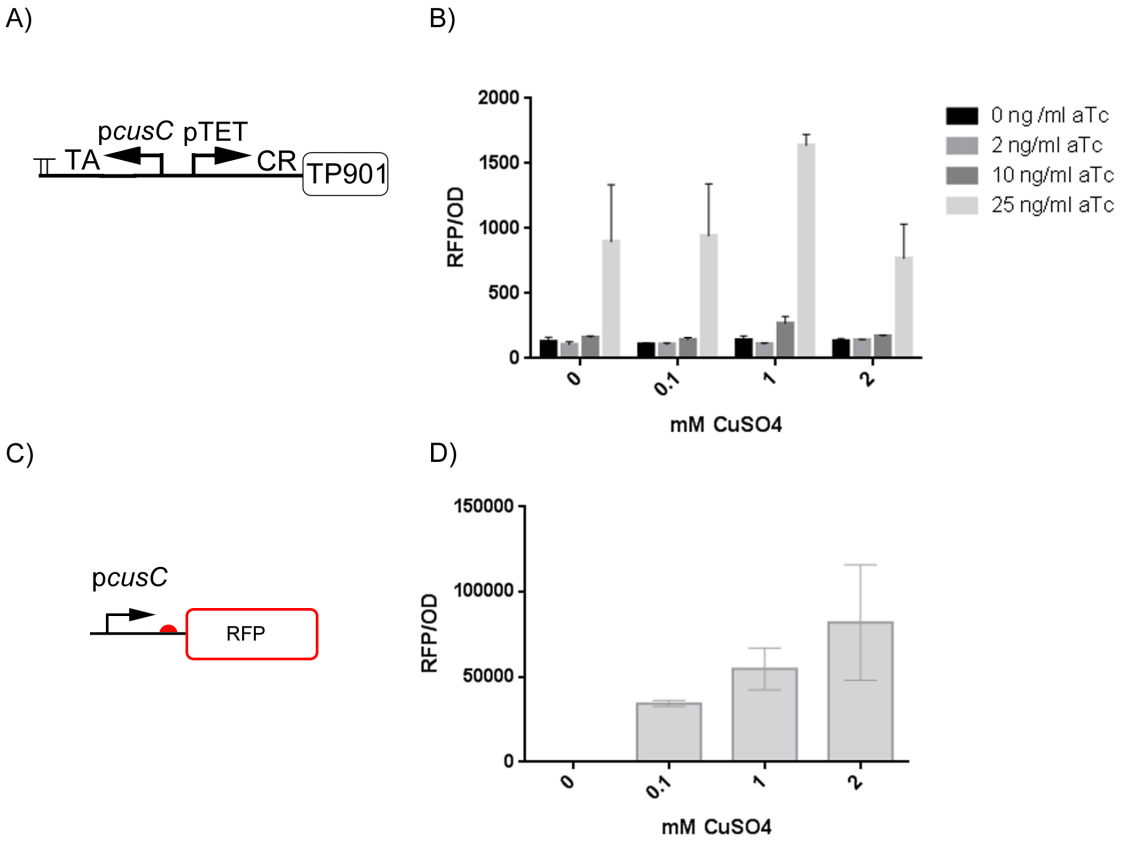


Figure 3.13: **Introduction of a copper biosensor to the ribo-regulators** A) The trans-activating RNA is placed under control of the copper biosensor. B) RFP expression in response to 0-2 mM $CuSO_4$ with the riboregulator and integrase over 0-25 ng/ml aTc. C) schematic showing the regular biosensor. D) RFP response to 2-3 mM $CuSO_4$ directly controlling transcription of an RFP.

3.7.8 Integrating an arsenite biosensor

An arsenite biosensor was also integrated into the ribo-regulated integrase. The schematic is shown in Figure 3.14A. As shown in Figure 3.14B, at low aTc concentrations no trend towards induction was seen, although at 10 and 25 ng/ml aTc an induction trend it seems to occur however it is not a significant one ($p => 0.05$, paired T-test).

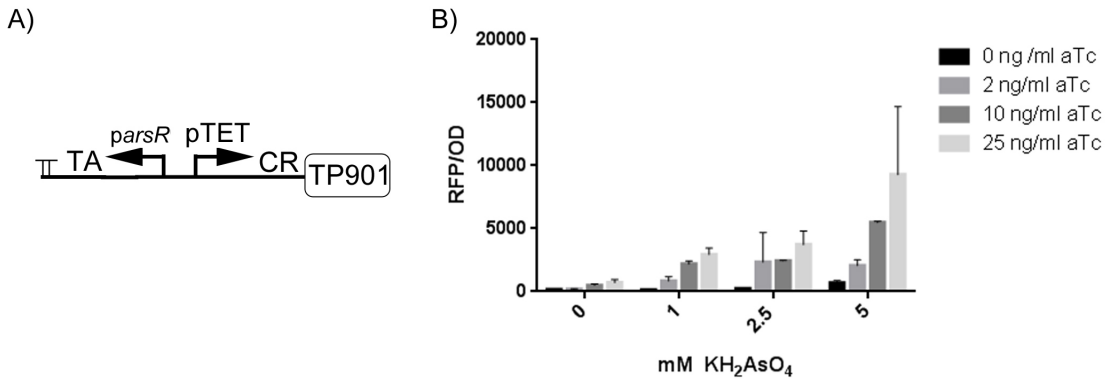


Figure 3.14: **Introduction of a arsenate biosensor to the ribo-regulators**A) The trans-activating RNA is placed under control of the arsenate biosensor. B) RFP expression in response to 0-5 mM KH_2 AsO_4 with the riboregulator and integrase over 0-25 ng/ml aTc.

3.7.9 Growth in the presence of heavy metals

Demonstrating that *E. coli* are still able to grow in the presence of described levels of heavy metals was an important control, I show the growth curves for 12 hours in Figure 3.15. From these data we show that *E. coli* are able to grow in the reported concentrations of heavy memtals and as such are a viable option for their detection.

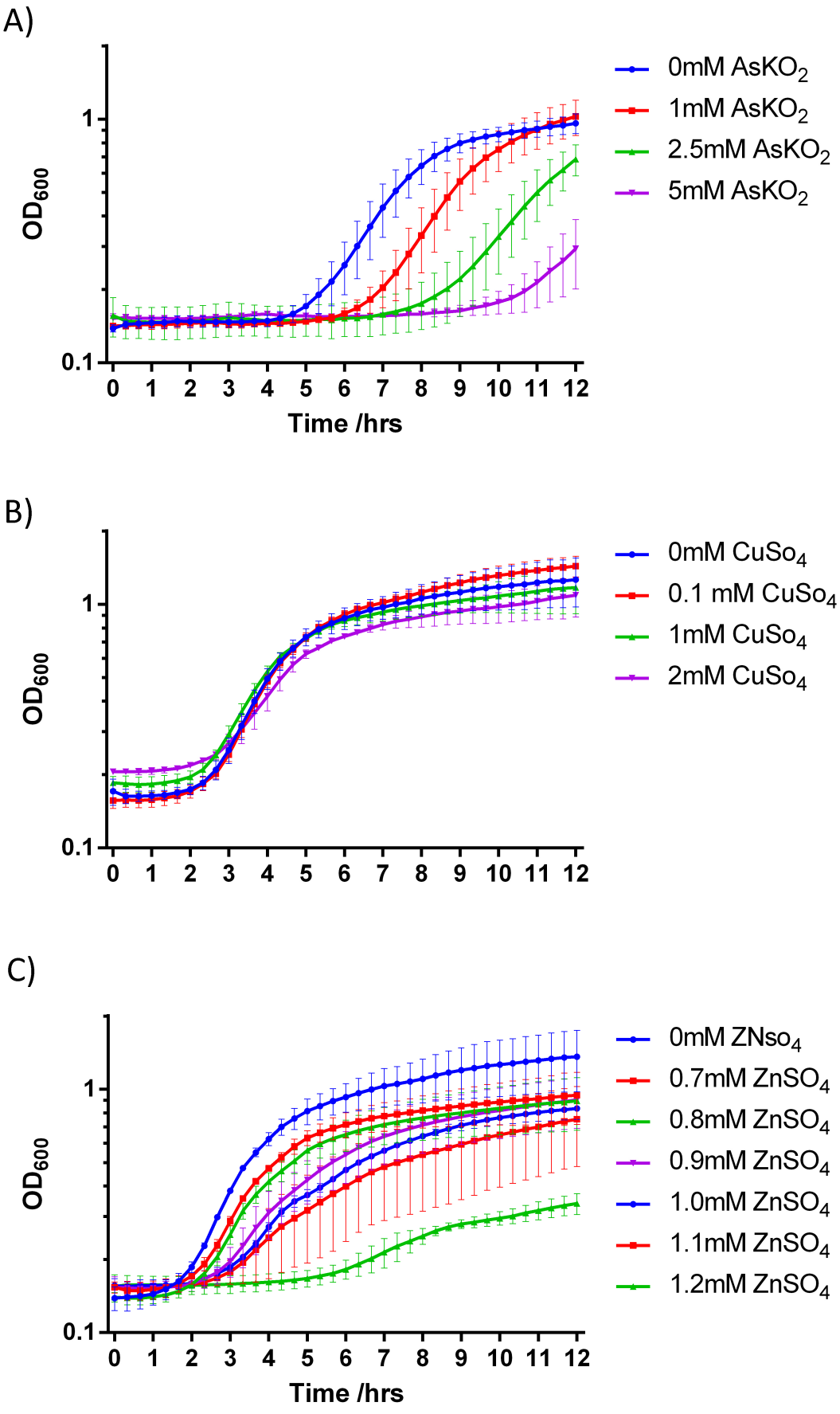


Figure 3.15: **Growth curves describing toxicity effect of heavy metal presence.** Growth over time for shown concentrations of A) Arsenite, B) Copper and c) Zinc

3.8 Discussion

The principle aims of this chapter were to investigate whether it was possible to combine two well documented technologies and apply them in a novel way to achieve more tunable biosensors or if other context dependencies would limit their applicability. I demonstrated that previously published integrase amplifiers [13] worked well with a zinc responsive promoter [124] but further tuning was required which would require lengthy and difficult RBS optimisation as previously documented when these Integrases had been optimised for different inputs [12]. I used a cis-repressing riboswitch which had previously been shown to work well to control the basal expression of this class of viral proteins [152]. Both parts of the riboswitch (the cis-repressed integrase and the trans-activating RNA) were expressed from the same promoter, the novel implementation involved placing only one of these post-transcriptional units under control of a transcriptional signal of interest while the other was placed under control of an orthogonal input. This input should then be able to be used to fine tune response from the biosensor. I first investigated that TP901-1 integrase expressed with the cis-repressed region under control of the pTET promoter and the trans-activating RNA under control of the pBAD promoter. While the construct gave an AND gate function as predicted, some leaky expression was observed in the presence of pTET induction and absence of pBAD induction. This basal expression was likely from a small population of cis-repressing hairpins mis-folding allowing expression of TP901-1 integrase in the absence of both inputs of the AND gate.

To further characterise this construct the pBAD promoter was replaced with the Rhamnose promoter pRHA. As this substitute promoter is a linear promoter it allowed for full characterisation over a full induction range. This showed the same basal expression in the presence of only aTc induction suggesting that the previous assumption of a small subset of mis-folded hairpins was correct. To fur-

ther investigate this leaky expression the promoters were reversed so that pTET now controlled the trans-activating RNA while pRHA now controlled the cis-repressed integrase RNA. This resulted in a construct that failed to respond to aTc, only expressing in response to L-Rhamnose induction and as such was not an AND gate. This could be the result of a changed secondary structure across the cis-repressed region however both promoters have known transcriptional start sites and folding analysis did not predict a change in secondary structure. This was the first indication that the post-transcriptional riboswitches were sensitive to contextual changes.

To further characterise this sensitivity a zinc responsive promoter pZNTA was integrated to control the trans-activating RNA. This yielded the predicted AND function when induced with zinc and aTc giving an induction range many orders of magnitude higher than the conventional biosensor. We also saw an increase in expression in response to zinc when a greater concentration of aTc was used as expected. Single cell analysis revealed only a small population was flipping suggesting inefficient pairing of the trans-activating RNA with the cis-repressing region. Integration of a further copper and arsenite biosensor yielded mixed results. The copper biosensor, despite working when controlling direct expression of an RFP, failed to transcribe a functional trans-activating RNA giving an induction response not associated with copper induction. The arsenite biosensor did work with the trans-activating RNA but its induction response was much smaller than that of the zinc responsive construct suggesting it too was not transcribing effective trans-activating RNA.

While amplifying biosensors using integrases has many advantages the use of post-transcriptional engineered riboswitches suffers from contextual sensitivity limiting its novel application. It is likely that tuning integrase amplifiers through RBS optimisation is the more reliable technique. In chapter 4 I look at how contextual changes in riboswitches affects their utilisation.

Chapter 4

Ribo-attenuators: novel elements for reliable and modular riboswitch engineering

I devised, designed and conducted the experiments. I designed the model with assistance of Thomas Prescott and Antonis Papachristodoulou , the model parameters, mathematics and computation was done by Thomas Prescott and Antonis Papachristodoulou. The OMPF expression and immunoblotting was carried out by Barbara Mertins.

'"Ribo-attenuators: novel elements for reliable and modular riboswitch engineering"' was submitted to Nature Scientific Reports on the 6^{th} of October 2016 (currently in review).

4.1 Aims

Riboswitches are structural genetic regulatory elements that directly couple the sensing of small molecules to gene expression. They have considerable potential for applications throughout synthetic biology and bio-manufacturing as they are able to sense a wide range of small molecules and thus regulate gene expression. The aims of this chapter are:

1. To investigate riboswitch sensitivity to changes in down stream Open Reading Frame (ORF).
2. To design and charaterise a genetic element to allow for the contextual isolation of a riboswitch from any ORF change.

4.2 Introduction

In the previous chapter I showed previously unknown contextual sensitivity in a class of trans-activating riboswitch. In this chapter I investigate this through two more riboswitches. Riboswitches are structural regulatory elements generally found in the 5′ UTR of messenger RNA[143] which allow the regulation of a downstream gene or operon in response to the binding of small molecules such as cellular metabolites or metal ions [63, 98, 142, 164]. Relying solely on RNA for structure and appearing in all three domains of life, it is possible that they arose as an early regulatory element in a hypothesized RNA-based world[14]. Riboswitches can regulate a variety of processes including transcription, translation, and splicing in eukaryotes[89]. Riboswitches that regulate translation do so through the allosteric effects of small molecules binding to their aptamer domain. This causes a structural rearrangement which usually opens up or sequesters away a ribosome binding site (RBS) core as shown in Figure 4.1A [133, 125]. A range of riboswitches evolved to control the expression of enzymes in natural systems, from which synthetic riboswitches have been adapted to respond to specific molecules [70, 134, 169, 167, 39].

Ideally, riboswitches should be treated as modular "plug and play" devices to allow construction of genetic networks that respond to small molecules. However, despite significant promise and over a decade of research, limitations inherent in their function have restricted their applicability. For example, many riboswitches selectively bind a small molecule of interest using an aptamer whose specific RNA secondary structure is influenced by its own sequence, as well as the surrounding genetic context including the proximal open reading frame (ORF) under the riboswitch's control. Thus, substituting the original ORF with a new one on a 'start codon for start codon' basis can nullify the desired riboswitch response

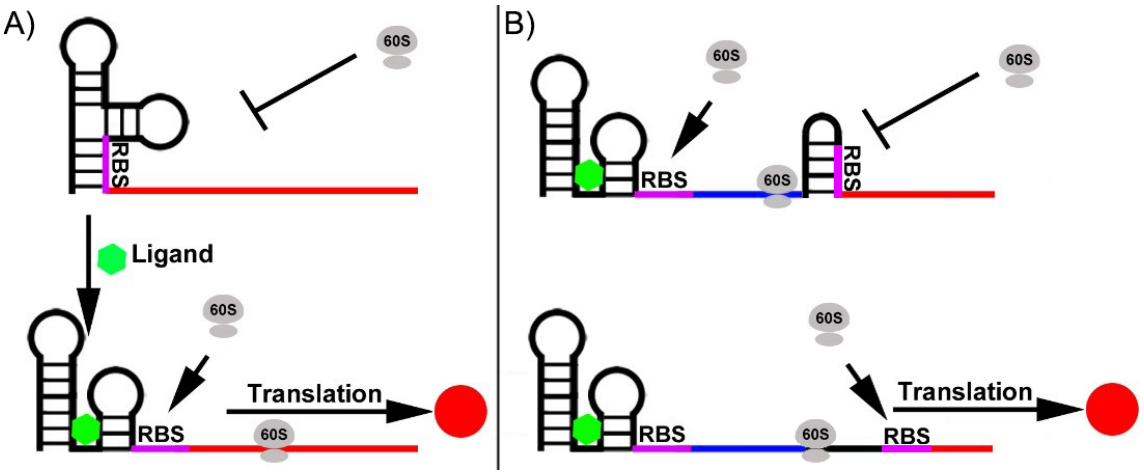


Figure 4.1: **Typical activating riboswitch function:** An RBS (purple) is sequestered away within a riboswitch preventing ribosome recruitment. (A) A standard riboswitch in which binding of a ligand (green) causes a conformational change exposing the RBS, allowing translation of the gene of interest (red). (B) A ribo-attenuator adds a second RBS, sequestered away by a local hairpin. The hairpin can be opened by a Ribosome travelling from the riboswitch RBS, exposing the attenuator RBS and allowing translation of the gene of interest.

to a given ligand [24]. To overcome this lack of modularity many studies have created fusions comprised of a riboswitch, the first few hundred base pairs of its working ORF, and a gene of interest [52, 111, 168, 40]. However, this approach fails in many circumstances as it can alter the gene's functionality, since many enzymes will not work with the inclusion of large 5' fusions [129, 1, 27, 170, 172].

Despite advances towards the *in silico* design of riboswitches [163, 5, 45], adjusting their functional range in response to a given ligand remains challenging. This difficulty arises because a riboswitch's activation or repression response is determined by both its RBS strength and secondary structure (which is itself dependent on the RBS sequence) [45]. Therefore, due to the specificity of the secondary structure surrounding the aptamer domain of riboswitch (and a poor understanding of *in vivo* RNA structures), changing the riboswitch induction range (via alteration of the RBS, generally a more predictable approach than promoter modification) without destroying the induction response is difficult. Further lim-

itations to application of many synthetic riboswitches arise due to variation in the expression of genes under their control. This variation is thought to originate from ligand-dependent RBS accessibility bursts [133] and limits their use with many high throughput screening techniques (such as Fluorescence Activated Cell Sorting (FACS) sorting) that rely on a distinct separation between a positive and negative population. Several solutions to these individual problems have been suggested. For example, the use of T7 polymerase could amplify a riboswitch's output and improve its functional range [43], however, its application is limited by its reliance on low basal expression, and does nothing to improve expression noise.

To overcome the limitations of large 5' fusions, fixed induction ranges, and sensitivity to ORF changes of both engineered and natural riboswitches, I designed and tested ribo-attenuators (Atts) a schematic of which is shown in Figure 4.2. These novel genetic elements are placed after 150 base pairs of a riboswitch's working ORF. They consist of an RBS, over which a hairpin is engineered to silence translation independent of upstream riboswitch activity, followed by a negative one-shifted transcriptionally-fused stop and start codon (TAATG) as shown in Figure 4.1. The passage of ribosomes recruited by the upstream riboswitch opens up the introduced hairpin, before dissociation at the proximal in-frame stop codon (TAA) in the transcriptionally-fused junction. Additional ribosomes can then assemble at the ribo-attenuator RBS and initiate translation at the first start codon of the introduced gene of interest as shown in Figure 4.1B. Therefore, instead of directly controlling the translation of a gene as in its natural setting, the riboswitch instead controls the translational initiation rate from the downstream attenuator RBS. As such ribo-attenuators facilitate tuning of a riboswitch's response, the orthogonal expression of a novel gene of interest, and by improving the reliability with which the downstream RBS remains exposed, they also re-

duce expression variability within populations. Ribo-attenuators thus overcome many of the issues preventing widespread use of re-engineered or synthetic riboswitches.

In this chapter I employ two riboswitch systems; the activating addA riboswitch from *Vibrio vulnificus* [39, 144] that binds 2-aminopurine, and the repressing btuB riboswitch from *E. coli* [52] that binds adenosylcobalamin. For both of these riboswitches I show that replacement of their working ORF results in a loss of ligand response, though this can be overcome by fusing an introduced gene-of-interest (GOI) to 150 bp of their working ORF. I then design and test nine potential ribo-attenuators within the addA riboswitch system, from which I select five attenuators that give a range of output responses. These five attenuators are used with the addA riboswitch to express sfGFP and OmpT, demonstrating that ribo-attenuators can reduce expression variability (which I support with a theoretical model), avoid build-up of inclusion bodies, and (in the case of OmpT) achieve proper targeting of membrane proteins. The same five attenuators are then modularly inserted into the btuB riboswitch system. These are used to express eGFP and ColE9, demonstrating again that ribo-attenuators can reduce expression noise, and that they allow amplification and tuning of a riboswitchs output response. The experimental results thus support the theoretical mechanisms that motivated the design of ribo-attenuators, and demonstrate that ribo-attenuators can be modularly employed in different riboswitch systems to control the expression of a range of functional outputs.

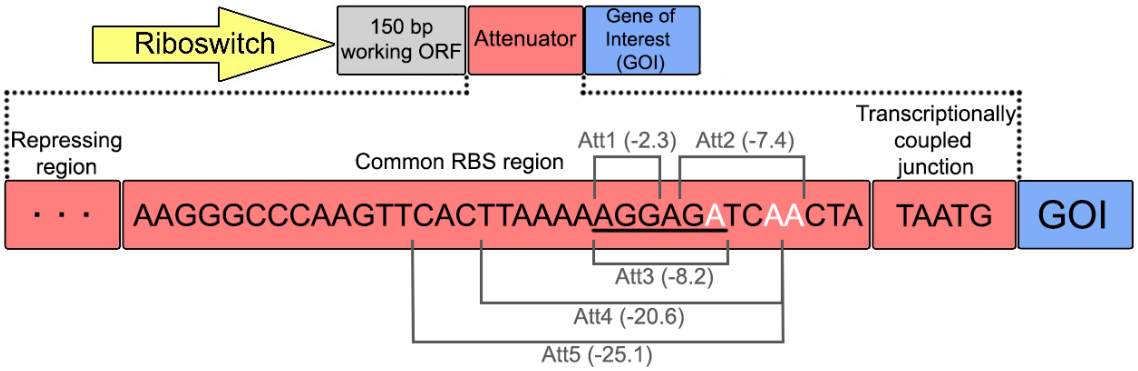


Figure 4.2: **Ribo-attenuator context and schematic:** A ribo-attenuator is inserted after a riboswitch and 150 bp of its native ORF, and preceeds the desired gene of interest (GOI). I defined a ribo-attenuator element to include a repressing region, a common RBS region (within which the RBS is underlined), and a transcriptionally coupled junction. The repressing region, which for all attenuators except Att2 is the reverse complement of a subsequence (indicated in grey) of the common RBS region (see table 4.2 for sequences), is selected so that the attenuator region as a whole forms a hairpin with appropriate ΔG (see Figure 4.3 for predicted conformations). For Att2 no repressing region was included. Instead, bases highlighted in white were altered to encourage formation of a hairpin (which involved the sequence indicated in grey) downstream of the RBS.

4.3 Materials and methods

4.3.1 Plasmids

All riboswitch cassettes were cloned into the J64100 plasmid [13] (ColE1 origin; 50-60 copies, chloramphenicol resistance) under control of the tetracycline promoter. The addA riboswitch sequence was the kind donation of Dr Neil Dixon (Department of Chemistry, University of Manchester) and the btuB riboswitch was taken from the *E. coli* genome. All riboswitches were synthesised by IDT (Skokie, IL, USA) as gBlock gene fragments, assembled by Gibson assembly [57, 55] and sequence confirmed. A full list of plasmids and Addgene references is provided in table 4.1. A list of ribo-attenuator sequences is provided in table 4.2.

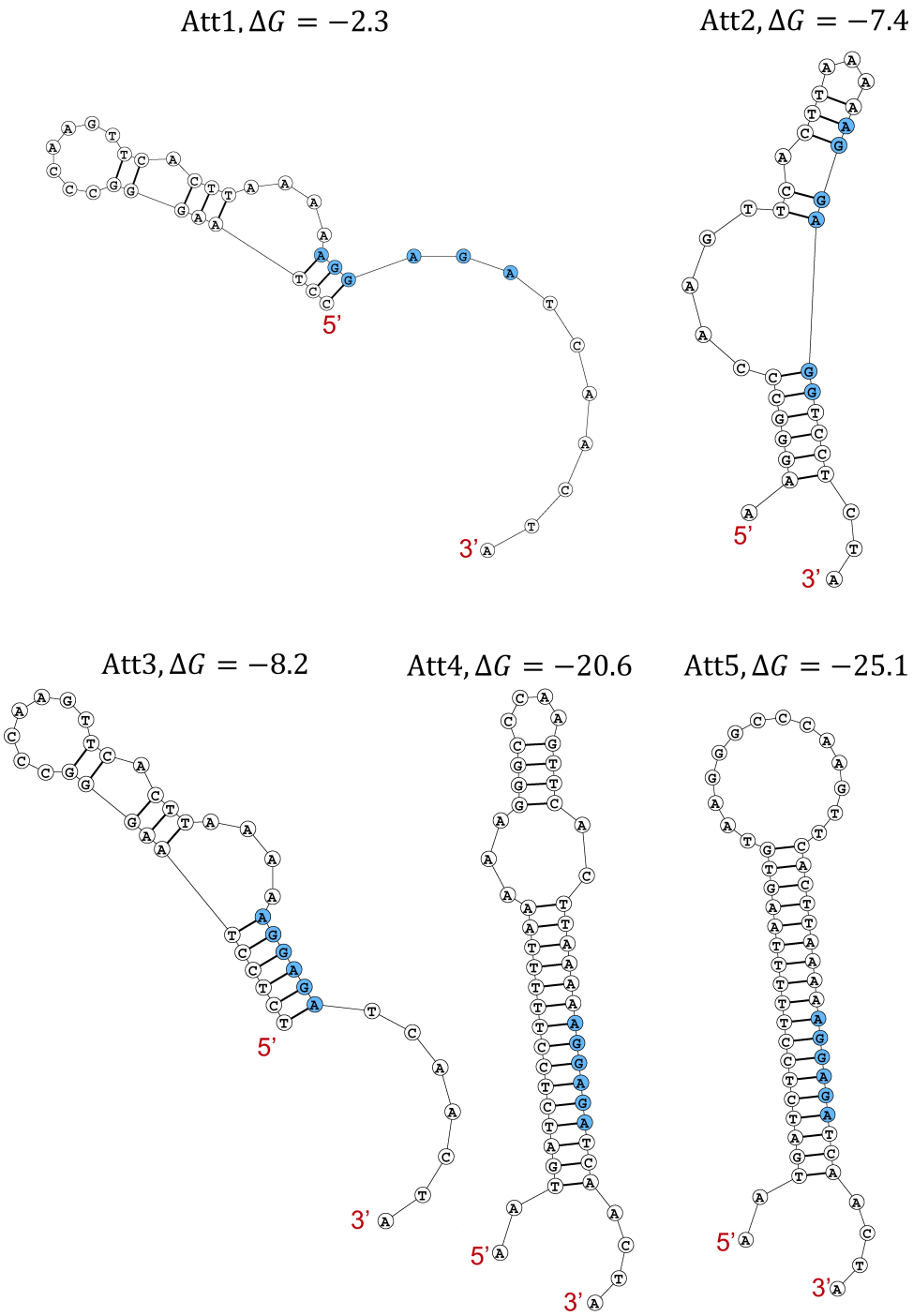


Figure 4.3: **Ribo-attenuator secondary structures:** Lowest free energy conformations were calculated for each ribo-attenuator using the RNAstructure Fold Web Server [130]. These structures were calculated for the attenuators in isolation, and thus may vary depending on corresponding genetic contexts, as well as due to limitations inherent in the theoretical prediction of RNA structures. ΔG values are in kcal/mole, and the RBS core is highlighted in blue. For isolated sequences see Table 4.2.

4.3.2 Growth and induction

All experiments were performed in *E. coli* DH5αZ1 [94]. For each experiment three freshly transformed colonies were inoculated in LB medium (10g/L tryptone, 5g/L yeast extract, 10g/L NaCl) with chloramphenicol (25 μg/mL) and grown overnight. For the functional assay and riboswitch characterisation, overnight cultures were diluted 1:100 into 1 mL LB in a deep well plate (Greiner bio-one) with chloramphenicol (25 μg/mL), anhydrotetracycline (150 ng/mL) and stated concentrations of riboswitch inducer. For protein assays overnight cultures were diluted 1:50 into 50ml LB and the same antibiotic and inducer concentrations were used. Cultures were induced for 8 hours at 37 °C with shaking before data collection or harvesting for protein analysis. Biological triplicates were generated to assess behavioural variations between cultures.

4.3.3 Data collection

After induction, 500 μL of each culture was centrifuged at 4000 RPM, washed with PBS and resuspended in 500 μL of PBS. OD_{600} measurements were taken in clear well plates and GFP measurements in black well plates in a CLARIOstar platereader (BMG Labtech, Ortenberg, Germany). GFP expression was quantified by normalising by OD_{600}. Flow cytometry was performed on an Attune flow cytometer (Lifetech Scientific, Basingstoke, UK) for each data point and 50,000 events were measured. Raw data and plasmid sequences are available online at [50]

4.3.4 Fluorescence microscopy

Fluorescence microscopy was performed on cells grown and induced as described above. Thin, flat agarose pads (1% w/v agarose in Milli-Q water) were generated

on microscope slides. 2 μL of cells were added to the pad immobilising them on the surface of the agarose. Cells were then imaged with a Nikon Eclipse Ti microscope (with NIS Elements software), 100$\times$ phase contrast objective (Nikon), GFP filter set (Chroma), Andor iXON CCD camera.

4.3.5 Cell disruption

Cells were harvested (3700g, 30 min) and pellets were resuspended in Lysis Buffer (PBS-Buffer with 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4, with addition of lysozyme (Sigma Aldrich), DNAse (Sigma Aldrich), and 1 protease inhibitor tablet per 50 ml solution (Complete EDTA free protease inhibitor, Roche)) relative to the wet weight of the cell pellet (1 mL buffer per 50 mg pellet). Cells were disrupted using a fast prep device (2$\times$ 6 m/s, 30s, FastPrep-24 5G, MP Biomedical) or sonication (1 min, 10s on, 4s off, 50%, Vibra-Cell VCX130, Sonics), and insoluble material was removed using centrifugation (8,000g (OmpT) or 20,000g (ColE9), 20 min).

4.3.6 Functional assay of ColE9

Supernatants of the lysed cells were separated and diluted from 10^0 - 10^{10} using PBS buffer. Their potency was analysed by a plate killing assay [131]: An LB plate containing 0.8% Agar was warmed to 42 °C before an established culture was added in a dilution of 1:10,000. The plate was poured and left to dry for 30 min at 37 °C. 2 μL of the supernatants at the given dilutions were spotted on a plate and incubated at 37 °C for 16h. The potency of ColE9 was assessed by observing the clearance around each spotted point on the plate.

4.3.7 Preparation of strep-tagged OmpT

After cell disruption, the supernatant containing membrane-bound OmpT and the pellet containing improperly folded OmpT were treated as follows. To precipitate the membranes, the supernatant was spun down (120,000g, 1.5h) and membranes were resuspended in 50 μL PBS. To extract membrane bound OmpT, the resuspended membranes were diluted with 50 μL PBS containing n-dodecyl-β-D-maltopyranoside (Anatrace) giving a final concentration of 0.635 mM (5 critical micelle concentration). The extraction was performed overnight by shaking at 4°C, and insoluble material was spun down (12,000g, 20 min) yielding soluble OmpT. The remaining pellets harbouring OmpT as inclusion bodies (IB) were washed as described previously [101] using PBS with 1% Triton-X-100 for removal of residual *E. coli* lipids. Washed IB were dissolved using 6M Gunanidinium chloride in PBS and incubated for 10 min at 95°C. Insuloble material was removed (12,000g, 5 min) yielding the IB fraction. Whole cells (WC), extracted membrane bound OmpT (Mem) and solubilized misfolded OmpT (IB) were resolved with 4-20% SDS-PAGE (Mini-PROTEAN TGX, BIORAD), using the Laemmli buffer compositions [85]. Immunoblotting was performed as described previously [160] using anti-strep antibodies (Thermo Fisher Scientific) and Luminata Western HRP Substrate (Merck Millipore).

4.3.8 Model Methods

The cumulative distribution of the random variable T denoting the time elapsed between an arbitrary time-point τ and the next production of a GFP molecule was calculated according to the model described in the appendix. This model treats the transition between closed (OFF) and open (ON) states of both the riboswitch and ribo-attenuator as stochastic random processes, with the rate of attenuator opening taking an increased value when the riboswitch is ON as seen

in Figure 4.12A. This dependency captures the system's intended mechanism, whereby a ribosome initiated at the riboswitch RBS opens the attenuator, revealing the second RBS. The primary assumptions made are therefore the model's parameterised rate values, for which I used the following:

$$\lambda(ON) = 50, \mu(ON) = 4, \lambda(OFF) = \mu(OFF) = 0.01K_{-} = 0.5m_{-} = 1$$
$$K_{+[I]} = \frac{[I]^2}{[I]^2 + [10]^2} \tag{4.1}$$

For $0 \leq [I] \leq 50$ (where $[I] = 50$ corresponds to 100% induction). The inverse of T is the random variable I call "Expression Rate". For the calculation of the CDFs, see appendix Information and the corresponding MATLAB code.

We consider two models: the one-component system, and the two-component system.

4.3.9 One-Component System

Here the RNA has a single riboswitch covering the RBS for GFP. There are two possible rates at which ribosomes bind to the RNA, depending on the state of the riboswitch covering the RBS. If the riboswitch is in the OFF position, GFP is translated at a rate λ_{OFF}. If the riboswitch is in the ON position, the translation rate of GFP is $\lambda_{ON} \gg \lambda_{OFF}$. For stability, GFP also degrades at a constant rate δ. The riboswitch randomly flips between OFF and ON, depending on the inducer concentration $[I]$, such that

$$OFF \underset{k_{-}}{\overset{k_{+}([I])}{\rightleftharpoons}} ON, \tag{4.2}$$

where $k_{+}([I])$ is assumed to be an increasing function of $[I]$. This is a random walk, where increasing inducer concentrations correspond to a bias towards the ON state.

4.3.10 Two-Component System

Here there are two switches on the RNA: the original riboswitch, and the attenuator. There are four possible states of the two switches: (OFF, OFF), (OFF, ON), (ON, OFF), and (ON, ON). In each of these states, the first ON or OFF corresponds to the state of the first switch, while the second ON or OFF corresponds to the state of the second switch. The intuition behind this system is that a ribosome binds to the RBS downstream of the first switch at the rate λ_{OFF} or λ_{ON} defined above, depending on its state. The ribosome then opens up the downstream switch. Ribosomes also bind to another RBS downstream of the first, and translate GFP at two rates μ_{OFF} and $\mu_{ON} \gg \mu_{OFF}$, now depending on the state of the second switch. Again, GFP degrades at the rate δ for stability.

The possible transitions between the four states of the two switches are:

$$(OFF, OFF) \underset{k_-}{\overset{k_+([I])}{\rightleftharpoons}} (ON, OFF), \tag{4.3a}$$

$$(OFF, OFF) \underset{m_-(L)}{\overset{\lambda_{OFF}}{\rightleftharpoons}} (OFF, ON), \tag{4.3b}$$

$$(ON, OFF) \underset{m_-(L)}{\overset{\lambda_{ON}}{\rightleftharpoons}} (ON, ON), \tag{4.3c}$$

$$(OFF, ON) \underset{k_-}{\overset{k_+([I])}{\rightleftharpoons}} (ON, ON). \tag{4.3d}$$

Here $k_\pm$ are as in the previous model, describing the dynamics of the upstream switch. There are two rates that the downstream region switches on: λ_{OFF} and λ_{ON}, depending on the upstream switch's state. The downstream switch spontaneously switches off at a rate $m_-(L)$ that I assume is an increasing function of L, the length of the attenuator region.

To emphasise: the two-component system has production rates of GFP of μ_{OFF} in the states (OFF, OFF) and (ON, OFF), and $\mu_{ON} \gg \mu_{OFF}$ in the states (OFF, ON) and (ON, ON).

4.3.11 Analysis

Our analysis of these random processes focused on a single random variable. I defined T to be the length of time elapsing from an arbitrary time τ until the next GFP molecule was produced. I assumed that both systems had reached stationarity, so that this was well-defined independently of τ. Importantly, this random variable is *not* equal to the time elapsed between two consecutive GFP production events: instead, the start time τ is arbitrary. In the main text, the inverse $1/T$ of this random variable is what I call the expression rate.

4.3.12 One-Component System

To find the CDF of T, I condition on the state of the RNA switch, so that

$$\mathbb{P}(T \leq t) = \mathbb{P}(T \leq t \mid OFF)\mathbb{P}(OFF) + \mathbb{P}(T \leq t \mid ON)\mathbb{P}(ON).$$

It is simple to see that, at stationary, I have $\mathbb{P}(ON) = k_+/(k_+ + k_-)$ and $\mathbb{P}(OFF) = k_-/(k_+ + k_-)$, where the notation of $[I]$ is dropped for now. I define two conditional CDFs $F_{OFF}(t) := \mathbb{P}(T \leq t \mid OFF)$ and $F_{ON}(t) := \mathbb{P}(T \leq t \mid ON)$, with boundary conditions $F_{OFF}(0) = F_{ON}(0) = 0$ and $\lim_{t\to\infty} F_{OFF}(t) = \lim_{t\to\infty} F_{ON}(t) = 1$.

Assume that the switch at time t is ON, and consider the auxiliary random variable X, which is the waiting time until the switch switches OFF. Conditioning on

$X \geq 0$, the conditional CDF $F_{ON}(t)$ satisfies the equation

$$\begin{aligned} F_{ON}(t) &= \mathbb{P}(T \leq t \mid ON) = \int_0^\infty \mathbb{P}(T \leq t \mid ON, X = x) k_- e^{-k_- x} \, \mathrm{d}x \\ &= \int_t^\infty \mathbb{P}(T \leq t \mid ON, X = x) k_- e^{-k_- x} \, \mathrm{d}x + \int_0^t \mathbb{P}(T \leq t \mid ON, X = x) k_- e^{-k_- x} \, \mathrm{d}x \\ &= \left(1 - e^{-\lambda_{ON} t}\right) e^{-k_- t} + \int_0^t \left[\left(1 - e^{-\lambda_{ON} x}\right) + e^{-\lambda_{ON} x} F_{OFF}(t - x)\right] k_- e^{-k_- x} \, \mathrm{d}x \\ &= \frac{\lambda_{ON}}{\lambda_{ON} + k_-}(1 - e^{-(\lambda_{ON} + k_-)t}) + e^{-(\lambda_{ON} + k_-)t} \int_0^t F_{OFF}(s) k_- e^{(\lambda_{ON} + k_-)s} \, \mathrm{d}s. \end{aligned}$$

We multiply this equation by $e^{(\lambda_{ON} + k_-)t}$, and take the derivative of both sides with respect to t. Re-arranging the result, I find that the PDF $f_{ON}(t) = \frac{\mathrm{d}}{\mathrm{d}t} F_{ON}(t)$ is equal to

$$f_{ON}(t) = -(\lambda_{ON} + k_-) F_{ON}(t) + k_- F_{OFF}(t) + \lambda_{ON}. \tag{4.4a}$$

By the symmetry of our notation (or a similar argument), the PDF of T conditional on the OFF switch state satisfies

$$f_{OFF}(t) = -(\lambda_{OFF} + k_+) F_{OFF}(t) + k_+ F_{ON}(t) + \lambda_{OFF}. \tag{4.4b}$$

Hence, matrix notation gives a first-order, two-dimensional ODE to solve for F_{ON} and F_{OFF}:

$$\begin{bmatrix} f_{ON} \\ f_{OFF} \end{bmatrix} = \begin{bmatrix} -(\lambda_{ON} + k_-) & k_- \\ k_+ & -(\lambda_{OFF} + k_+) \end{bmatrix} \begin{bmatrix} F_{ON} \\ F_{OFF} \end{bmatrix} + \begin{bmatrix} \lambda_{ON} \\ \lambda_{OFF} \end{bmatrix}. \tag{4.5}$$

Denote the two negative eigenvalues of the system matrix as $-\omega_1$ and $-\omega_2$, with corresponding eigenvectors v_1 and v_2. The solution to this ODE that satisfies the

boundary conditions is

$$\begin{bmatrix} F_{ON}(t) \\ F_{OFF}(t) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha_1 v_1 e^{-\omega_1 t} - \alpha_2 v_2 e^{-\omega_2 t},$$

where α_1 and α_2 are determined by the eigenvectors such that

$$\alpha_1 v_1 + \alpha_2 v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Finally, the CDF $F(t) = \mathbb{P}(T \leq t)$ is equal to $F(t) = 1 - A_1 e^{-\omega_1 t} - A_2 e^{-\omega_2 t}$, for the scalar coefficients

$$A_i = \alpha_i \begin{bmatrix} \frac{k_+}{k_+ + k_-} & \frac{k_-}{k_+ + k_-} \end{bmatrix} v_i.$$

From this CDF we can calculate quantiles, and also determine the mean of T as

$$\mathbb{E}(T) = \frac{A_1}{\omega_1} + \frac{A_2}{\omega_2}.$$

4.3.13 Two-Component System

We can now perform a similar analysis on the two-component system, but now need to consider the CDF of T conditional on the four possible states of the RNA. Following the example above, we write

$$\begin{aligned} F(t) &= \mathbb{P}(T \leq t) \\ &= \mathbb{P}(T \leq t \mid (OFF, OFF))\mathbb{P}(OFF, OFF) + \mathbb{P}(T \leq t \mid (OFF, ON))\mathbb{P}(OFF, ON) \\ &\quad + \mathbb{P}(T \leq t \mid (ON, OFF))\mathbb{P}(ON, OFF) + \mathbb{P}(T \leq t \mid (ON, ON))\mathbb{P}(ON, ON). \end{aligned}$$

We write the conditional CDFs as $F_{(OFF,OFF)}(t) := \mathbb{P}(T \leq t \mid (OFF, OFF))$ and so on. Given that the RNA is in state (OFF, OFF) at time t, we now let the auxiliary

random variable X denote the time elapsing between the arbitrary time t and a switch away from (OFF, OFF). The first conditional CDF $F_{(OFF,OFF)}$ then satisfies the equation

$$\begin{aligned}
F_{(OFF,OFF)}(t) &= \mathbb{P}(T \leq t \mid (OFF, OFF)) \\
&= \int_0^\infty \mathbb{P}(T \leq t \mid (OFF, OFF), X = x)(k_+ + \lambda_{OFF})e^{-(k_+ + \lambda_{OFF})x}\, \mathrm{d}x \\
&= \left(1 - e^{-\mu_{OFF} t}\right) e^{-(k_+ + \lambda_{OFF})t} + \int_0^t (1 - e^{-\mu_{OFF} x})(k_+ + \lambda_{OFF})e^{-(k_+ + \lambda_{OFF})x}\, \mathrm{d}x \\
&\quad + \int_0^t \left[e^{-\mu_{OFF} x}\left(k_+ F_{(ON,OFF)}(t - x) + \lambda_{OFF} F_{(OFF,ON)}(t - x)\right)\right] e^{-(k_+ + \lambda_{OFF})x}\, \mathrm{d}x \\
&= \frac{\mu_{OFF}}{\mu_{OFF} + k_+ + \lambda_{OFF}}\left(1 - e^{-(\mu_{OFF} + k_+ + \lambda_{OFF})t}\right) \\
&\quad + e^{-(\mu_{OFF} + k_+ + \lambda_{OFF})t} \int_0^t \left[k_+ F_{ON,OFF}(s) + \lambda_{OFF} F_{OFF,ON}(s)\right] e^{(\mu_{OFF} + k_+ + \lambda_{OFF})s}\, \mathrm{d}s.
\end{aligned}$$

We then multiply both sides by $e^{(\mu_{OFF} + k_+ + \lambda_{OFF})t}$ and take derivatives. Re-arranging for the PDF $f_{(OFF,OFF)}$, we find that

$$f_{(OFF,OFF)}(t) = \mu_{OFF} + \lambda_{OFF} F_{(OFF,ON)}(t) + k_+ F_{(ON,OFF)}(t) - (mu_{OFF} + k_+ + \lambda_{OFF}) F_{(OFF,OFF)}(t).$$

Performing a similar analysis for the other CDFs, we find that

$$\begin{bmatrix} f_{(OFF,OFF)}(t) \\ f_{(OFF,ON)}(t) \\ f_{(ON,OFF)}(t) \\ f_{(ON,ON)}(t) \end{bmatrix} = \begin{bmatrix} -(\mu_{OFF} + k_+ + \lambda_{OFF}) & \lambda_{OFF} & k_+ & 0 \\ m_- & -(\mu_{ON} + m_- + k_+) & 0 & k_+ \\ k_- & 0 & -(\mu_{OFF} + k_- + \lambda_{ON}) & \lambda_{ON} \\ 0 & k_- & m_- & -(\mu_{ON} + m_- + k_-) \end{bmatrix} \begin{bmatrix} F_{(OFF,OFF)}(t) \\ F_{(OFF,ON)}(t) \\ F_{(ON,OFF)}(t) \\ F_{(ON,ON)}(t) \end{bmatrix} + \begin{bmatrix} \mu_{OFF} \\ \mu_{ON} \\ \mu_{OFF} \\ \mu_{ON} \end{bmatrix}.$$

This is now a four-dimensional first-order ODE, but the analysis mirrors that of the simpler case above. If $-\phi_i$ and w_i, for $i = 1, \ldots, 4$, are the negative eigenvalues and corresponding eigenvectors of this system matrix, then the solution of this

ODE satisfying the boundary conditions is

$$\begin{bmatrix} F_{(OFF,OFF)}(t) \\ F_{(OFF,ON)}(t) \\ F_{(ON,OFF)}(t) \\ F_{(ON,ON)}(t) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \sum_{i=1}^{4} \beta_i w_i e^{-\phi_i t},$$

where the coefficients β_i are chosen such that $\sum_{i=1}^{4} \beta_i w_i = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$. These conditional CDFs map to the unconditional CDF for T to give

$$F(t) = \mathbb{P}(T \leq t) = 1 - \sum_{i=1}^{4} B_i e^{-\phi_i t},$$

where

$$B_i = \beta_i \begin{bmatrix} \mathbb{P}(OFF,OFF) & \mathbb{P}(OFF,ON) & \mathbb{P}(ON,OFF) & \mathbb{P}(ON,ON) \end{bmatrix} w_i$$

for the probabilities of the stationary distribution of the random walk with transitions given in the model (4.3) above. Again, T has a mean

$$\mathbb{E}(T) = \sum_{i=1}^{4} \frac{B_i}{\phi_i}.$$

Strain or Plasmid	Description	Reference or Source
E. coli		
DH5 alphaZ1	Expression and cloning host	[94]
Plasmids		
pSB50C1	addA Riboswitch:150 bpORF:SFGFP fusion	This study
pSB50C2	addA Riboswitch:SFGFP	This study
pSB50C3	btuB riboswitch:150 bpORF eGFP fusion	This study
pSB50C4	btuB riboswitch eGFP	This study
pSB50C5	addA riboswitch Att1:SFGFP	This study
pSB50C6	addA riboswitch Att2:SFGFP	This study
pSB50C7	addA riboswitch Att3:SFGFP	This study
pSB50C8	addA riboswitch Att4:SFGFP	This study
pSB50C9	addA riboswitch Att5:SFGFP	This study
pSB50C10	btuB riboswitch Att1:eGFP	This study
pSB50C11	btuB riboswitch Att2:eGFP	This study
pSB50C12	btuB riboswitch Att3:eGFP	This study
pSB50C13	btuB riboswitch Att4:eGFP	This study
pSB50C14	btuB riboswitch Att5:eGFP	This study
pSB50C15	(RBS only) Att0:SFGFP	This study
pSB50C16	Att1:SFGFP	This study
pSB50C17	Att2:SFGFP	This study
pSB50C18	Att3:SFGFP	This study
pSB50C19	Att4:SFGFP	This study
pSB50C20	Att5:SFGFP	This study
pSB50C21	btuB riboswitch: Cole9,immunity	This study
pSB50C22	btuB riboswitch:150 bp fusion:Cole9,immunity	This study
pSB50C23	btuB riboswitch:150 bp ATT2:Cole9,immunity	This study
pSB50C24	addA riboswitch:ompT	This study
pSB50C25	addA riboswitch:150 bp fusion:ompT	This study
pSB50C26	addA riboswitch:150 bp ATT2 ompT	This study

Table 4.1: List of strains and plasmids used

Attenuator	repressing region	RBS region	G	N. repressing base pairs
att 1	CCTAA	GGGCCCAAGTTCACTTAAAAAGGAGATCAACTA	-2.9	3
att 2	-	GGGCCCAAGTTCACTTAAAAAGGAGGTCCTCTA	-6.3	3 after RBS core
att 3	TCTCCT	GGGCCCAAGTTCACTTAAAAAGGAGATCAACTA	-7.84	6
att 4	TGATCTCCTTTTTAA	GGGCCCAAGTTCACTTAAAAAGGAGATCAACTA	-20.6	15
att 5	TGATCTCCTTTTTAAGTG	GGGCCCAAGTTCACTTAAAAAGGAGATCAACTA	-25.1	18

Table 4.2: List of ribo-attenuators

4.4 Results

4.4.1 Change in open reading frame demonstrate sensitivity of 5 regulatory elements to secondary structure.

Two previously reported riboswitches were investigated for sensitivity to changes in ORF. The addA riboswitch, characterised from *Vibrio vulnificus* [39, 144], is an activating purine riboswitch [59, 132, 58] reported to selectively bind 2-aminopurine. Fusing sfGFP to the first 150 base pairs of the previously reported working ORF [39] yielded an induction response (to 2-aminopurine concentrations between 0 and 250 μM) very similar to previously reported studies [39]. However, directly replacing the reported ORF with sfGFP (start codon for start codon) gave no induction response. These data were supported by single cell analysis, which demonstrated a response to the inducer for the fusion construct, but no such output when the ORF was completely substituted as seen in Figure 4.4A.

The second riboswitch is the btuB riboswitch, a repressive riboswitch notable for its response to the high-value small metabolite adenosylcobalamin, an active form of vitamin B_{12} [52]. As with the addA riboswitch, replacing the working ORF with eGFP (start codon for start codon) resulted in a loss of induction response (between 0 and 50 nM of adenosylcobalamin) for both colony and single-cell measurements. However, fusion of eGFP to 150 base pairs of the working ORF yielded the desired repressive induction response as seen in Figure 4.4B.

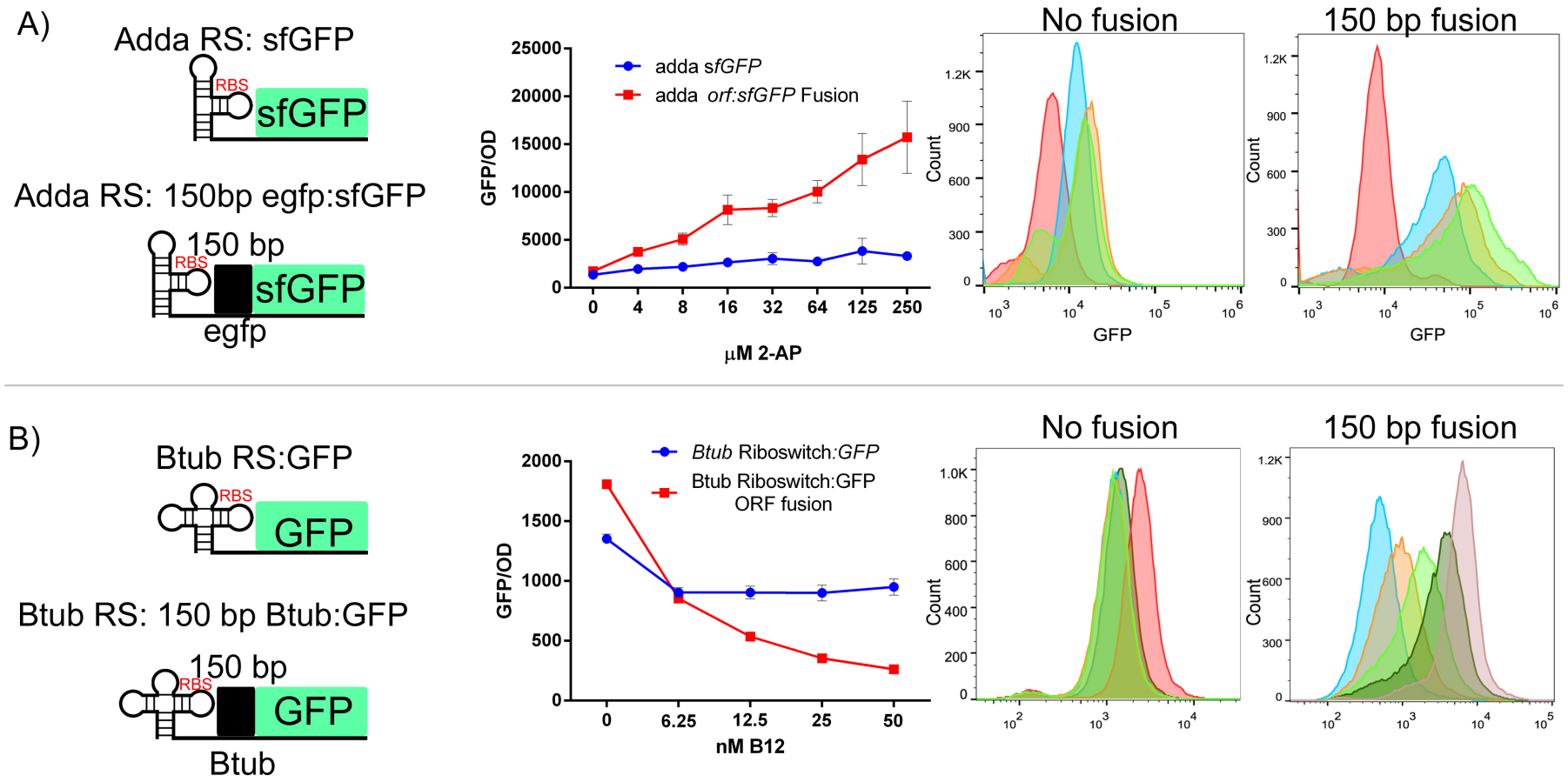


Figure 4.4: **addA and btuB riboswitches exhibit contextual sensitivity:** (A) The Adenine riboswitch had sfGFP introduced directly after the riboswitch replacing its reported ORF (previously eGFP), and also fused to the first 150 bp of the previous ORF. Both constructs were analysed for population and single cell response to 2-aminopurine. Error bars indicate standard deviation of measurements for biological triplicates. Single cell colours; Red (0 μM) Blue (8μM) Orange (32μM) Green (250μM). (B) The btuB riboswitch had eGFP introduced directly after the riboswitch replacing its reported ORF (previously btuB), and also fused to the first 150 bp of the previous ORF. Both constructs were analysed for population and single cell response to adenosylcobalamin. Single cell colours; Red (0 nM) Dark Green (6.25 nM) Light Green (12.5 nM) Orange (25 nM) Blue (50 nm).

4.4.2 Design of ribo-attenuators

Ribo-attenuators were designed within the addA riboswitch system, placed between 150 bp of the system's previous working ORF (eGFP) and the new gene of interest (sfGFP), as illustrated in Figure 4.4. This particular length of ORF (150 bp) was selected by trading-off the requirement that it be long enough to include the secondary structure of most riboswitches (which commonly extends 100bp into the open reading frame [24]), but short enough to minimise the overall size of the system. A common RBS region was designed to include the SD1 RBS from the BIOFAB parts library [105], which was placed upstream of a transcriptionally coupled junction (TAATG) that included a stop codon in-frame with any ribosome originating from the riboswitch, and a negative one base shifted start codon. Hairpins with ever lower ΔG were introduced over the RBS on the 5′ end by altering the attenuator's repressing region as shown in Figure 4.5. For one attenuator (Att2) I took a different approach, engineering a hairpin on the 3′ side by altering bases within the common RBS region. Nine potential hairpins with stems of between 3 and 24 bp in length were screened and omitted hairpins are shown in Figure 4.6. From these I selected five (labelled Att1-5) that gave a range of output responses. Further ribo-attenuators could easily be designed rationally around a new RBS by following this process, or by generating random sequences and screening for a ribo-attenuator of the required strength. The selected ribo-attenuators (without upstream riboswitch) were expressed under the control of the tetracycline promoter, demonstrating that with the exception of the smallest hairpin (Att1), the hairpin structures caused a significant drop in the RBS's translational efficiency as seen in Figure 4.5. Thus, when a riboswitch is placed upstream of the ribo-attenuator, translation from the attenuator's RBS independent of riboswitch activity will be minimal.

To demonstrate the universal applicability of ribo-attenuators, the addA riboswitch

was used in isolation, as a 150 bp fusion with its working ORF (eGFP), and with Att2, to express and analyse the targeting of OmpT, an outer membrane protein. For all constructs basal expression of OmpT was observed 4.8. The riboswitch in isolation failed to produce membrane bound OmpT, whilst the fusion system yielded mainly insoluble inclusion bodies. However, the ribo-attenuated system yielded correctly targeted OmpT in the cell membrane, demonstrating the ribo-attenuator's ability to express functional protein.

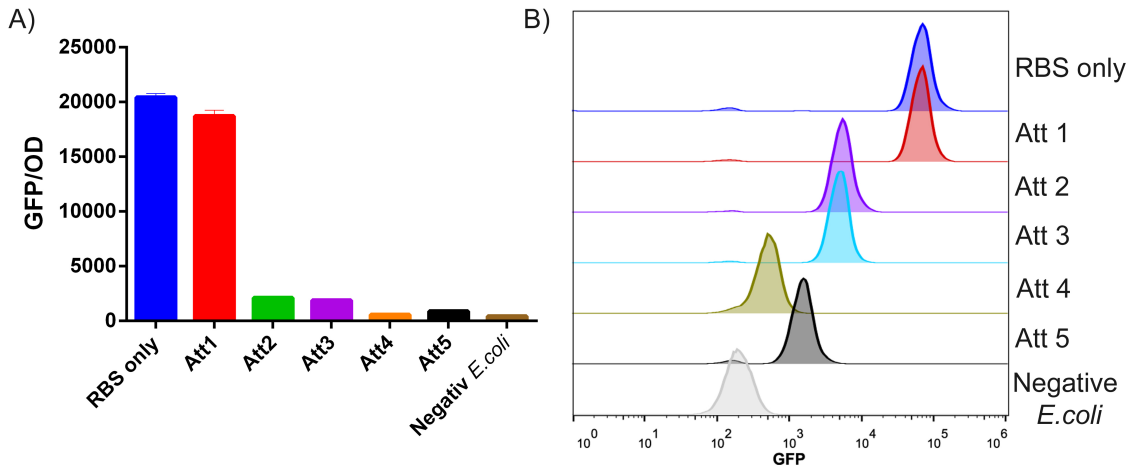


Figure 4.5: **Repressive effects of the Ribo-attenuators without any riboswitch** Each attenuator was introduced between the tetracycline promoter and sfGFP to assess its basal expression rate in the absence of an upstream riboswitch. A positive control (RBS only) was created by including only the Common RBS region without the upstream repressing region. Introduction of the repressing regions resulted in a reduction in expression on a population (A) and single cell (B) level, though only a minor reduction was demonstrated by Att1 due to its weak hairpin. Error bars indicate standard deviation of measurements for biological triplicates.

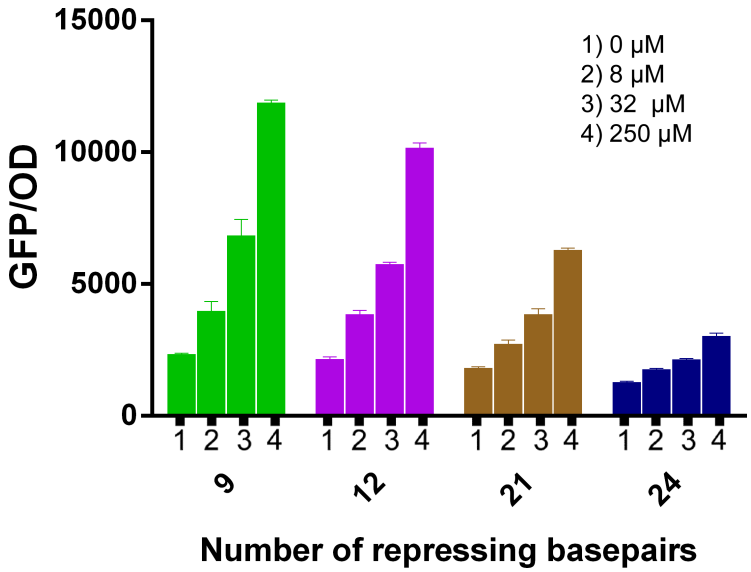


Figure 4.6: **Additional ribo-attenuator screening data:** Response of ribo-attenuators not used in the main chapter to varying levels of induction by 2-aminopurine. This experiment was performed using the above ribo-attenuators (which were designed to have hairpins consisting of 9, 12, 21, and 24 repressing bp respectively), as well as Att1 (3 repressing bp), Att2 (4 bp), Att3 (6 bp), Att4 (15 bp), and Att5 (18 bp), for which data is presented in Fig. 4. Att1-5 were selected from the nine candidates to provide a wide range of induction response behaviours. Error bars indicate standard deviation of measurements for biological triplicates.

4.4.3 Ribo-attenuators can shift addA riboswitch induction response.

Five ribo-attenuators were introduced downstream of the adenine riboswitch and 150 bp of its working ORF (eGFP). The ribo-attenuators were able to shift the system's induction profile in a manner that correlated with the strength of the attenuator; a stronger hairpin leads to a narrower and lower induction response (Figure 4.7A). The direct fusion of sfGFP to the working ORF yielded very clear inclusion bodies, manifested as distinct spots present at one pole of the cell as seen in Figure4.7B, which are formed by misfolded insoluble proteins.

By comparison, the inclusion of a ribo-attenuator between the riboswitch and sfGFP yielded soluble protein (Att1 shown) since the translated reporter gene is

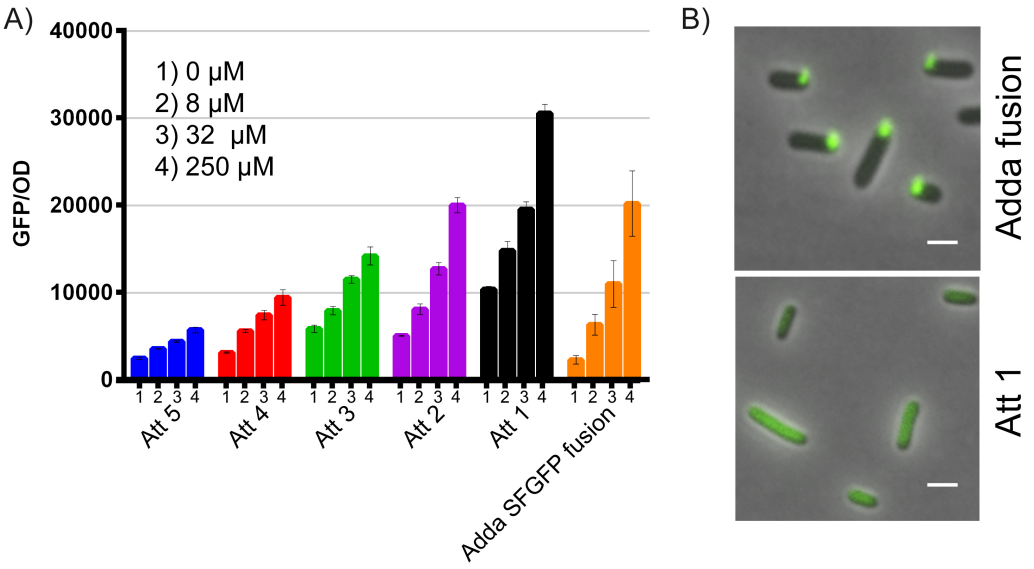


Figure 4.7: **Ribo-attenuators allow addA riboswitch rational re-engineering:** (A) Response of each ribo-attenuator to 0, 8, 32 and 250 μM 2-aminopurine as compared to the fusion construct (orange). Error bars indicate standard deviation of measurements for biological triplicates. Full single-cell data is presented in Figure 4.11. (B) Fluorescence microscopy of the addA fusion showing clear inclusion bodies and the Att1 ribo-attenuator showing soluble dispersed GFP. White bars represent 1μm.

expressed orthogonally from the upstream 150 base pairs of the working ORF. The absence of the fusion domain in the ribo-attenuated case was confirmed via immunoblotting, which demonstrated a size difference between the protein products in systems with and without the attenuator as shown in Figure 4.9.

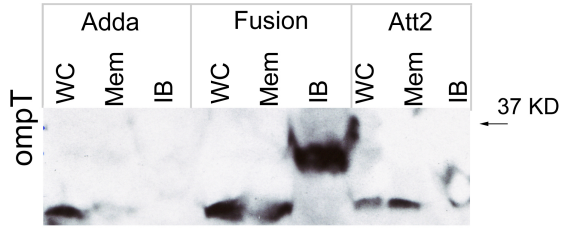


Figure 4.8: **OmpT targeting analysis:** Immunoblot of strep-tagged OmpT expressed from the addA riboswitch directly, as a fusion with the first 150 bp of the addA system's working ORF, and with Att2 isolating OmpT from the fusion domain. To assess targetting of OmpT, membranes (Mem) and inclusion bodies (IB) were separately dissolved on the blot demonstrating that the fusion approach to expression leads to a prevalence of insoluble inclusion bodies, whereas the ribo-attenuator system promotes proper membrane targetting. Whole cell (WC) fractions were included to demonstrate the basal expression of the system. Prepared by Barbara Mertins

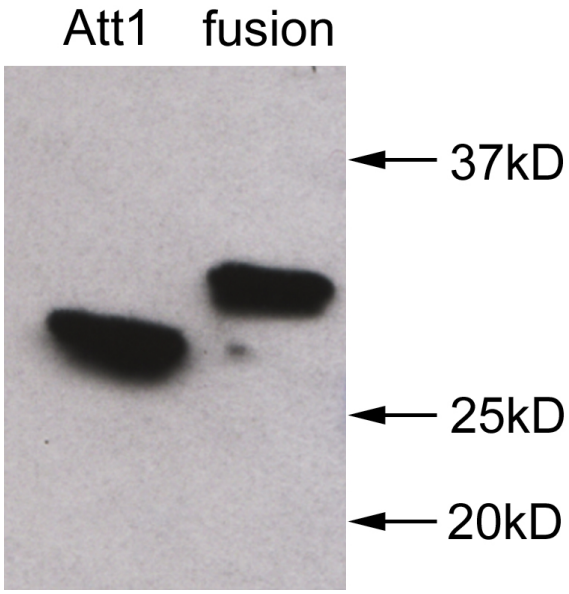


Figure 4.9: **sfGFP size analysis:** Immunoblot of sfGFP translated from the addA riboswitch with the Att1 ribo-attenuator, and as a direct fusion to the first 150 bp of the riboswitch's working ORF (eGFP). The effectiveness of the transcriptionally coupled junction **TAA**TG is observed in the approximate 5 kD difference between constructs, demonstrating that the attenuator separates the introduced gene of interest from the 150 bp fusion. Immunobloting was performed as described in the materials and methods using a monoclonal GFP antibody (Clonetech). Done by Barabara Mertins

4.4.4 Ribo-attenuators can shift and amplify btuB riboswitch induction response.

The same five ribo-attenuators were introduced downstream of the btuB riboswitch and 150 bp of its working ORF (btuB). The ribo-attenuators were able to shift and substantially amplify the system's induction profile, with (as for the addA riboswitch) increased hairpin strength corresponding with reduced expression as shown in Figure 4.10A. Introduction of the attenuators resulted in a notable difference between the activated and unactivated systems when observed by fluorescence microscopy as seen in Figure 4.10B. The use of the same five ribo-attenuators demonstrates the modularity of our system: Despite these attenuators being designed within a different genetic context (the addA riboswitch system), when transplanted into the btuB system their functionality is maintained.

To demonstrate further application of ribo-attenuators, the btuB riboswitch was used in isolation, as a 150 bp fusion with its working ORF (btuB), and with Att2, to express ColE9, a bacteriocin that is imported into the cytoplasm (via OmpF and btuB) and degrades the genome of *E. coli* cells [81], killing the cells. The amount of active ColE9 in a sample can be directly correlated with the clearance of *E .coli* grown on a soft agar plate when ColE9 is spotted onto it [131]. As in Fig. 4.10B, it was found that inserting ColE9 directly after the riboswitch resulted in minimal expression as seen in Figure 4.10C. Expression was increased by a factor of $\sim$ 1000 in the fusion system, however, with the inclusion of Att2 a substantially higher yield of active protein was achieved: Clearances were observable up to dilution by a factor of 10^{-10}, demonstrating the attenuator's ability to significantly enhance protein production from this riboswitch.

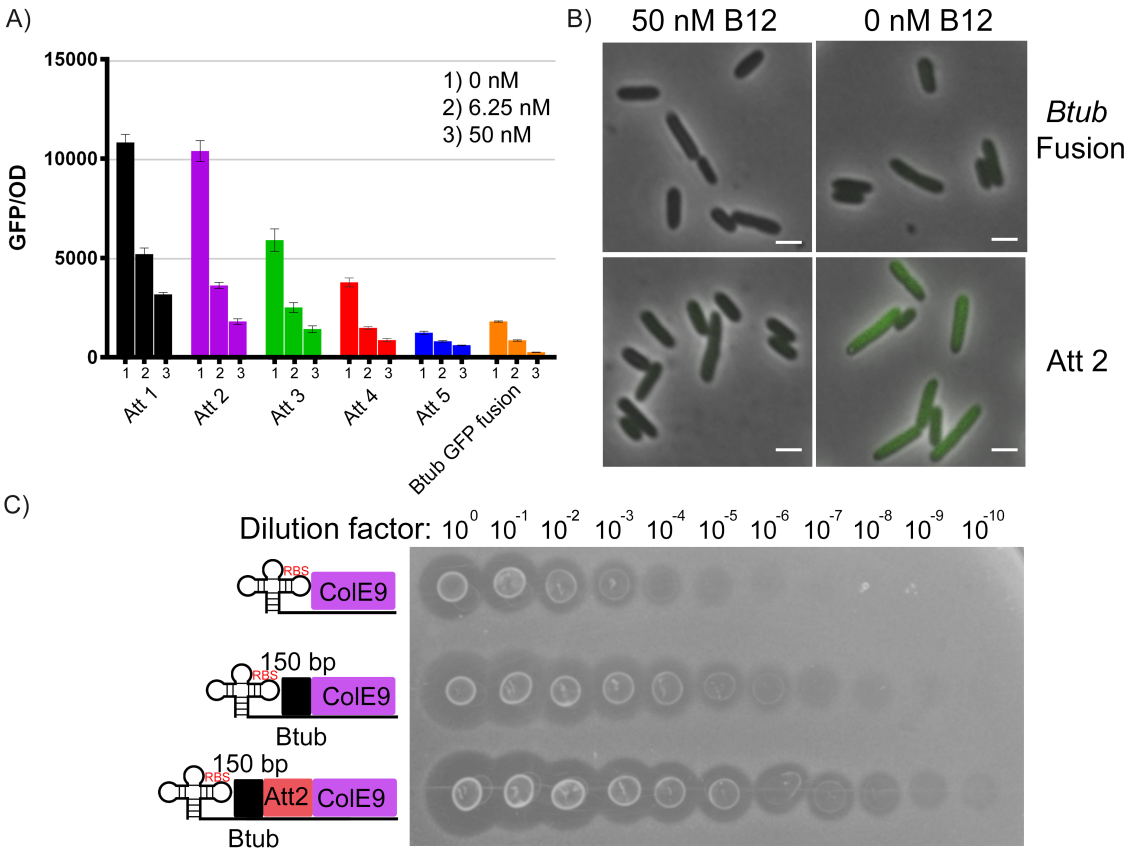


Figure 4.10: **Ribo-attenuators allow btuB riboswitch rational re-engineering:** (A) Response of each ribo-attenuator to 0, 6.25, and 50 μM 2-aminopurine as compared to the fusion construct (orange). Error bars indicate standard deviation of measurements for biological triplicates. Full single cell data is shown in Figure 4.11. (B) Fluorescence microscopy of the uninduced and induced btuB fusion and Att2 ribo-attenuated btuB riboswitch, with the latter demonstrating a greater induction difference. White bars represent 1μm. (C) Plate-killing assay for production of bacteriocin ColE9 from the btuB riboswitch with and without fusion and attenuator domains. Clearances in the *E. coli* soft agar plate correlate with ColE9 production level, demonstrating a substantially increased protein yield in the ribo-attenuator system.

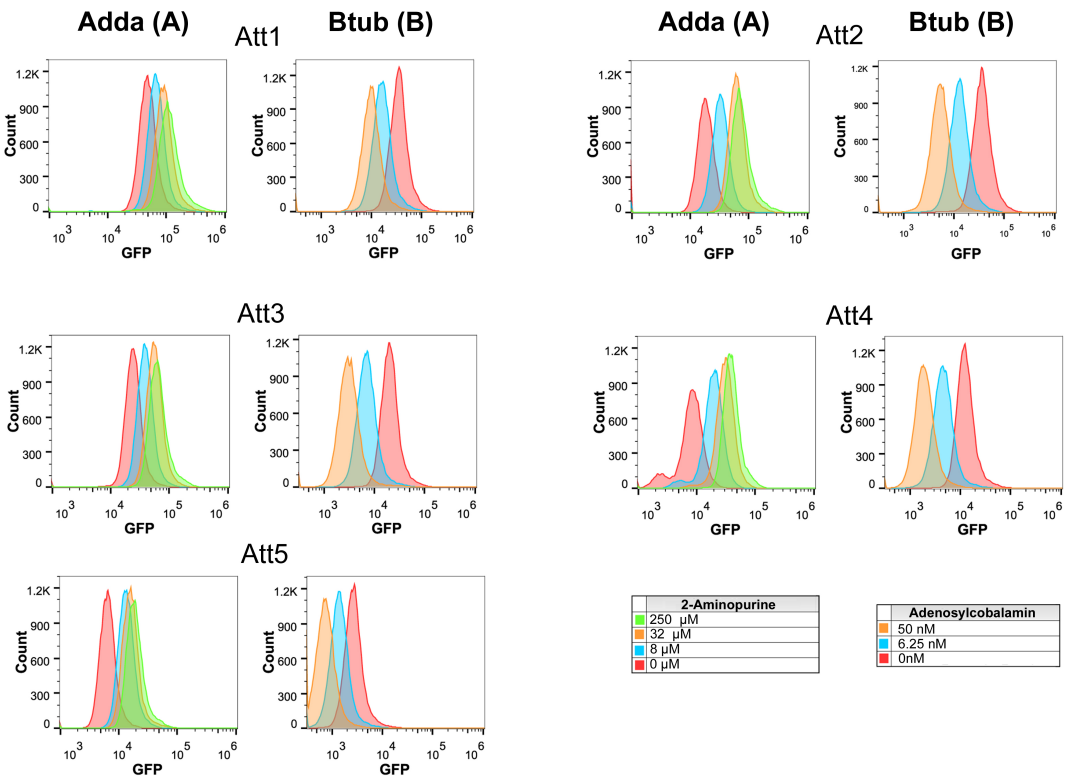


Figure 4.11: **Full single cell analysis for all Ribo-attenuators** Single cell analysis for both addA and btuB riboswitches over reported inducer concentrations for all Ribo-attenuators used.

4.4.5 Ribo-attenuators reduce uncertainty of riboswitch function

The translation of GFP from mRNA in the modelled riboswitch system at any given time depends on the state of the riboswitch, either λ_{ON} or λ_{OFF} (Fig. 4.12A). Over time the riboswitch stochastically switches between ON and OFF, with inducer concentration biasing the stochastic switching towards one or the other state. Assuming that the two translation rates are at different orders of magnitude, the waiting time from any given time until the production of a GFP molecule is thus highly uncertain. In the modelled ribo-attenuator system, the accessibility of the two RBSs can be in one of four states (OFF, OFF), (OFF, ON), (ON, OFF), and (ON, ON) (Fig. 4.12A). The rates λ_{ON} and λ_{OFF} of ribosome transit now determine the rates at which the second switch turns on, given the

state of the first. The scale separation of those rates means that, to a first order approximation, the first switch ON implies the second switch will turn ON, and the first switch OFF implies the second switch is likely to turn OFF. However, the GFP translation rates μ_{ON} and μ_{OFF} resulting from the state of the second switch can now be set independently of the translation parameters of the upstream riboswitch via adjustment of the ribo-attenuator's RBS. Expression level is also dependent upon the rate at which the ribo-attenuator switches back to the OFF position (m_(L)), which is itself a function of the length (L) and hence strength of the attenuator region's secondary structure.

From this stochastic model, probability distributions for total GFP expression were generated, demonstrating a substantial reduction in gene expression variability (at maximal induction) for the ribo-attenuator system when compared to the un-attenuated riboswitch (Fig. 4.12B), a trend which is maintained over a range of induction levels (Fig. 4.12C). The same behaviour was observed experimentally for both the addA and btuB riboswitches (Fig. 4.13), which demonstrates that inclusion of a ribo-attenuator substantially reduces the width of the expression distribution. Figure 4.14 displays the single cell distinction between uninduced and fully induced populations for the addA and Btu riboswitchs with and without the ribo-attenuator, demonstrating that the reduced variation in the attenuated system makes overlapping populations substantially more distinct.

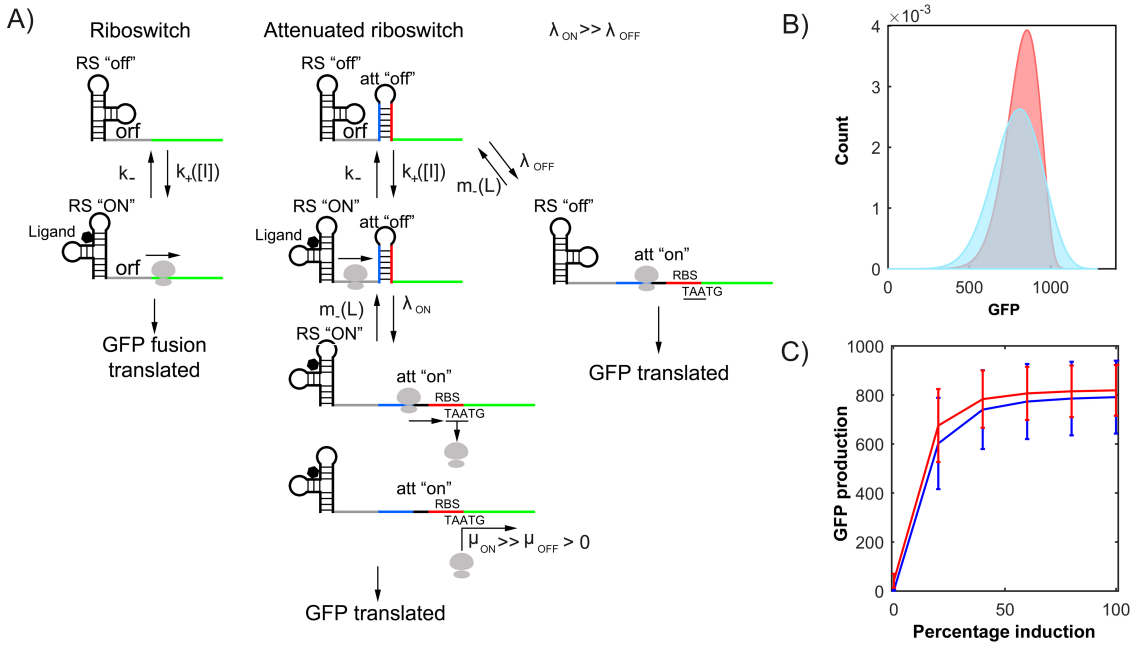


Figure 4.12: **Ribo-attenuators reduce riboswitch expression noise:** (A) Random walk dynamics of the modelled system. The riboswitch randomly switches between OFF and ON states at rates $k_+[I]$ and k_-, and GFP is produced at rate λ_{OFF} or λ_{ON} depending on the state of the riboswitch. In the ribo-attenuator system, the riboswitch has similar dynamics; the rates λ_{OFF} and λ_{ON} are now the rates at which the downstream attenuator is switched ON. The downstream attenuator can spontaneously switch off at rate $m_-(L)$, which is an increasing function of the length L of the attenuator's hairpin stem. GFP is produced at rates μ_{OFF} and μ_{ON} depending on the state of the downstream attenuator. (B) Probability distribution for the random variable defined as the number of GFP molecules expressed by the stochastic model in (A) over a unit time interval starting from an arbitrary (long-term) time-point at maximal (100%) induction for the riboswitch system (blue) and the ribo-attenuator system (red). (C) The expression distribution mean plus/minus one standard deviation (error bars) for varying induction concentrations defined as a percentage of maximal induction (the data points at 100% induction therefore correspond to the data in (B)). Comparison of the riboswitch (blue) and ribo-attenuator system (red) demonstrates substantially reduced variability across the induction range.

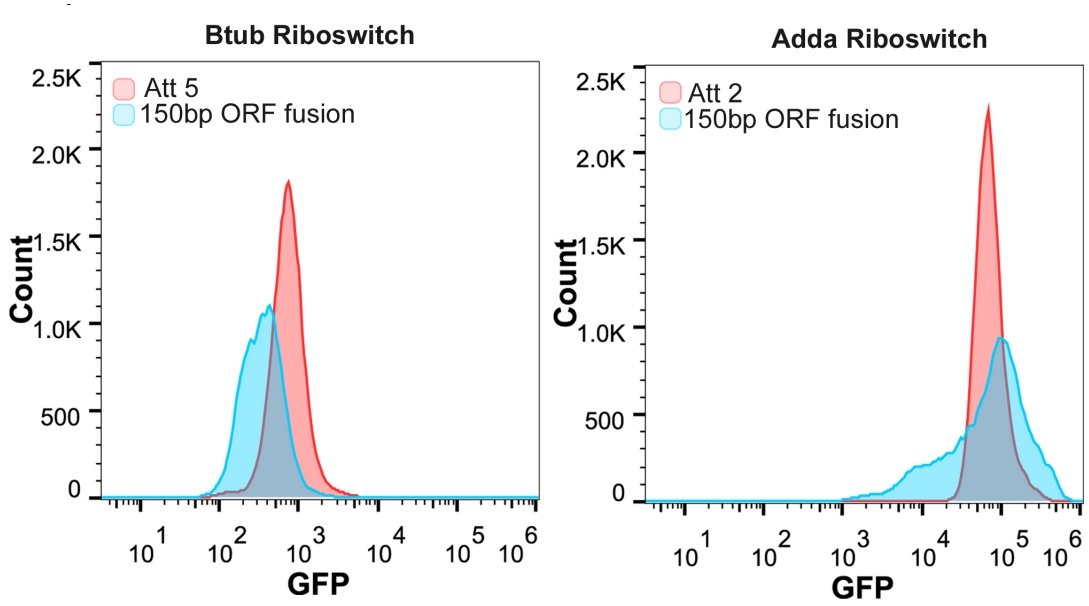


Figure 4.13: **Single cell comparison between similar expressing ribo-attenuators and the fused constructs** Overlay of the btuB riboswitch with a functional fusion (blue) and with att5 (red) both in the presence of 50 nM B12, and overlay of the addA riboswitch with a functional fusion (blue) and Att2 (red)both in the presence of 250 M 2-aminopurine

4.5 Discussion

We designed a novel and widely applicable tool that allows riboswitches to be used as modular, tunable components with a highly predictable outcome, enhancing applicability by overcoming previously identified drawbacks. I chose two well known, representative riboswitches; the addA riboswitch, an activating riboswitch that respond to a cellular metabolite (Adenine), and the btuB riboswitch, a repressing riboswitch which responds to adenosylcobalamin. These have the potential to facilitate the biological production of high value micronutrients such as vitamins, many of which are still produced via expensive and inefficient chemical synthesis. Recently described methods applying riboswitches to high value vitamins could also be used with the ribo-attenuators described herein to elicit a high throughput approach to pathway elucidation [183].

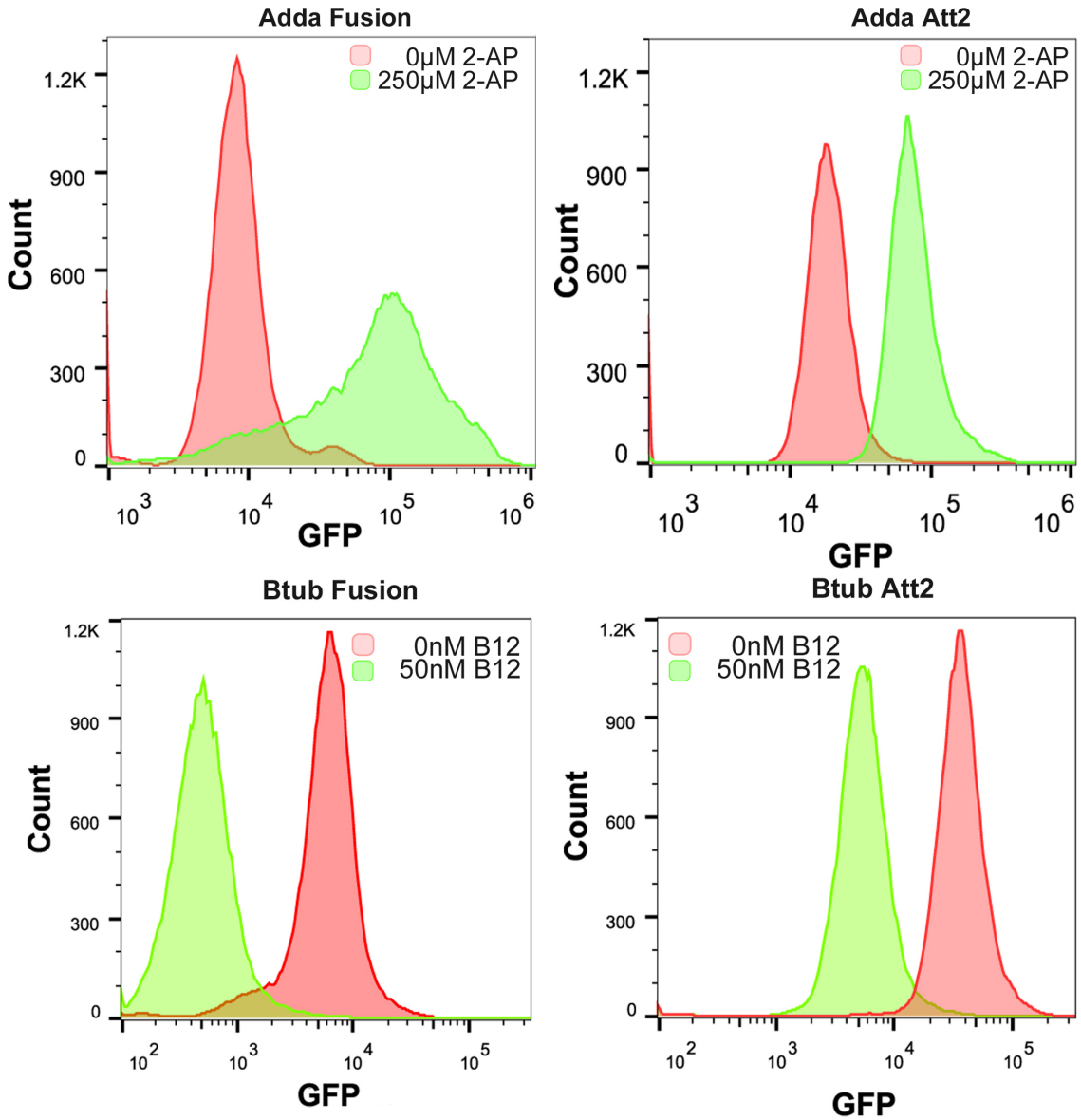


Figure 4.14: **Single cell distinction between induced and un-induced riboswitches** Single cell distinction between the lower and upper limits of the addA and btuB riboswitches fused to GFP and with the Att2 Ribo-attenuator showing more distinct populations with the addition of the Ribo-attenuator.

The first advantage of the developed ribo-attenuator system is that it enables riboswitch-controlled expression of a new gene of interest without the inclusion of a 5′ fusion. Typically such fusions have been employed [52, 111, 168, 40] be-

cause direct replacement of a riboswitch's ORF can degrade its functionality [24] (as demonstrated in Fig. 4.4). However, re-engineering a riboswitch in this fashion gives rise to a protein with a large fusion, which can severely impact function. This was demonstrated in Fig. 4.7B, where the fusion-based approach led to formation of inclusion bodies. As a second example, expression of the outer membrane protein OmpT demonstrated that fusion to the first 150 bp of btuB resulted in formation of inclusion bodies, whereas the attenuator-produced protein was correctly targeted to the membrane. I hypothesise that inclusion bodies may form due to the fusion domain causing protein mis-folding. Alternatively, in the case of OmpT the fusion domain's presence may disrupt the signal sequence required for membrane transport [84]. Ribo-attenuators overcome this challenge by translating a gene of interest from a second downstream RBS, avoiding inclusion of any 5′ fusion and thereby improving functionality.

A second benefit provided by the inclusion of a ribo-attenuator is the ability to modularly tune the induction response of a riboswitch. This was demonstrated by ribo-attenuators shifting and re-scaling the output response of the addA system in which they were designed (Fig. 4.7A), in a manner correlating with *a priori* predictions of their relative strength. This functionality was maintained when the same ribo-attenuators were employed in a different genetic context in the btuB system (Fig. 4.10A), demonstrating their modularity. In both contexts an increased expression level was observed (compared to the fusion system) when the riboswitch was in its repressing induction state. This may arise due to spontaneous accessibility of the ribo-attenuator's RBS, or re-initiation of terminated ribosomes at the ribo-attenuator's start codon due to 70s scanning [174]. Furthermore, an increase in repressed expression level could occur due to the fact that any ribosome originating from the riboswitch RBS would, in the fusion case, produce a single copy of GFP, whilst in the ribo-attenuator system it would open

the attenuator's hairpin for a time period during which multiple translation initiation may events occur. Particularly in the btuB system, ribo-attenuators were able to substantially increase the maximal induction output (for both GFP and ColE9). This is an important capability [43] that would be difficult to achieve via re-engineering of the riboswitch's RBS (which could influence its ligand-binding ability), or via increasing translational promoter strength (our system already employed the strong tetracycline promoter at full induction). Though attempts have been made to overcome these problems, such as boosting riboswitch efficiency through RNA amplification [43] or *de novo* design of synthetic riboswitches [45], as-yet they do not provide a reliable, modular, and predictable approach to tuning of riboswitch induction. The ribo-attenuator system described here achieves this, and new variants can be developed with different RBS strengths and easily predicted secondary structures, whilst preserving riboswitch ligand response.

A third advantage provided by ribo-attenuators is their ability to reduce variability of a riboswitch's output. This limitation had restricted the application of riboswitches in a number of ways: For example, unpredictable behaviour may produce downstream variance in processes under a riboswitch's control, whilst also making separations of populations by FACS challenging, limiting the ability to screen libraries for novel compounds. The modelled ribo-attenuator system provided a theoretical underpinning for its noise-reduction capability, demonstrating that inclusion of a ribo-attenuator can cause a substantial reduction in population expression variability, as well as a slight increase in mean expression (Fig. 4.12B). This was observed experimentally (Fig. 4.13), demonstrating that the reduction in variability can provide more distinct output distributions (Fig. 4.14).

In conclusion, by fine-tuning sensitivity, and reducing uncertainty in expression

thereby separating cell populations, ribo-attenuators facilitate the use of riboswitches in applications ranging from FACS screening of component libraries to rapid elucidation of novel biosynthesis pathways. Initially I examined the sensitivity of two well-known riboswitches to the introduction of a new open reading frame: Introducing a novel reporter as part of a fusion yielded functional riboswitches, but inclusion bodies that significantly limit application were observed. I also found a high degree of variation in the addA riboswitch, which made identification of specific concentrations of ligand difficult. Using ribo-attenuators I allowed both riboswitches to maintain the original open reading frame and as a result their ligand response, whist expressing the introduced gene of interest orthogonally over a predictable dynamic range. Using these ribo-attenuators it should be possible to overcome limitations to riboswitches identified in previous studies. For example, the Glycine riboswitch is a similar riboswitch from *Bacillus subtilis* identified as having an unsuitable induction range [119], which limits its use as a tool for maintaining vectors in the absence of an antibiotic, or as a cheap Glycine-based induction system for bioreactors. As shown in Fig. 4.10 for the btuB riboswitch, ribo-attenuators could amplify and tune this system's static response, aiding its industrial application. In summary, ribo-attenuators represent a breakthrough step towards allowing riboswitches to be treated as truly modular and tunable devices. Moving forward I look at reducing noise in biological devices other than riboswitches through the application of feedback.

Chapter 5

A synthetic viral feedback loop results in robust expression

A Papachristodoulou devised the approach. The experiments were designed and conducted by me. The principle of the model, the model parameters, mathematics and its computation was done by Thomas Prescott and Antonis Papachristodoulou. I am grateful to Dr Drew Endy from Stanford University for providing the excisionase sequence.

'"A synthetic viral feedback loop results in robust expression"' was submitted to Nature Scientific Reports on the 5^{th} of October 2016 (currently in review).

5.1 Aims

Accurate control of a biological process is essential for critical functions in biology, from the cell cycle to proteome regulation. Many of these events use a form of feedback to achieve a highly robust and reliable output. Feedback is found throughout biology and technology, but due to problems caused by low tunability and complexity of implementation, it is yet to be widely adopted in synthetic biology. The aims of this chapter are:

1. To design a tunable synthetic feedback network using a class of viral proteins called integrases with excisionases, which can flip the orientation of DNA segments in a digital manner.

2. To characterise the novel feedback network in terms of its ability to reduce noise within a system.

5.2 Introduction

Many important biological processes, such as cell cycle control, use feedback to ensure highly robust and reliable operation. Feedback control is employed extensively in engineered systems that need regulation. By using measurements of a system's output to influence its input (known as *closing the loop*) it is possible reduce a system's response time, its input-output gain, and the dependence of its response to certain system parameters, which makes its overall performance more robust to fluctuations in both the system's properties and its environment [2, 157, 115]. These advantages have motivated researchers to implement similar features in synthetic biological systems [69, 31, 149, 11, 16]. The use of transcription factors to regulate expression dynamics is a long-established approach to implementing feedback [75, 136, 96, 33, 49, 109, 108], however, this approach has potential limitations. For example the reliance of a closed-loop system on endogenous host proteins can result in cross-talk, which arise from interference between the constituent regulators and other cellular processes [22]. Furthermore, the resulting closed-loop system may have very low gain (depending on the repression strength of the regulating transcription factor), which can be undesirable and can make tuning of downstream processes challenging. One approach to overcoming the low gain of negative feedback systems is to redesign protein expression rates to compensate for the reduced steady state [136], however, this leads to further design challenges and uncertainties. Thus, it can often be very difficult to engineer a predictable orthogonal output using genetic regulation. The challenges posed by the implementation of the feedback systems discussed result in a severe under-utilisation of this otherwise clearly advantageous principle, widespread in natural and engineered systems, and as such there is significant need for a robust, orthogonal, and easily applicable feedback architecture.

Serine integrases and their associated excisionases are a class of viral proteins that mediate integration and excision of the phage genome during the lysogenic and lytic phases of the phage life cycle [153, 15, 77, 71]. Integrases allow phages to integrate their genome into the host cell and enter the lysogenic phase. They catalyse the insertion of DNA through the recognition of specific $\sim$50 base pair long binding sites referred to as attB and attP (attachment Bacteria and attachment Phage respectively). Integrases from many phages, such as TP901-1, Bxb1 and ϕC31, recognise orthogonal attB and attP recognition sites via a helix-turn-helix domain [103]. The process of integration occurs via strand cleavage, inversion and ligation to form attL and attR (attachment Left and attachment Right respectively) sites [62]. A central catalytic domain with a conserved serine residue catalyses this reaction [62].

Excisionases, via interaction with integrases, allow phages to enter the lytic phase of the life cycle. Passage from lysogenic to lytic phase is usually controlled by the cleavage or phosphorylation of a transcription factor, which flips the bias of a high gain bistable switch towards the expression of the system's integrase and excisionase proteins. The expression of an excisionase cofactor allows the integrase-excisionase complex to recognise attL and attR sites and subsequently catalyse the reverse reaction back to attP and attB [140, 54]. Research has shown a potentially symbiotic nature in some such phage–bacterium interactions: For example, in A118-like prophage the phage regions are excised, remain non-integrated yet also non-lytic, and may subsequently be re-integrated into the host genome to regulate host genes depending on the state of the cell and its environmental conditions [48].

Integrases and excisionases have been shown to mediate site-specific integration,

excision, and inversion depending on the orientation of their binding sites [62], which has led to a particular focus on their characterisation and application in the field of Synthetic Biology. This class of proteins confers the full range of logic functions at the genetic level [13, 152] (XOR, AND, OR, XNOR etc) and allows for a wide range of applications such as highly sensitive medical assays in healthcare [177, 32, 12, 47] as well as metabolic pathway assembly [30]. Given the digital and high gain nature of integrase devices, their use in Synthetic Biology tends to require careful tuning of system parameters to avoid a high level of output uncertainty, which arises due to variability in the integrase-mediated flipping of DNA regions, both between cells and between cell populations [12]. As a consequence the expression level of a particular protein of interest in an integrase circuit is often suboptimal and substantially variable, a problem currently limiting the wide application of integrases as predictable and robust biological components for biomanufacturing, healthcare, or environmental remediation systems.

Here we present an engineered synthetic feedback loop based on an integrase and excisionase system which could be applied in a diversity of biological settings. We re-design a previously reported DNA register [12] (which expressed Green Fluorescent Protein (GFP) in response to DNA inversion) to co-express excisionase with the protein of interest, in our case a fluorescent reporter, as illustrated in Figure 5.1. The excisionase closes the loop around the high gain integrase system by enabling registers in an ON state to stochastically and transiently switch to an OFF state. Over a sufficiently long time period the distribution of registers in either state stabilises, which due to the coupling of excisionase and GFP production results in a reduced variation in GFP expression between cells (both in absolute terms, and when normalised by mean expression levels) when compared to an open-loop system (which shows substantial variability due to gene expression noise). In both cases variability arises due to cell-to-cell heterogeneity, for exam-

ple due to variations in plasmid copy number or differences in proteome. In the closed loop case, as will be demonstrated later, the correlation of excisionase and GFP production results in a reduced variation in GFP expression when compared to the open loop systems.

5.3 Methods

5.3.1 Plasmids

The target plasmid containing GFP flanked by inverted Bxb1 recognition sites in the opposite orientation to the P7 constitutive promoter [105] was used as a control without feedback (P15A origin, 15-20 copies. The excisionase gene [12] with a tuned ssrA degradation tag was synthesized by IDT and inserted after the GFP output by Gibson assembly [57, 56]. The second plasmid from [13] containing Bxb1 integrase under control of the pTet promoter was co-transformed in each experiment to provide Bxb1 integrase expression. Plasmid sequences are available online [51].

5.3.2 Cell culture and experimental conditions

All experiments were performed using DH5αZ1 [94]. For each experiment three colonies were picked from freshly transformed plates into LB media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) [7] with appropriate antibiotics and grown overnight, each was then diluted 1/1000 with reported concentrations of inducer and appropriate antibiotics. For end point experiments (Figure 5.3 and Figure 5.6) cultures were grown for 12 h at 37°C and 225 RPM in 1.5 mL deep well plates (Fisher). For time course experiments 200 μL microcultures were prepared in 300 μL black well, clear bottom plates (Fisher) and run for 12 h in a BMG clar-

iostar platereader with a gain of 1000 at 37°C and shaken at 200 RPM. Antibiotics used were carbenicillin (100 μg/mL) and chloramphenicol (25 μg/mL) (Sigma). Anhydrotetracycline (Sigma) was used in reported concentrations.

5.3.3 Measurement and data analysis

End point experiments were centrifuged and re-suspended in PBS [42]. Individual GFP and Optical Density (OD) OD600 measurements were taken using a Clariostar plate reader with a gain of 1000 before the data was normalised by OD600. Time course data was collected using sequential measurement of GFP and OD600. Flow cytometry was performed on an Attune flow cytometer; for each sample 50,000 cells were recorded. GFP intensities were quantified by removing any data points below the negative control population, to remove cell fragments and other artefacts.

5.3.4 Mathematical modelling

Considering our system at the single-cell level, we assumed that each cell contains D copies of the plasmid depicted in Figure 5.1, where the attB and attP integrase recognition sites are separated by either a reporter coding region (i.e. open-loop), or both reporter and excisionase coding regions (i.e. closed-loop).

Integrase is expressed on a different plasmid of copy number N, under the control of aTc. We modelled the dynamics of integrase concentration as the net balance of its expression under the (saturating) induction of aTc with degradation and dilution. This results in the deterministic differential equation

$$\frac{\mathrm{d}[I]}{\mathrm{d}t} = N\mu_I \frac{[aTc]}{K_I + [aTc]} - \delta_I[I],$$

where $[I]$ and $[aTc]$ denote the concentrations of integrase and aTc respectively. Here μ_I and δ_I are parameters representing the maximal integrase expression per plasmid, and the degradation/dilution rates of integrase molecules respectively. The parameter K_I is the dissociation constant for aTc promoting integrase expression.

Similarly to integrase expression above, when d of the D plasmids have flipped to the ON state, we have expression of GFP and (in the closed-loop case) excisionase, and also degradation of each. We thus model the concentrations $[GFP]$ and $[X]$ of GFP and excisionase respectively by the differential equations

$$\frac{\mathrm{d}[X]}{\mathrm{d}t} = d\alpha\beta\mu_X - \delta_X[X], \quad \frac{\mathrm{d}[GFP]}{\mathrm{d}t} = d\alpha\beta\mu_G - \delta_G[GFP], \tag{5.1}$$

where the scaling d of the expression rates μ_X and μ_G reflects an assumption that the expression of excisionase and GFP is linearly proportional to the number of ON registers. The combined parameter $\alpha\beta$ is a cell-specific scaling factor for expression rates, chosen as discussed in the following section, which models the cell-to-cell uncertainty in protein expression around the nominal rates μ_X and μ_G. Note that setting $\mu_X = 0$ models the open-loop behaviour, while $\mu_X > 0$ models closed-loop behaviour.

Consider now the closed-loop circuit. Presence of integrase causes the D registers to begin to flip from OFF to ON. When d of the D registers have flipped ON, both GFP and excisionase are expressed. Integrase and excisionase together cause some of the d ON registers to flip back to OFF, thus slowing expression of GFP and excisionase. Hence, in addition to the linear deterministic models described above, we also need to model the stochastic, digital behaviour of the D DNA registers. The variable $d(t)$ representing the number of ON registers at time t takes a random walk on the set of integers $d \in \{0, 1, \ldots, D\}$. Given $d(t)$, the rate

at which d jumps to $d+1$ is modelled as

$$r_{ON}(D-d)\frac{([I]/D)^4}{K_{ON}^4+([I]/D)^4}.$$

This rate is constructed as a scalar parameter r_{ON}, representing a switching rate, multiplying the expected number of OFF registers that have two integrase dimers bound, thereby being in the configuration able to stochastically switch to the ON state. In particular, the integrase concentration-dependent fraction in this rate is a Hill function with parameter 4, representing the fact that a total of four integrase molecules need to bind to the DNA to be in the correct configuration to switch the register to the ON position. The parameter K_{ON} is a combined parameter that is a function of the dissociation constant of integrase dimerization and the dissociation constant of integrase dimer binding to each DNA recognition site. Note that we normalise $[I]$ by D to reflect the intuitive fact that increasing D for a fixed integrase concentration $[I]$ will reduce the probability that four integrase molecules are bound onto any given register at any given time.

The rate at which d jumps to $d-1$ is modelled as

$$r_{OFF}d\frac{([I]/D)^4}{K_{ON}^4+([I]/D)^4}\frac{([X]/D)^4}{K_{OFF}^4+([X]/D)^4}.$$

Similarly to the case above, this rate is constructed as a scalar parameter r_{OFF}, representing a switching rate, multiplying the expected number of ON registers that have two integrase dimer/excisionase dimer complexes bound, thereby being in the configuration able to stochastically switch to the OFF state. In this expectation, both the integrase concentration-dependent fraction and the excisionase concentration-dependent fraction are Hill functions with parameter 4, since four integrase and four excisionase molecules are required for an ON register to flip to OFF. Similarly to the previous case, the parameter K_{OFF} represents the dissoci-

ation constant of excisionase with integrase. Again, both protein concentrations are normalised by D to reflect the fact that the concentration is considered relative to the number of DNA registers to be switched.

A diagram summarising the structure of our model is depicted in Figure 5.2. The number d of DNA registers in the ON state follows the dynamics of a random walk, with time-varying transition rates that depend on the concentrations $[I]$ and $[X]$ of integrase and excisionase. In the absence of feedback, $[X]$ is assumed to equal zero for all time, so that the switching process only increases the number of ON registers.

5.3.5 Simulation Setup and Parameters

For input levels 0, 5, 15, 25, 50, 75, and 100 ng/mL, and for feedback both on and off ($\mu_X > 0$ and $\mu_X = 0$ respectively), we simulated $10,000$ single cells using the mathematical model described above with the parameters given in Table S6. This was done with values of $\alpha\beta$ spaced evenly with $\alpha\beta \in [0.25, 2.25]$, which gave us a distribution of of cell behaviours from which we could artificially simulate triplicates. At every hour during the simulation the following process was repeated 2000 times to form and analyse an equivalent to the biological triplicates: We sampled values $\alpha_i \in \text{Unif}(0.5, 1.5)$ for $i = 1, 2, 3$, and then (uniformly) randomly selected 2500 cells with $\alpha\beta \in [\alpha_i \times 0.5, \alpha_i \times 1.5]$ from the $10,000$ single cells simulations to form each of the three triplicate populations. This process is equivalent to setting $\beta_{i,j} \in \text{Unif}(0.5, 1.5)$ for $j = 1, \ldots, 2500$, but by selecting cells with an appropriate range of $\alpha\beta$ we are able to sample from a constant pool of cells, rather than simulating a large number of distinct triplicates. For each simulated triplicate the mean fluorescence of each of the three populations was calculated, as was the standard deviation of these means. After the 2000 repetitions of this process the mean of the individual triplicate means was calculated,

as was the mean of the triplicate standard deviations, which is the data presented in (for example) Figure 5.4a. To give the steady-state data in Figure 5.3a the simulation described above was run until both the open- and closed-loop systems reached equilibrium: to do this we (conservatively) set the simulation length for concentrations $aTc \leq 15$ ng/ml to 100 hours, and for $aTc > 15$ ng/ml to 30 hours. Parameter values are given in Table 5.1

Parameter	Value / Distribution	Unit
D	4	Scalar (Copy Number)
N	10	Scalar (Copy Number)
μ_I	0.2	nM/s
μ_X	10	nM/s
μ_G	1	nM/s
δ_I	0.01	1/s
δ_X	0.1	1/s
δ_G	0.001	1/s
K_I	15	nM
K_{ON}	25	nM
K_{OFF}	50	nM
r_{ON}	0.005	1/s
r_{OFF}	0.01	1/s
α	Unif$(0.1, 1.9)$	Scalar

Table 5.1: **Simulation Parameter Values**

5.4 Results

5.4.1 Assembly of the integrase/excisionase system

I implemented two synthetic systems based on the integrase DNA switching mechanism, depicted in Figure 5.1. Both systems placed a DNA coding region between the attachment sites attB and attP, counter-aligned from the constitutive promoter such that the enclosed proteins would not be initially expressed; I call this state the OFF state. The presence of integrase, expressed from its own plasmid in response to anhydrotetracycline (aTc) inducer, inverts the DNA coding

region to align with the constitutive promoter, thereby enabling expression; I call this state the ON state. In the *open-loop* system, i.e. the system *without feedback*, the invertible DNA region codes only for GFP reporter, whilst in the *closed-loop* system it codes for both GFP and the cognate excisionase. This means that in the closed-loop system the presence of excisionase with the available integrase enables the reversion of the coding region back to its initial OFF state. Thus, negative feedback is achieved by using excisionase as a proxy for fluorescence; when GFP concentration is large, the correlated concentration of excisionase acts to bias the random walk towards lower numbers of ON registers, thereby reducing GFP expression. To prevent the closed-loop system from heavily biasing towards the OFF state I also incorporated a strong ssrA degradation tag (already reported [12]) on the excisionase to ensure that it was degraded quickly.

5.4.2 System modelling and simulation

To investigate the performance of our integrase/excisionase-based implementation of negative feedback we developed a mathematical model that combines both stochastic processes and differential equations, for which a block diagram is presented in Figure 5.2 (for further details see the Methods section). Representative model parameters were chosen in line with the assumptions made during the design of our system, and were subsequently tuned so that the model's behaviour would qualitatively capture that exhibited in the experimental data. We simulated cell populations by introducing two random parameters to capture cell-to-cell heterogeneity that would result in variability in the production rate of GFP and (in the closed-loop case) excisionase. These random parameters are α, which was chosen to simulate inherited variations in behaviour found between populations grown from single cells, and β, which introduces further cell-to-cell variation within each population. Thus, α accounts for some inherited intrinsic

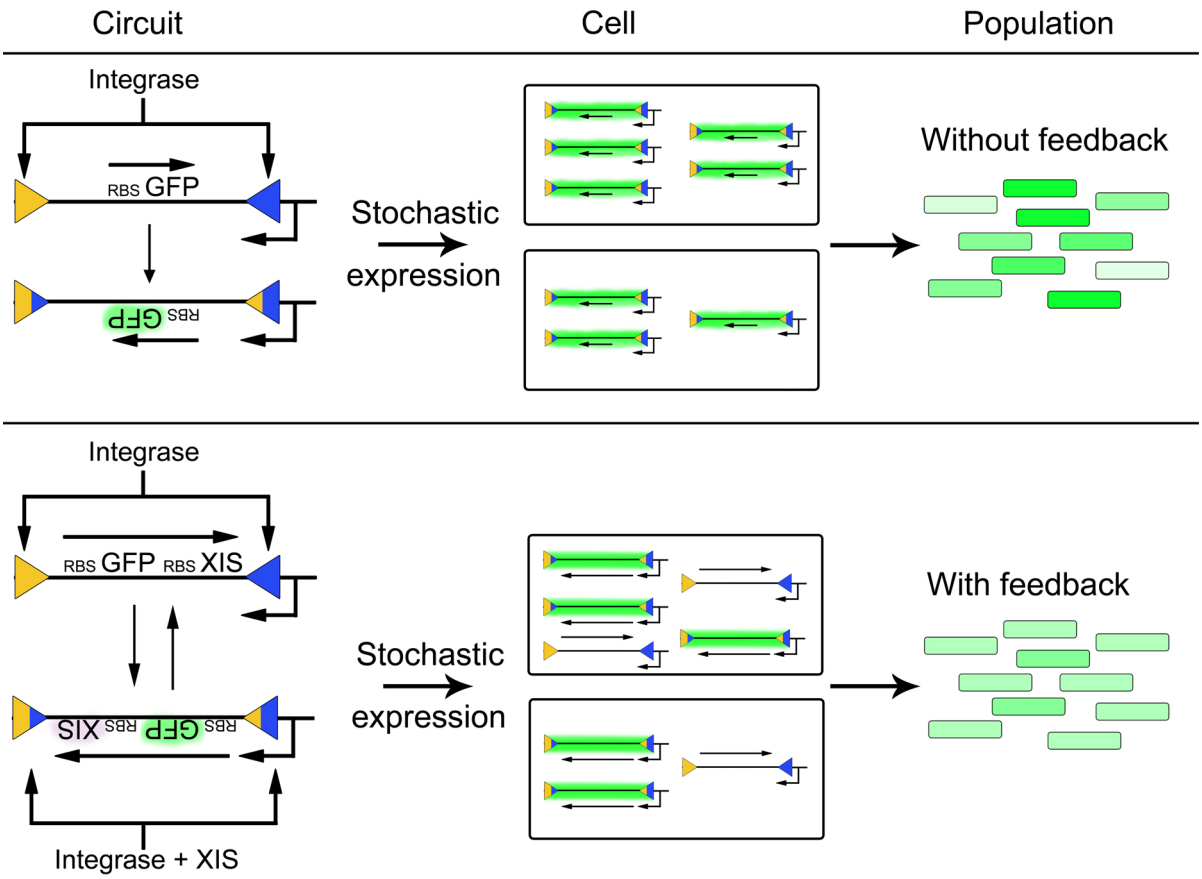


Figure 5.1: The schematic architecture of the open-loop (top) and closed-loop (bottom) expression systems developed in this chapter. The correlated expression of excisionase and GFP closes the loop around the integrase's switching process, allowing stochastic switching between ON and OFF states, thereby implementing negative feedback. Thus, whilst all DNA registers in the open-loop system eventually switch to the ON position, those in the closed-loop system are regulated by excisionase concentration leading to smaller variability in the number of ON registers per cell: If GFP concentration is high (low), excisionase concentration will be likewise, resulting in more OFF (ON) registers and down (up) regulation of GFP expression. The yellow and blue triangles represent attP and attB recognition sites, whilst half-coloured triangles represent attL and attR.

noise sources (such as long-term inherited differences in proteome) as well as extrinsic noise on a large scale (such as environmental differences between biological triplicates), whilst β accounts for cell-to-cell intrinsic noise sources (such as variations in plasmid copy number) and small scale extrinsic noise (such as local environmental inhomogeneities). Triplicates were generates by splitting cells into three populations, each with a different value for α, and within each population

values for β were sampled for individual cells. With this minimal model we were able to predict (from a theoretical standpoint) many of the benefits our negative feedback system would provide, which were then confirmed by experimentation as discussed in the following sections.

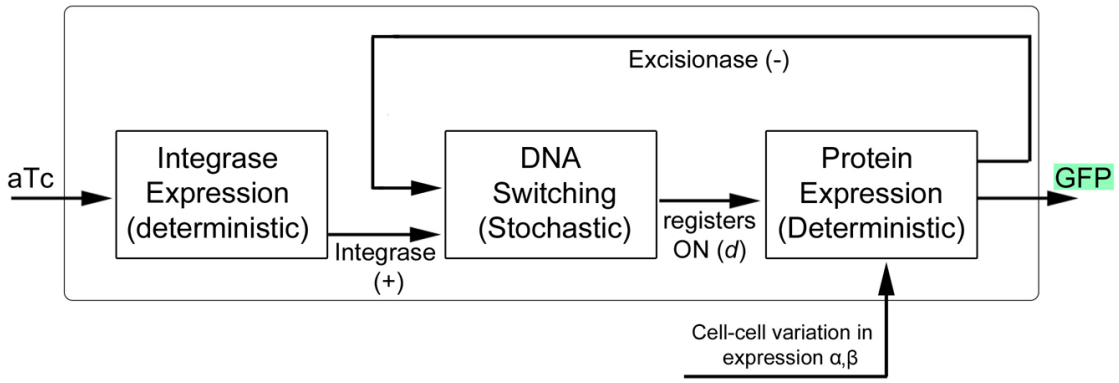


Figure 5.2: Block diagram for the modelled feedback system. The inducer (aTc) acts as an input to the deterministic expression of integrase. Integrase (and excisionase) concentration determines the transition probabilities of the random walk of the integer-valued state d representing the number of ON registers. Excisionase acts as negative feedback, because it biases the random walk towards transitions $d \rightarrow d-1$. The state d then acts as a random input to the deterministic protein expression dynamics, together with randomly selected constant scaling parameters α and β to represent variability between populations and cells.

5.4.3 Introduction of excisionase reduces the system's expression level and response time

Using our mathematical model we simulated the steady-state response of the system to a range of anhydrotetracycline (aTc) concentrations. Figure 5.3a shows the steady-state expression per cell, which is uniformly decreased in the closed-loop when compared to the open-loop system. This result was then confirmed experimentally, with the data (presented in Figure 5.3b) exhibiting a near-uniform trend towards lower expression in the closed-loop system. To investigate the impact of feedback on the system's response time we simulated time-course data for GFP expression-per-cell with 50 ng/ml aTc concentration (Figure 5.4a), and then

performed the corresponding experiment for comparison (Figure 5.4b, for other aTc concentrations see Figure 5.3). As seen in the experimental results, the time-course data from the closed-loop and open-loop systems match relatively closely for the early part of the response, but the steady state of the closed-loop system is realised earlier (at approximately 7 h post-induction) than in the open-loop system (approximately 10 h post-induction).

5.4.4 Synthetic feedback reduced output variation

Figure 5.3 also demonstrates that over a range of aTc concentrations, in both simulation and experiment, the variation in end-point expression of the closed-loop system is substantially lower than in the open-loop. For the end-point experimental data in Figure 5.3b, if I normalise the standard deviation of the triplicates' mean fluorescence (which is plot in the error bars) by the mean fluorescence of each triplicate set (for aTc$\geq$ 15 ng/ml), I find that on average the inclusion of negative feedback reduces the standard deviation of fluorescence between triplicates by a factor $\sim$2.8. Furthermore, Figure 5.4 demonstrates that in both simulated and experimental data the closed-loop system provides a reduction in expression uncertainty: the size of the error bars, again when normalised by the mean expression level at a given point, is substantially reduced for the closed-loop system, even early in the response when the mean expression levels do not significantly differ.

5.4.5 Intrinsic noise is reduced by the synthetic feedback loop

Via flow cytometry I sought to investigate the effect of the negative feedback system on the distribution of GFP expression levels between cells, beyond the population-level comparisons revealing closer alignment of population means as

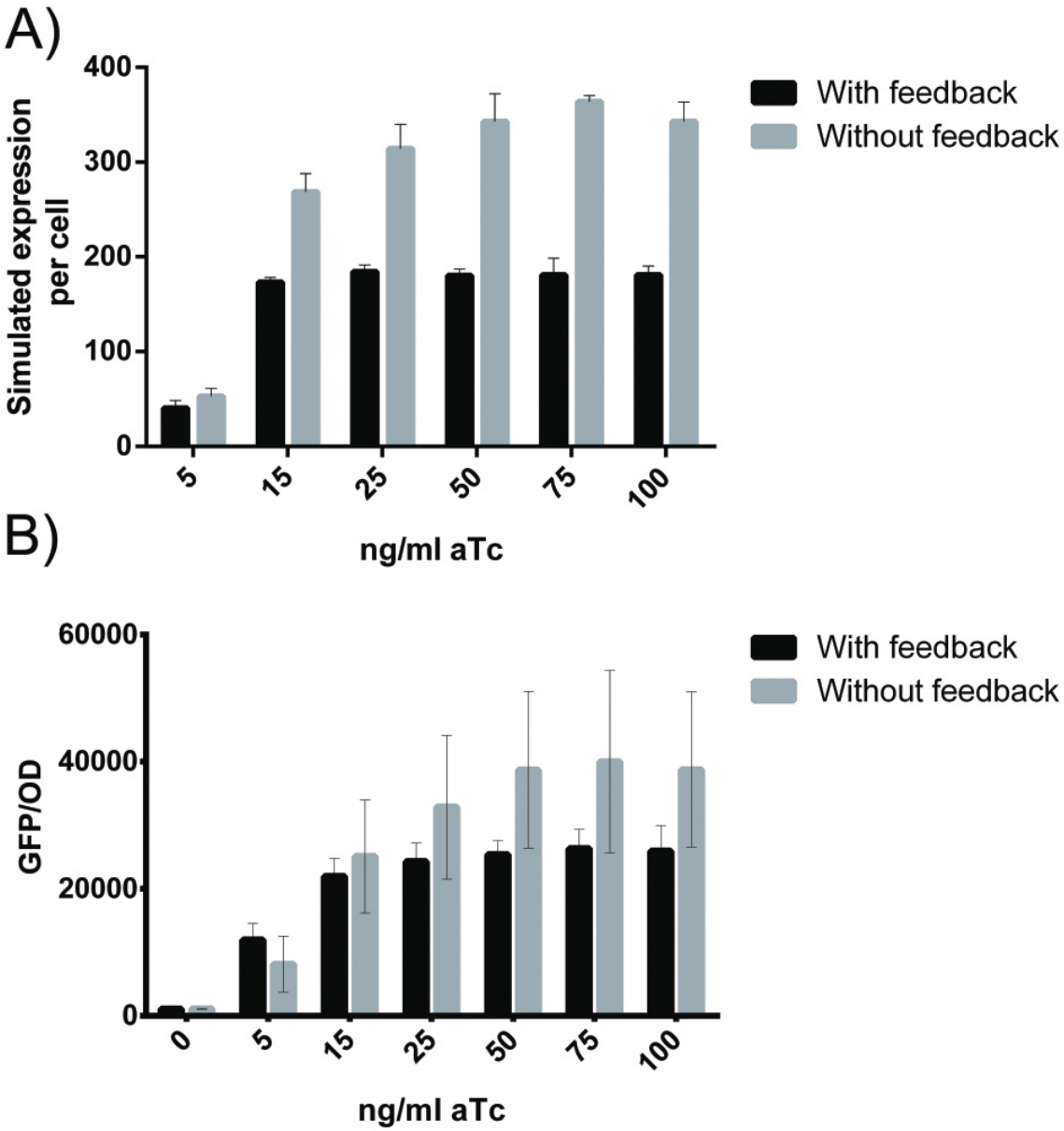


Figure 5.3: Steady state responses after for the open- and closed-loop systems over a range of aTc inputs for: (A) The simulated system run until equilibrium is reached, where the mean end-point GFP concentration per cell (in arbitrary units) for 2000 sets of triplicates (each consisting of three populations of 2500 cells) is plot, with error bars indicating the mean standard deviation in fluorescence within each triplicate set. (B) Averaged end point (taken to be 12 hours) GFP expression from triplicate biological repeats, where error bars indicate the standard deviation in fluorescence of the triplicates. The y-axis is GFP expression (measured via culture fluorescence) divided by Optical Density (OD), in arbitrary units. For both experimental and simulated data the feedback system demonstrates a reduction in steady state fluorescence, as well as a reduced variation (when normalised by mean expression level) in expression between triplicates.

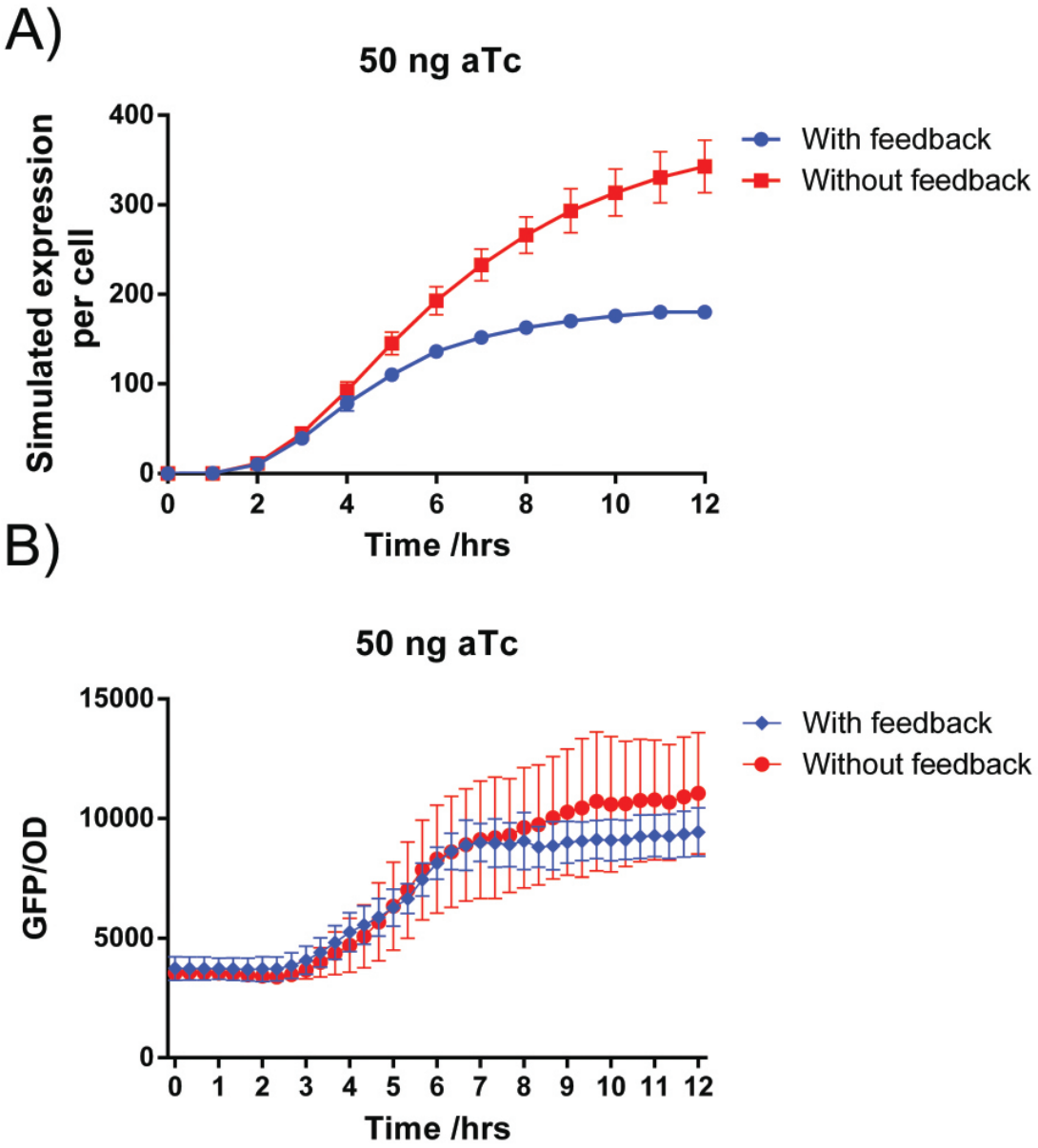


Figure 5.4: **Steady state is achieved more quickly and robustly in the presence of feedback** dynamic behaviour at 50 ng/mL aTc. (A) Simulation as in Fig. 5.3. The mean and standard deviation of the subpopulation means are taken at time points from 0 to 12 h. (B) GFP expression from unregulated circuit compared with the regulated circuit plotted against time. Error bars denote standard deviation around the mean for triplicate biological repeats. Equivalent plots for the full range of inducer concentrations are shown in Figure 5.5.

depicted in Figures 5.3 and 5.4. In particular, I aimed to investigate the effect of the negative feedback system on the distribution of GFP expression within a cell population. I used single cell flow cytometry to obtain the histograms in Figure 5.6, where each plot depicts the GFP expression distribution of three pop-

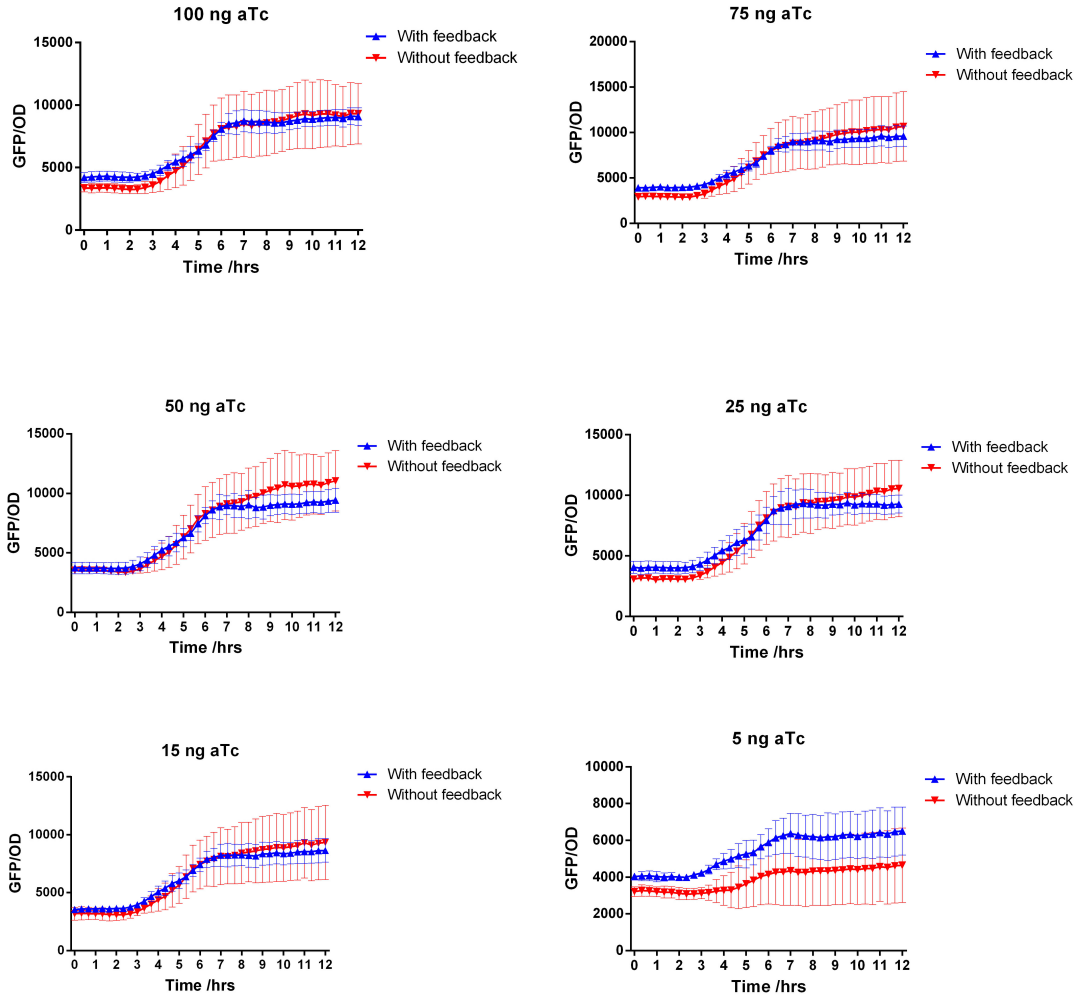


Figure 5.5: Time-course fluorescence data for the open- and closed-loop systems induced with 50 ng/mL aTc over 12 hours for: (A) The simulated system as in Figure 5.3, where the overall average of the mean and standard deviation of 2000 triplicates is plot each hour. (B) Averaged time-course GFP expression from triplicate biological repeats, where error bars denote the standard deviation of the individual triplicate means. The y-axis is GFP expression (measured via culture fluorescence) divided by OD, in arbitrary units. Equivalent plots for the full range of inducer concentrations are shown in Figure 5.3. For both experimental and simulated data the feedback system demonstrates a faster response, as well as reduced variation in fluorescence levels.

ulations (biological triplicates) of an open-loop or closed-loop system at a given inducer concentration. In the presence of feedback I find a closer alignment of the triplicates' means, as seen in earlier experiments, and also observe that variation between populations in terms of the overall shape of the expression profile

is substantially decreased in the closed-loop system. Furthermore, compared to the open-loop histograms, those of the closed-loop system demonstrated a substantially narrower distribution of fluorescence within each population. This is made clear in Figure 5.6, in which the data from the closed- and open-loop cases for one of the triplicates has been overlaid.

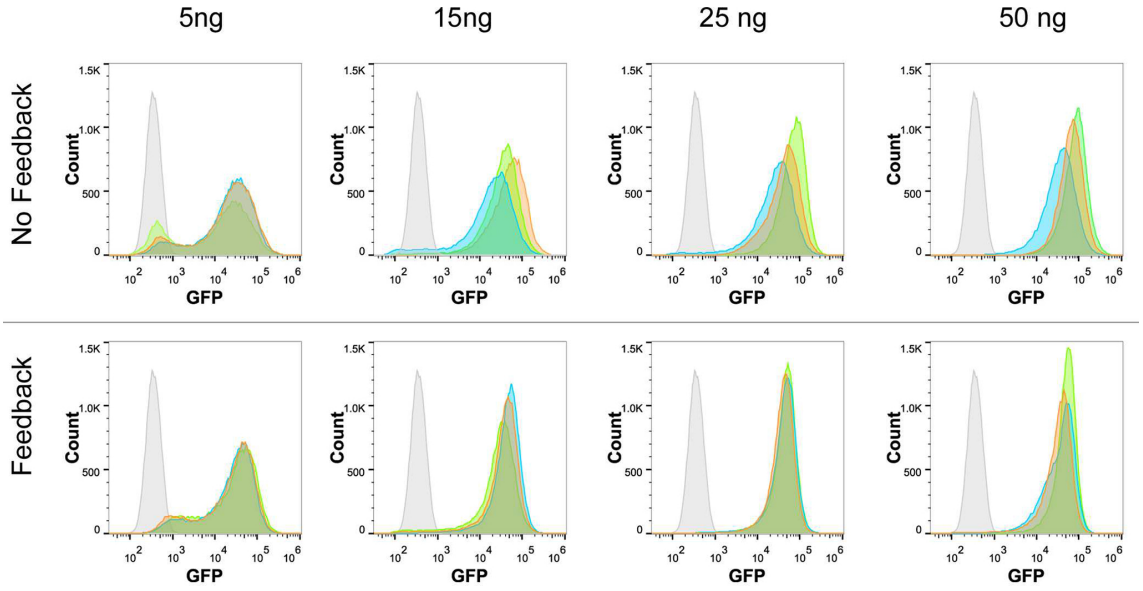


Figure 5.6: Single-cell flow cytometry data for triplicates of both the open- and closed-loop systems at a range of inducer concentrations. The x-axis is the level of GFP expression (measured via fluorescence) per cell, in arbitrary units. The triplicate biological repeats are overlaid and coloured green, blue and orange, and the grey distribution is a negative *E. coli* control. I observe that feedback results in higher reproducibility between replicates, both in terms of the location of the histogram's peak, as well as its overall shape. The system with feedback also results in a narrower histogram , indicating a reduction in cell-to-cell variability. The reduction in population vari-ation is quantised by calculating coecients of variation (CV) for each distribution, which are on average 25 percent lower in the closed-loop system when compared to the open-loop

To ensure the feedback loop was acting as modelled I implemented two controls, one with a non-functional excisionase and one with an excisionase with no SSRA tag. The excisionase mutant showed a slight increase in GFP expression indicating that this DNA conformation might be easier to invert for the integrase but noise was not reduced compared with the open loop system as seen in Figure

5.7A. Removal of the SSRA tag resulted in a substantial reduction in expression and a widely distributed population (as seen in Figure 5.7B) indicating that the increased degradation rate of the excisionase was, as predicted, a central aspect of the feedback system.

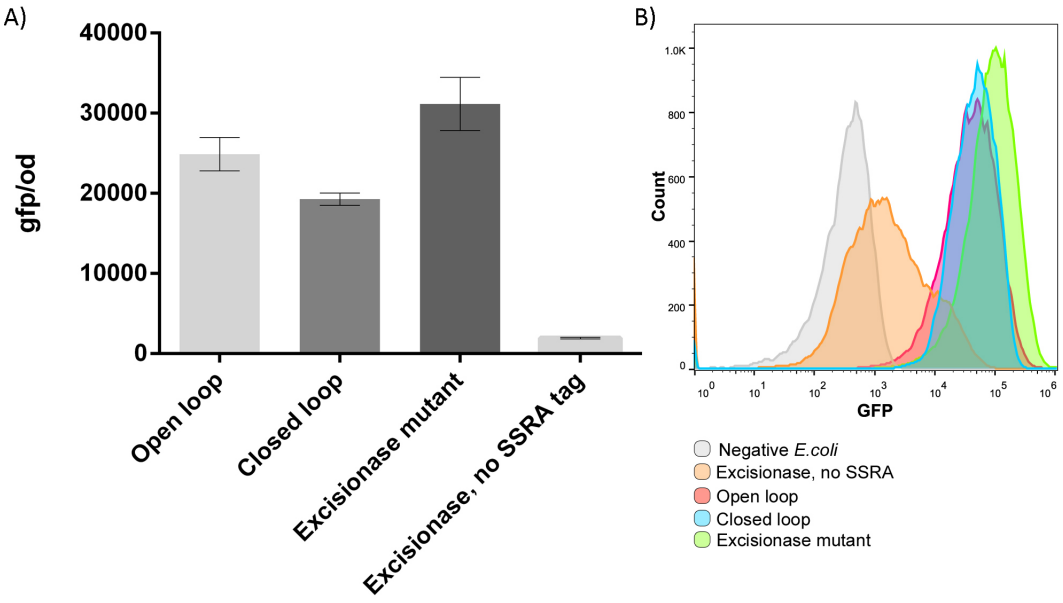


Figure 5.7: Effect of an excisionase mutant or removal of the SSRA tag on the feedback system. A) comparison between the open loop, closed loop systems and an excisionase non-functional mutant and an excisionase with no SSRA degradation tag. B) Single cell representative populations showing a broad population resulting from the removal of the SSRA tag.

5.4.6 Downstream process is easily tuned to a predictable threshold

Using the Salis Ribosome Binding Site (RBS) calculator [139] I generated a small library of GFP expression variants (that is, GFP genes with different RBSs) over an approximate 10-fold range in strength, and integrated each one into the closed-loop system. For each of these four variants I observed a reliable and distinct level of expression in the downstream process reporter as seen in Figure 5.8 , which correlated with the relative theoretical RBS strengths. This demonstrates a

straight-forward approach via which the implementation of our feedback system can be tuned to reliably set a controlled gene's output to a desired level.

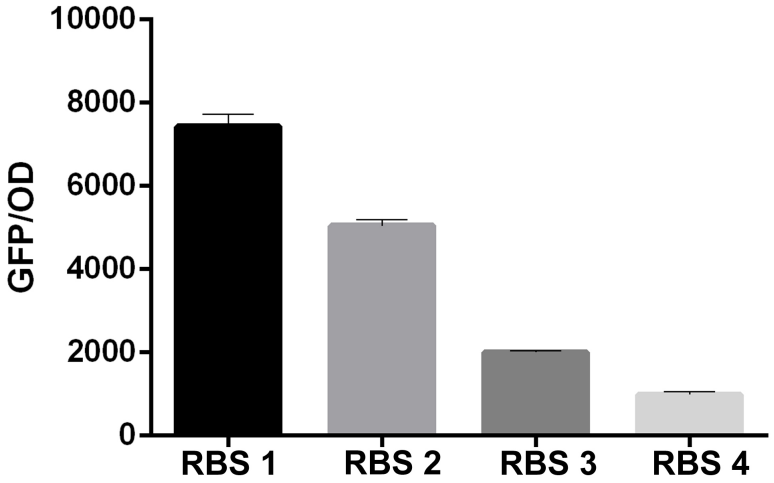


Figure 5.8: Output response from the closed loop system is predictably tunable. The down stream process controlled by the integrase and excisionase loop is orthogonally tunable through well known RBS screening methods. Here I integrate a library of RBS stregths generated by the Salis calculator over a 10-fold range. Error bars represent standard deviation aroun the mean for biological triplicates

5.5 Discussion

Our results demonstrate a good correspondence between the qualitative behaviour of the simulated and experimental systems, in both cases highlighting the beneficial properties of negative feedback (closed-loop) systems. Figure 5.3 demonstrates a reduced steady-state response in the closed-loop system, which can be justified as follows: in the open-loop system the DNA registers can only switch ON, without possibility of reversion, due to the absence of excisionase. Thus, the GFP expression of the open-loop system is maximal given the induction levels and constitutive expression rates of GFP. In contrast, for the closed-loop system the excisionase expression provides negative feedback resulting in reversion of DNA registers, and thus on average a reduced fraction will be in the ON posi-

tion, leading to a lower expression level of the GFP reporter. Figure 5.4 (and 5.5 for other inducer concentrations) shows that the presence of excisionase feedback also allows the closed-loop system to realise its steady state earlier than open-loop systems, which is also intuitive: in the open-loop system, a steady state is only achieved once all DNA registers in the entire population have flipped ON. However, the closed-loop system reaches steady state when excisionase concentration is sufficient to balance the system's stochastic DNA flipping between ON and OFF, which requires less time as only a fraction of the registers need be flipped to achieve equilibrium. Though the overall responses of the closed- and open-loop systems differ as discussed, they do match relatively closely for the first few hours of the experiment due to the delay in the system's initiation which arises from the time needed for aTc induction to prompt transcription and translation of the integrase. For the experimental steady-state data (Figure 5.3b) the reduction in fluorescence for lower aTc concentrations may be partially attributable to insufficient time being given for the cell culture to fully reach steady state. This again supports the claim that negative feedback reduces the response time of our system, since the closed loop system is able to reach steady state for aTc concentrations of 15 ng/ml and above, whilst the open-loop system's output only levels off within the same time period (12 hours) for aTc concentrations of 50 ng/ml and above.

In all experiments and simulations it was found that variability was reduced in the closed-loop system. For example, at steady state in Figure 5.3b the normalised standard deviation of fluorescence between triplicates was reduced by a factor of $\sim$2.8 in the closed-loop system when compared to the open-loop variant. This might be considered counter-intuitive, since one could initially hypothesise that expressing excisionase introduces extra stochasticity to the expression of the reporter protein, thereby increasing the variability of its output. However, to un-

derstand why this intuitive reasoning fails, first note that the regulated output is not the number of DNA registers in the ON state, but rather the concentration of GFP. In populations of cells with faster (slower) constitutive GFP expression rates than the mean, the correlation between excisionase and GFP expression (since they are in the same operon) means that excisionase will also be produced at a correspondingly faster (slower) rate, thereby speeding (slowing) the rate at which the populations' DNA registers revert back to OFF. Thus, populations with GFP expression that tends away from the mean in open-loop due to natural cell-to-cell variation will be regulated closer to the mean in closed-loop, providing negative feedback functionality. Considering this in the context of our model, if the random scaling factor $\alpha\beta$ is large (small), which might correspond to the GFP/excisionase plasmid having higher (lower) copy number in that cell, then for each ON register there will be a large (small) expression of both GFP and excisionase, which will reduce (increase) the number of ON registers. So, for cells with large (small) $\alpha\beta$ the level of expression of the target protein per ON register will be high (low), but there will be fewer (more) ON registers.

The variability between populations demonstrated by flow cytometry measurements in Figure 5.6 supports the approach taken to introduce random variations between cells in our model: I observe (particularly in the open-loop case) substantial variation between triplicates in terms of the centre of their GFP fluorescence distribution, which corresponds to varying values for the random parameter α, as well as a wide distribution of fluorescence values within each population, corresponding to variation in the parameter β. Both of these parameters reflect natural differences in behaviour that arise between cells: α accounts for differences (for example, in the proteome) that are inherited from the small number of cells used to initiate each triplicate, whilst β reflects the spread in cell properties within each triplicate during growth and experiment (for example, due to inhomogeneities in

environmental conditions or plasmid copy number fluctuations). Further work could be done to refine the probability distributions from which α and β are sampled; in our model both were sampled from uniform distributions to ease computation and analysis.

Though the negative feedback system developed in this work has achieved a reduction in both the system's response time and the variability of its output, further development and improvement is possible. Given that the feedback signal is excisionase concentration, the 'actuator' of GFP expression is the flipping of DNA registers to ON or OFF positions. Thus, the actuator signal controlling GFP expression is inherently quantized, since only an integer number d of the total D registers can be ON, leading to expression rates at fractions d/D of the maximum. In engineered control systems, quantized signals are known to degrade performance below the optimum [2]. I thus theorize that an increased number of DNA registers would reduce the effect of quantisation, thereby improving the noise-attenuating properties of our system further. However, without additional redesign of the excisionase expression/degradation rates a larger plasmid copy number is likely to also lead to an overly strong feedback signal, as well as increased cell burden, potentially causing a subsequent loss of output signal.

The synthetic feedback system developed herein is of particular interest as it centred on serine integrases: these components have been extensively characterised in Synthetic Biology, and libraries of integrase proteins with orthogonal binding domains exist. Thus our design is likely to be easily scalable, allowing a number of integrase-excisionase feedback circuits to function within the same cell. Given that phage-derived integrases have been used in systems with a wide variety of output proteins over the last three decades [21, 26, 92], there should

be no significant limitations in replacing the GFP reporter in our design with an alternate target protein, making this framework widely adaptable to a broad range of applications in a plug and play fashion. Furthermore, the phage proteins are non-native and have been shown to function in many different domains of life [138, 173]. Therefore, there should be minimal crosstalk between cellular processes and the feedback circuit (as is often found in transcription-based feedback designs), which makes this framework widely transferable between different hosts and applications. It would therefore be especially useful in systems that require highly accurate and reproducible protein expression, such as for the regulation of degradase expression for degradation of orthogonally tagged proteins, or to uniformly express a complemented gene in knock-in studies.

Chapter 6

Conclusions

Biosensor development is an area of significant scientific interest but progress is limited by issues such as low signal to noise and low absolute levels of expression as I showed in Chapters 3 and 4. This has lead to considerable affects to optimise and modulate the outputs from biosensors.

Previously, transcription factor based biosensor output has been modulated by either RBS screening to find a better expression rate as in *Ravikumar et al* [124] which focused on the development of an orthogonal zinc and copper dual biosensor, or through the employment of a T7 amplifier as in *Kim et al* [80] which focused on making a highly sensitive lead and cadmium biosensor. Both these approaches have disadvantages, RBS opimisation can be very time and resource consuming and may need to be repeated every time an alternative expression rate is needed. The use of a T7 polymerase to amplify a biosensor signal is limited by the high activity of T7 polymerase. Biosensors amplified in this fashion must have very low basal expression as even low expression of T7 could saturate the output signal. Often these issues prevent the use of a biosensor in an applied setting. Low signal to noise ratio can cause issues when trying to differentiate between two different concentrations of ligand.

Previously, riboswitches have been modulated to express a new gene of interest by fusing the new gene behind the 5' region of the native gene, allowing for the expression of a fusion protein under the control of the riboswitch [52, 111, 168, 40]. This approach is widely used as it prevents new secondary structure from displacing the structure of the riboswitch and altering its response to the ligand. Some advances have been made towards the *in silico* design of riboswitches by *Borujen et al* [45] however the types of riboswitches and the ligands are very limited.

To reduce inter and intra cellular variation, feedback modules have been designed around transcription factor autoregulation. These function by employing a repressive transcription factor to implement feedback from the dowstream protein expression to the promoter as in *Rosenfeld et al* [136]. This approach does however require a transcription factor to be characterised for each use and as such is not easy to treat in a modular fashion.

In Chapters 3 and 4, I demonstrated that RNA based components changed their behaviour when their surrounding sequences were altered i.e. suffered from contextual limitations. This issue of contextual limitation bears significant problems for one of the core principles of synthetic biology - the use of standardised parts. These standardised parts are usually comprised of a single biological component which is projected to work uniformly across different contexts. For example, *BBaJ*61100 is an RBS predicted to have a strong translation initiation rate however *Mutalik et al* [105] showed that this part is likely to be adversely affected by its context. I discovered this contextual sensitivity throughout the RNA parts I used.

For example, the copper biosensor in Figure 3.13C was unable to statistically differentiate between 1 and 2 mM copper sulphate due to its low signal to noise ratio and the ribo-regulators used in Chapter 3 were unable to mediate this problem

with the copper biosensor due to contextual sensitivity as seen in Figure 3.13B.

I continued to characterise previously unknown contextual issues surrounding the use of riboregulators to amplify conventional zinc, copper and arsenite transcriptional biosensors using TP901-1 integrase. While this approach worked with the zinc biosensor as seen in Figure 3.12 there were contextual limitations that prevented this approach being expanded to the arsenite and copper biosensors as I showed in Figure 3.13. I hypothesise that the change in context when the copper or arsenite promoters were introduced to the ribo-regulators caused a change in the structure of the ribo-regulators causing them to cease functioning.

I went on to characterise a new genetic element called a ribo-attenuator in Chapter 4 which reduced noise and the chance of context sensitivity when changing a down stream gene controlled by a riboswitch. I suggest that this context sensitivity comes from the change in secondary structure around the riboswitch reducing affinity for the ligand when a new gene is introduced. The ribo-attenuators reduced this by isolating the riboswitch and the leader region of the gene it functions with and the new gene to be introduced.

I looked at the addA and btuB riboswitches which gave high degrees of noise and showed context sensitivity when a new gene of interest was introduced as seen in Figure 4.7A. Ribo-attenuators allowed the addA riboswitch to more accurately control gene expression and the btuB riboswitch to greatly increase protein yield.

Finally I characterised a new feedback architecture centred around the relationship between BXB1 integrase and its associated excisionase enabling less within and between population variation enabling more reliable and repeatable function from an integrase. This could rapidly be employed to improve existing integrase biological devices such as the medical diabetes detector discussed in *Corbet et al* [32]. An overview of all three results Chapters is shown in Figure 6.1

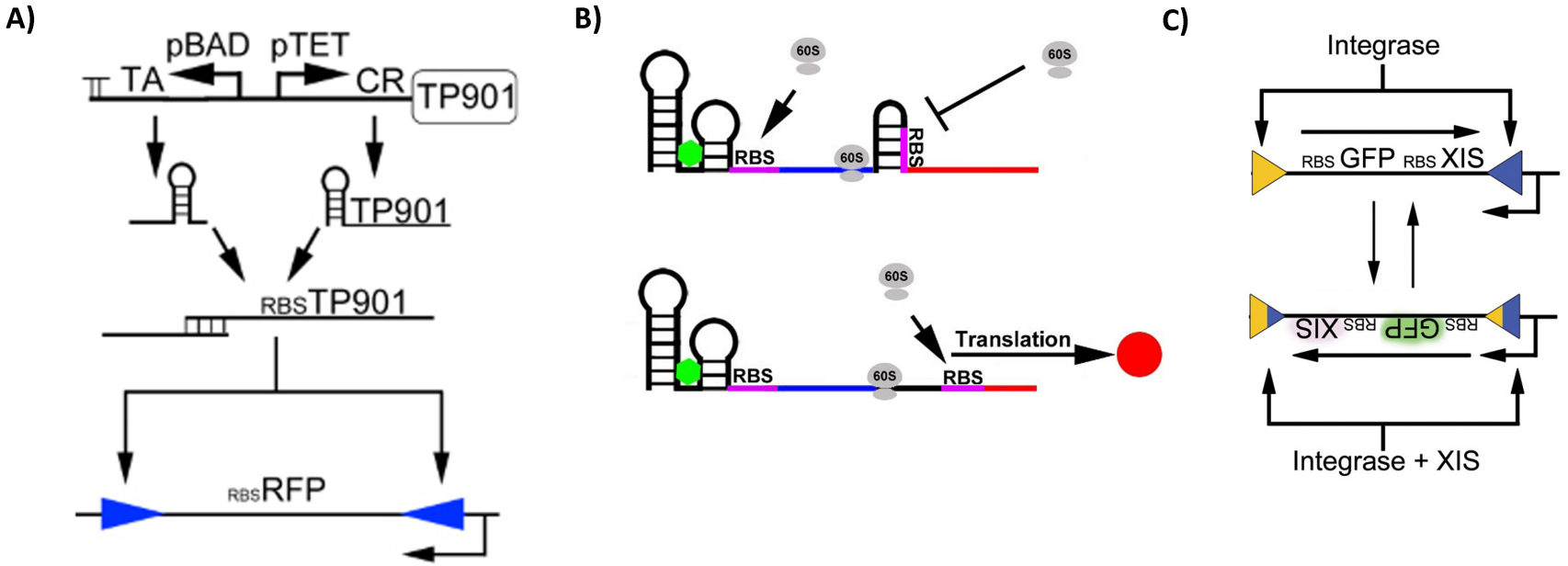


Figure 6.1: **Thesis chapters overview** A) Overview schematic of trans-activating/cis-repressing system showing dual expression of both components as required for DNA inversion as discussed in Chapter 3. B) A ribo-attenuator adds a second RBS, sequestered away by a local hairpin allowing for the introduction of a down stream gene of interest to be introduced without interruption of the upstream riboswitch. The hairpin can be opened by a Ribosome travelling from the riboswitch RBS, exposing the attenuator RBS and allowing translation of the gene of interest as discussed in Chapter 4 C) The schematic architecture of the closed-loop expression system developed in chapter 5. The correlated expression of excisionase and GFP closes the loop around the integrases switching process, allowing stochastic switching between ON and OFF states, thereby implementing negative feedback.

In Figure 6.2 the combined approach to re-engineering a biosensor is shown. If a biosensor with a digital response is required the ribo-regulators can be applied as discussed in Chapter 3. If context dependencies are seen the ribo-attenuators discussed in Chapter 4 have been shown to alleviate this problem. If the digital biosensor has a low signal to noise ratio, the excisionase feedback could reduce this noise. In table 6.1 I tabulate where the advances made fit within the field of biosensor engineering.

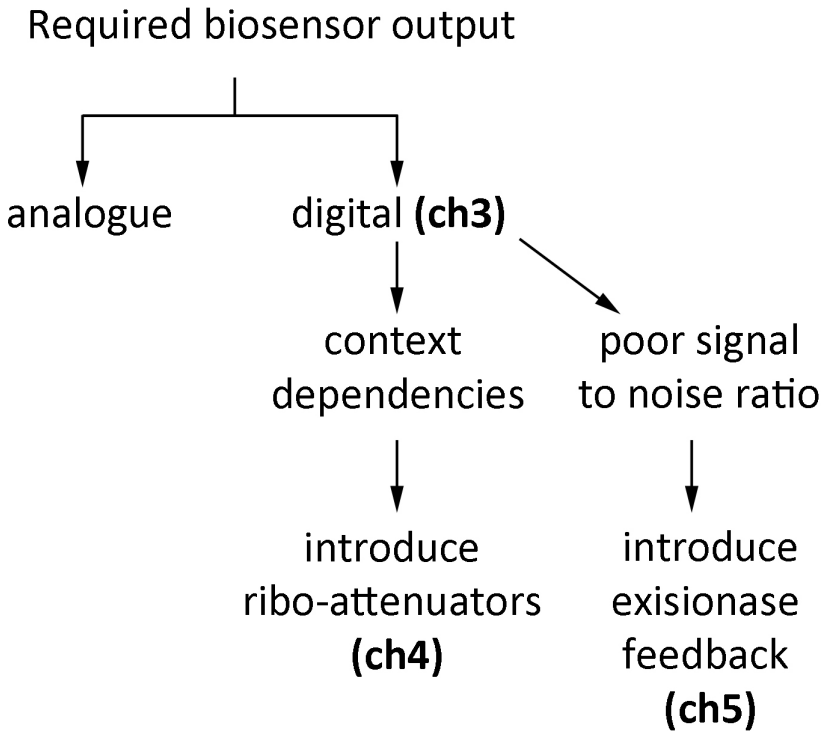


Figure 6.2: **Thesis conclusion overview**

Table 6.1: Biosensor engineering approaches - advantages and disadvantages

Biosensor type	Example	Advantages	Disadvantages
Analogue biosensor	Zinc biosensor [124]	Gradual response to higher concentration of ligand	Low expression levels and low signal to noise
Digital biosensors	pBAD [78]	Distinct ON or OFF response	no ability to modulate expression
Traditional riboswich	btuB riboswitch [52]	Low basal expression	Low modularity
Riboswitch ribo-attenuator	Chapter 4	High signal to noise, better modularity	higher basal expression
Transcription factor feedback	TETr autorepressor [136]	Fast response time	Low absolute expression and modularity
Integrase feedback	Chapter 5	Higher signal to noise, high modularity	Slow response time

6.1 Further work

Several challenges should be addressed for future work in order to extend the utilities of the characterised devices.

6.1.1 Chapter 3

As discussed in Chapter 3 contextual sensitivity prevented the use of the post-transcriptional riboswitches, there was no obvious explanation of this sensitivity. Moving forward with this chapter would likely require characterising an alternative approach to using these post-transcriptional riboswitches. One alternative approach is to avoid the riboswitch and employ graded degradation of the integrase using tunable degradation tags [20]. These tags would allow for the isolation of the two parts of the circuit, the integrase and the inducible promoter.

6.1.2 Chapter 4

As discussed in Chapter 4 an issue with using various strength hairpins is that the shorter hairpins are less stable and prone to misfolding. Future work on this chapter should implement internal competing secondary structures for shorter hairpins that do not bind the down stream RBS. With this approach the stabilising hairpin could allow for a more accurate prediction of RBS hairpins as shown in Figure 6.3. The strength of both hairpins would require fine tuning.

Further to tuning internal hairpins, showing the ribo-attenuators working with several other riboswitches and several other open reading frames to fully characterise the component.

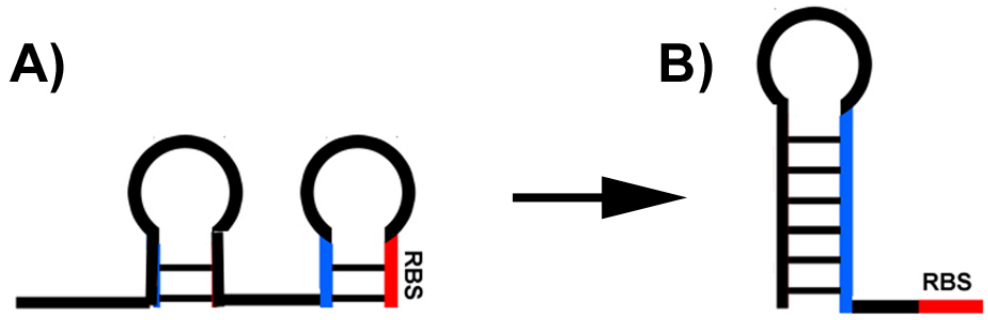


Figure 6.3: **Internal competing hairpins could stabilise smaller hairpins** A) Smaller hairpins with strong internal complimentary could give a more stable secondary structure. B) when the hairpin reforms it has a chance to for a strong stabilising hairpin opening up the RBS.

6.1.3 Chapter 5

As discussed in chapter 5, expanding the novel feedback architecture I described would require the addition of a second and third integrase to enable multiple feedback devices to be used simultaneously. Further to this altering the promoter that drives the expression of the reporter (in our system this is a constitutive strong promoter) to an inducible promoter could allow for fine tuning of the expression level. Alternatively tuning of the SSRA degradation tag on the excisionase could achieve the same outcome although this tuning may be time and resource heavy.

6.2 Appendix

Table 6.2: List of primers

Primer name	Sequence
TF_VF2:SFGFP.FWD	acccgtgaaggtgagccagttcatcatttgtacagttcatccatacc
TF_bbone:GDP.REV	atgaactgtacaaatgatgaactggctcaccttcacg
TF_GFPint.fwd	cgtcaccttcaccctc
TF_GFPint.REV	ggtcataagttttccgtgcg
TF_RAD:bbone.REV	tgactttctgctaatctaacgcgagtttttatttcgtttatttcaattaagg
TF_bbone:RAD.FWD	taaacgaaataaaaactcgcgttagattagcagaaagtcaaaagcc
TF_001.FWD	ttcgctaaggatgatttctgaccttg
TF_002.REV	cactgaacaccataagtcagcg
TF_003.REV	ttaccagagtcggccaaggtcagaaatcatccttagcgaaagc
TF_004.FWD	atgaactgtacaaatgatgaggcttcctcgctcactg
TF_005.FWD	ggccgactctggtaacg
TF_006.REV	agtcagtgagcgaggaagcctcatcatttgtacagttcatccatacc
TF_0015.fwd	ctggatggtgatgtcaacgg
TF_016.rev	cttagacgtcgatatctggcg
TF_017.fwd	gattgaccttgccgacgc
TF_018.rev	cacgcacggaaaacttatgac
TF_019.rev	tgtatctgttaccagttcacggg
TF_020.fwd	ctaaggatgatttctgcccgat
TF_021.rev	gctgagcgtcaaggaattatcggg
TF_022.fwd	aatttctggcagatgtacccg
TF_023.rev	ttagtaaagctggtatgcgcg
TF_024.fwd	attctcaccaataaaaaacgcccgg
TF_025.rev	atttgtcctactcaggagagcgttca
TF_026.fwd	tccggcaaaaaaacgggc
TF_27.fwd	cctgtacttccaatccatggtccggcaaaaaaacgggc
TF_28.rev	atccgtatccacctttactgtccggcaaaaaaacgggc
TF_29.fwd	AAAAAGACGACAAGAATTAGcagtaaaggtggatacggatc
TF_30.rev	TTGTTTTGCTTTTTCCCCATccatggattggaagtacagg
TF_034.rev	atgtagccgtcaagttgtcataagaattgccagctggggc
TF_035.fwd	ggcgccccagctggcaattcttatgacaacttgacggctacatcattc
TF_036.rev	ccttgggcccaaagacgtcggacttttctgccgtgattatagacacttttg
TF_58	GAGATATACAaataatgcgtaaaggcgaagagctgttcac
TF_59	cttcgcctttacgcattattTGTATATCTCCTTC
TF_60	gtgatatgggatagtgttcacccttgttacacc
TF_61	ggtgaacactatcccatatcaccagctca
TF_62	GCGGCTTTCACACTTGGGGGCTAGCatgcgtaaaggcgaagagctgttcac

Bibliography

[1] Amor-Mahjoub, M., Suppini, J.-P., Gomez-Vrielyunck, N., and Ladjimi, M. (2006). The effect of the hexahistidine-tag in the oligomerization of HSC70 constructs. *Journal of Chromatography B*, 844(2):328–334.

[2] Aström, K. J. and Murray, R. M. (2010). *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, New Jersey, USA.

[3] Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J., Smith, J. A., and Struhl, K., editors (2001). *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

[4] Beard, S. J., Hashim, R., Membrillo-Hernandez, J., Hughes, M. N., and Poole, R. K. (1997). Zinc(II) tolerance in Escherichia Coli K-12: evidence that the zntA gene (o732) encodes a cation transport ATPase. *Molecular Microbiology*, 25:883–91.

[5] Beisel, C. L. and Smolke, C. D. (2009). Design Principles for Riboswitch Function. *PLoS Computational Biology*, 5(4):e1000363.

[6] Berg, J. M. and Shi, Y. (1996). The galvanization of biology: a growing appreciation for the roles of zinc. *Science (New York, N.Y.)*, 271:1081–5.

[7] Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology*, 62(3):293–300.

[8] Bibb, L. a., Hancox, M. I., and Hatfull, G. F. (2005). Integration and excision by the large serine recombinase Rv1 integrase. 55(6):1896–1910.

[9] Binet, M. R. and Poole, R. K. (2000). Cd(II), Pb(II) and Zn(II) ions regulate expression of the metal-transporting P-type ATPase ZntA in Escherichia coli. *FEBS letters*, 473:67–70.

[10] Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature*, 435(7045):1059–66.

[11] Bloom, R. J., Winkler, S. M., and Smolke, C. D. (2015). Synthetic feedback control using an RNAi-based gene-regulatory device. *Journal of Biological Engineering*, 9:5.

[12] Bonnet, J., Subsoontorn, P., and Endy, D. (2012). Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proceedings of the National Academy of Sciences*, 109(23):8884–8889.

[13] Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013). Amplifying Genetic Logic Gates. *Science*, 340(6132):599–603.

[14] Breaker, R. R. (2012). Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology*, 4.

[15] Breüner, A., Brøndsted, L., and Hammer, K. (2001). Resolvase-like recombination performed by the TP901-1 integrase. *Microbiology*, 147(8):2051–2063.

[16] Briat, C., Gupta, A., and Khammash, M. (2016). Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks. *Cell Systems*, 2:15–26.

[17] Brocklehurst, K. R., Hobman, J. L., Lawley, B., Blank, L., Marshall, S. J., Brown, N. L., and Morby, A. P. (1999). ZntR is a Zn(II)-responsive MerR-like

transcriptional regulator of zntA in Escherichia coli. *Molecular microbiology*, 31:893–902.

[18] Brondsted, L., Ostergaard, S., Pedersen, M., Hammer, K., and Vogensen, F. K. (2001). Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. *Virology*, 283:93–109.

[19] Buhr, E. D. and Takahashi, J. S. (2013). Molecular components of the Mammalian circadian clock. *Handbook of Experimental Pharmacology*, pages 3–27.

[20] Cameron, D. E. and Collins, J. J. (2014). Tunable protein degradation in bacteria. *Nature Biotechnology*, 32:1276–81.

[21] Camilli, A. and Mekalanos, J. J. (1995). Use of recombinase gene fusions to identify *Vibrio cholerae* genes induced during infection. *Molecular Microbiology*, 18(4):671–683.

[22] Cardinale, S. and Arkin, A. P. (2012). Contextualizing context for synthetic biology–identifying causes of failure of synthetic biological systems. *Biotechnology Journal*, 7(7):856–866.

[23] Carlin, A., Shi, W., Dey, S., and Rosen, B. P. (1995). The ars operon of Escherichia coli confers arsenical and antimonial resistance. *Journal of Bacteriology*, 177:981–6.

[24] Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., and Lafontaine, D. A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50):E3444–53.

[25] Casjens, S. R. and Hendrix, R. W. (2015). Bacteriophage lambda: early pioneer and still relevant. *Virology*, 0:310–330.

[26] Castillo, A. R., Woodruff, A. J., Connolly, L. E., Sause, W. E., and Ottemann, K. M. (2008). Recombination-based in vivo expression technology identifies *Helicobacter pylori* genes important for host colonization. *Infection and Immunity*, 76(12):5632–5644.

[27] Chant, A., Kraemer-Pecore, C. M., Watkin, R., and Kneale, G. G. (2005). Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expression and Purification*, 39(2):152–159.

[28] Chen, Y. Y., Galloway, K. E., and Smolke, C. D. (2012). Synthetic biology: advancing biological frontiers by building synthetic systems. *Genome Biology*, 13(2):240.

[29] Christiansen, B., Johnsen, M. G., Stenby, E., Vogensen, F. K., and Hammer, K. (1994). Characterization of the lactococcal temperate phage TP901-1 and its site-specific integration. *Journal of bacteriology*, 176:1069–76.

[30] Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C. M., Osbourn, A., Keasling, J. D., and Rosser, S. J. (2014). Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Research*, 42:e23.

[31] Cosentino, C. and Bates, D. (2011). *Feedback Control in Systems Biology*, volume 2. CRC Press.

[32] Courbet, A., Endy, D., Renard, E., Molina, F., and Bonnet, J. (2015). Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Science Translational Medicine*, 7(289):289ra83.

[33] Daniel, R., Rubens, J. R., Sarpeshkar, R., and Lu, T. K. (2013). Synthetic analog computation in living cells. *Nature*, 497(7451):619–623.

[34] Daruwalla, K. R., Paxton, A. T., and Henderson, P. J. (1981). Energization of

the transport systems for arabinose and comparison with galactose transport in Escherichia coli. *The Biochemical journal*, 200:611–27.

[35] Datta, K. and von Hippel, P. H. (2008). Direct spectroscopic study of reconstituted transcription complexes reveals that intrinsic termination is driven primarily by thermodynamic destabilization of the nucleic acid framework. *The Journal of Biological Chemistry*, 283:3537–49.

[36] Degenkolb, J., Takahashi, M., Ellestad, G. A., and Hillen, W. (1991). Structural requirements of tetracycline-Tet repressor interaction: determination of equilibrium binding constants for tetracycline analogs with the Tet repressor. *Antimicrobial Agents and Chemotherapy*, 35(8):1591–5.

[37] Diorio, C., Cai, J., Marmor, J., Shinder, R., and DuBow, M. S. (1995). An Escherichia coli chromosomal ars operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria. *Journal of bacteriology*, 177:2050–6.

[38] Divies, C. (1975). [Remarks on ethanol oxidation by an "Acetobacter xylinum" microbial electrode (author's transl)]. *Annales de Microbiologie*, 126:175–86.

[39] Dixon, N., Duncan, J. N., Geerlings, T., Dunstan, M. S., McCarthy, J. E. G., Leys, D., and Micklefield, J. (2010). Reengineering orthogonally selective riboswitches. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7):2830–5.

[40] Dixon, N., Robinson, C. J., Geerlings, T., Duncan, J. N., Drummond, S. P., and Micklefield, J. (2012). Orthogonal riboswitches for tuneable coexpression in bacteria. *Angewandte Chemie (International ed. in English)*, 51:3620–4.

[41] Domian, I. J., Reisenauer, A., and Shapiro, L. (1999). Feedback control of

a master bacterial cell-cycle regulator. *Proceedings of the National Academy of Sciences of the United States of America*, 96:6648–53.

[42] Dulbecco, R. and Vogt, M. (1954). Plaque formation and isolation of pure lines with poliomyelitis viruses. *The Journal of Experimental Medicine*, 99(2):167–182.

[43] Emadpour, M., Karcher, D., and Bock, R. (2015). Boosting riboswitch efficiency by RNA amplification. *Nucleic Acids Research*, 43(10):e66–e66.

[44] Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O., and Salzberg, S. L. (2000). Prediction of transcription terminators in bacterial genomes. *Journal of molecular biology*, 301:27–33.

[45] EspahBorujeni, A., Mishler, D. M., Wang, J., Huso, W., and Salis, H. M. (2016). Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Research*, 44(1):1–13.

[46] Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K. S., Chowdhury, A., and Hossain, M. N. (2013). Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant E. coli. *ISRN biotechnology*, 2013:590587.

[47] Farzadfard, F. and Lu, T. K. (2014). Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science*, 346(6211):1256272.

[48] Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A. A. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nature Reviews Microbiology*, 13(10):641–650.

[49] Ferrell, J. E. (2002). Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Current Opinion in Cell Biology*, 14(2):140–148.

[50] Folliard, T. (2017a). Ribo-attenuators: novel elements for reliable and modular riboswitch engineering - data. *https://doi.org/10.5287/bodleian:mNNpdYbpN*.

[51] Folliard, T. (2017b). A synthetic viral feedback loop results in robust expression. *https://doi.org/10.5287/bodleian:wrNKzxaY2*.

[52] Franklund, C. V. and Kadner, R. J. (1997). Multiple transcribed elements control expression of the Escherichia coli btuB gene. *Journal of bacteriology*, 179(12):4039–42.

[53] Galkin, V. E., Yu, X., Bielnicki, J., Ndjonka, D., Bell, C. E., and Egelman, E. H. (2009). Cleavage of bacteriophage lambda cI repressor involves the RecA C-terminal domain. *Journal of molecular biology*, 385:779–87.

[54] Ghosh, P., Wasil, L. R., and Hatfull, G. F. (2006). Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biology*, 4(6):e186.

[55] Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A. r., Smith, H. O., and Venter, J. C. (2010a). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (New York, N.Y.)*, 329:52–6.

[56] Gibson, D. G., Smith, H. O., Hutchison, C. A. r., Venter, J. C., and Merryman, C. (2010b). Chemical synthesis of the mouse mitochondrial genome. *Nature Methods*, 7:901–3.

[57] Gibson, D. G., Young, L., Chuang, R.-y., Venter, J. C., III, C. A. H., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345.

[58] Gilbert, S. D., Montange, R. K., Stoddard, C. D., and Batey, R. T. (2006). Struc-

tural studies of the purine and SAM binding riboswitches. *Cold Spring Harbor Symposia on Quantitative Biology*, 71:259–68.

[59] Gong, S., Wang, Y., and Zhang, W. (2015). Kinetic regulation mechanism of pbuE riboswitch. *The Journal of Chemical Physics*, 142(1):015103.

[60] Grass, G., Fan, B., Rosen, B. P., Franke, S., Nies, D. H., and Rensing, C. (2001). ZitB (YbgR), a member of the cation diffusion facilitator family, is an additional zinc transporter in Escherichia coli. *Journal of Bacteriology*, 183:4664–7.

[61] Grey, B. and Steck, T. R. (2001). Concentrations of copper thought to be toxic to escherichia coli can induce the viable but nonculturable condition. *Applied and Environmental Microbiology*, 67(11):5325–5327.

[62] Grindley, N. D. F., Whiteson, K. L., and Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annual Review of Biochemistry*, 75:567–605.

[63] Groisman, E., Cromie, M., Shi, Y., and Lafti, T. (2006). A Mg2+-responding RNA That Controls the Expression of a Mg2+ Transporter. *Cold Spring Harbor Symposia on Quantitative Biology*, 71(0):251–258.

[64] Gruber, T. M. and Gross, C. A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual review of microbiology*, 57:441–66.

[65] Hecht, A., Endy, D., Salit, M., and Munson, M. S. (2016). When Wavelengths Collide: Bias in Cell Abundance Measurements Due to Expressed Fluorescent Proteins. *ACS Synthetic Biology*, 5:1024–7.

[66] Hillson, N. J., Hu, P., Andersen, G. L., and Shapiro, L. (2007). Caulobacter crescentus as a Whole-Cell Uranium Biosensor. *Applied and Environmental Microbiology*, 73(23):7615–7621.

[67] Housden, N. G., Loftus, S. R., Moore, G. R., James, R., and Kleanthous, C.

(2005). Cell entry mechanism of enzymatic bacterial colicins: porin recruitment and the thermodynamics of receptor binding. *Proceedings of the National Academy of Sciences of the United States of America*, 102:13849–54.

[68] Hughes, M. F. (2002). Arsenic toxicity and potential mechanisms of action. *Toxicology letters*, 133:1–16.

[69] Iglesias, P. A. and Ingalls, B. P., editors (2010). *Control Theory and Systems Biology*. MIT Press, Cambridge, MA, USA.

[70] Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., and Collins, J. J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7):841–7.

[71] Jain, S. and Hatfull, G. F. (2000). Transcriptional regulation and immunity in mycobacteriophage Bxb1. *Molecular Microbiology*, 38(5):971–985.

[72] Ji, N., Middelkoop, T. C., Mentink, R. A., Betist, M. C., Tonegawa, S., Mooijman, D., Korswagen, H. C., and van Oudenaarden, A. (2013). Feedback control of gene expression variability in the Caenorhabditis elegans Wnt pathway. *Cell*, 155:869–80.

[73] Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., and Ptashne, M. (1981). lambda Repressor and cro–components of an efficient molecular switch. *Nature*, 294:217–23.

[74] Johnson, C. M. and Schleif, R. F. (1995). In vivo induction kinetics of the arabinose promoters in Escherichia coli. *Journal of Bacteriology*, 177(12):3438–42.

[75] Jusiak, B., Daniel, R., Farzadfard, F., Nissim, L., Purcell, O., Rubens, J., and Lu, T. K. (2014). Synthetic Gene Circuits. *Reviews in Cell Biology and Molecular Medicine*, pages 1–56.

[76] Kasahara, M. and Anraku, Y. (1972). Inhibition of the respiratory chain of Escherichia coli by zinc ions. *Journal of Biochemistry*, 72:777–81.

[77] Keravala, A. and Calos, M. P. (2008). Site-specific chromosomal integration mediated by phiC31 integrase. *Methods in Molecular Biology*, 435:165–173.

[78] Khlebnikov, A., Risa, O., Skaug, T., Carrier, T. A., and Keasling, J. D. (2000). Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *Journal of Bacteriology*, 182:7029–34.

[79] Kim, A. I., Ghosh, P., Aaron, M. A., Bibb, L. A., Jain, S., and Hatfull, G. F. (2003). Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. *Molecular microbiology*, 50:463–73.

[80] Kim, H. J., Lim, J. W., Jeong, H., Lee, S.-J., Lee, D.-W., Kim, T., and Lee, S. J. (2016). Development of a highly specific and sensitive cadmium and lead microbial biosensor using synthetic CadC-T7 genetic circuitry. *Biosensors & bioelectronics*, 79:701–8.

[81] Kleanthous, C. (2010). Translocator hunt comes full Cir-Col. *Molecular Microbiology*, 75:529–33.

[82] Kleanthous, C. and Walker, D. (2001). Immunity proteins: enzyme inhibitors that avoid the active site. *Trends in biochemical sciences*, 26:624–31.

[83] Koch, A. L. (1961). Some calculations on the turbidity of mitochondria and bacteria. *Biochimica et Biophysica Acta*, 51:429–41.

[84] Kumamoto, C. A., Oliver, D. B., and Beckwith, J. (1984). Signal sequence mutations disrupt feedback between secretion of an exported protein and its synthesis in E. coli. *Nature*, 308:863–4.

[85] Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227:680–5.

[86] Lee, N., Wilcox, G., Gielow, W., Arnold, J., Cleary, P., and Englesberg, E. (1974). In vitro activation of the transcription of araBAD operon by araC activator. *Proceedings of the National Academy of Sciences of the United States of America*, 71:634–8.

[87] Legramante, J. M., Raimondi, G., Massaro, M., and Iellamo, F. (2001). Positive and negative feedback mechanisms in the neural regulation of cardiovascular function in healthy and spinal cord-injured humans. *Circulation*, 103:1250–5.

[88] Li, L., Liang, J., Hong, W., Zhao, Y., Sun, S., Yang, X., Xu, A., Hang, H., Wu, L., and Chen, S. (2015). Evolved bacterial biosensor for arsenite detection in environmental water. *Environmental Science & Technology*, 49(10):6149–55.

[89] Li, S. and Breaker, R. R. (2013). Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic acids research*, 41(5):3022–31.

[90] Lin, H.-L., Lin, C.-C., Lin, Y.-J., Lin, H.-C., Shih, C.-M., Chen, C.-R., Huang, R.-N., and Kuo, T.-C. (2010). Revisiting with a relative-density calibration approach the determination of growth rates of microorganisms by use of optical density data from liquid cultures. *Applied and environmental microbiology*, 76:1683–5.

[91] Lin, Y.-H. and Chang, K.-Y. (2016). Rational design of a synthetic mammalian riboswitch as a ligand-responsive -1 ribosomal frame-shifting stimulator. *Nucleic acids research*.

[92] Lowe, A. M., Beattie, D. T., and Deresiewicz, R. L. (1998). Identification of novel staphylococcal virulence genes by in vivo expression technology. *Molecular Microbiology*, 27(5):967–976.

[93] Lussier, A., Bastet, L., Chauvier, A., and Lafontaine, D. A. (2015). A kissing

loop is important for btuB riboswitch ligand sensing and regulatory control. *The Journal of Biological Chemistry*, 290:26739–51.

[94] Lutz, R. and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*, 25:1203–10.

[95] MacPherson, A. J., Jones-Mortimer, M. C., and Henderson, P. J. (1981). Identification of the AraE transport protein of Escherichia coli. *The Biochemical journal*, 196:269–83.

[96] Madar, D., Dekel, E., Bren, A., and Alon, U. (2011). Negative auto-regulation increases the input dynamic-range of the arabinose system of *Escherichia coli*. *BMC Systems Biology*, 5:111.

[97] Mahr, R. and Frunzke, J. (2015). Transcription factor-based biosensors in biotechnology: current state and future prospects. *Applied Microbiology and Biotechnology*, 100:79–90.

[98] Mandal, M. and Breaker, R. R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature structural & molecular biology*, 11(1):29–35.

[99] Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, 101:7287–92.

[100] Mediavilla, J., Jain, S., Kriakov, J., Ford, M. E., Duda, R. L., Jacobs, W. R. J., Hendrix, R. W., and Hatfull, G. F. (2000). Genome organization and characterization of mycobacteriophage Bxb1. *Molecular Microbiology*, 38:955–70.

[101] Mertins, B., Psakis, G., Grosse, W., Back, K. C., Salisowski, A., Reiss, P., Ko-

ert, U., and Essen, L.-O. (2012). Flexibility of the N-terminal mVDAC1 segment controls the channel's gating behavior. *PloS One*, 7:e47938.

[102] Michener, J. K., Thodey, K., Liang, J. C., and Smolke, C. D. (2012). Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metabolic engineering*, 14:212–22.

[103] Morita, K., Yamamoto, T., Fusada, N., Komatsu, M., Ikeda, H., Hirano, N., and Takahashi, H. (2009). In vitro characterization of the site-specific recombination system based on actinophage TG1 integrase. *Molecular Genetics and Genomics*, 282(6):607–616.

[104] Munson, G. P., Lam, D. L., Outten, F. W., and O'Halloran, T. V. (2000). Identification of a copper-responsive two-component system on the chromosome of Escherichia coli K-12. *Journal of Bacteriology*, 182(20):5864–5871.

[105] Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, 10(4):354–360.

[106] Narberhaus, F. and Balsiger, S. (2003). Structure-function studies of Escherichia coli RpoH (sigma32) by in vitro linker insertion mutagenesis. *Journal of bacteriology*, 185:2731–8.

[107] Nawab, J., Khan, S., Shah, M. T., Khan, K., Huang, Q., and Ali, R. (2015). Quantification of Heavy Metals in Mining Affected Soil and Their Bioaccumulation in Native Plant Species. *International Journal of Phytoremediation*, 17:801–13.

[108] Nevozhay, D., Adams, R. M., , and Balázsi, G. (2011). Linearizer Gene Circuits with Negative Feedback Regulation. In Becskei, A., editor, *Yeast Genetic Networks: Methods and Protocols*, pages 81–100. Springer.

[109] Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K., and Balázsi, G. (2009). Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 106(13):5123–5128.

[110] Nies, D. H. (1999). Microbial heavy-metal resistance. *Applied microbiology and biotechnology*, 51:730–50.

[111] Nou, X. and Kadner, R. J. (1998). Coupled changes in translation and transcription during cobalamin-dependent regulation of btuB expression in Escherichia coli. *Journal of bacteriology*, 180(24):6719–28.

[112] O'Donnell, S. M. and Janssen, G. R. (2001). The initiation codon affects ribosome binding and translational efficiency in Escherichia coli of cI mRNA with or without the 5′ untranslated leader. *Journal of bacteriology*, 183:1277–83.

[113] Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature structural biology*, 7:215–9.

[114] Outten, C. E., Outten, F. W., and O'Halloran, T. V. (1999). DNA distortion mechanism for transcriptional activation by ZntR, a Zn(II)-responsive MerR homologue in Escherichia coli. *The Journal of biological chemistry*, 274:37517–24.

[115] Oyarzn, D. A., Lugagne, J.-B., and Stan, G.-B. V. (2015). Noise Propagation in Synthetic Gene Circuits for Metabolic Control. *ACS Synth. Biol.*, 4(2):116–125.

[116] Papadakos, G., Wojdyla, J. A., and Kleanthous, C. (2012). Nuclease colicins and their immunity proteins. *Quarterly reviews of biophysics*, 45:57–103.

[117] Park, M., Tsai, S.-L., and Chen, W. (2013). Microbial biosensors: engineered microorganisms as the sensing machinery. *Sensors (Basel, Switzerland)*, 13:5777–95.

[118] Pedelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, 24:79–88.

[119] Phan, T. T. P. and Schumann, W. (2007). Development of a glycine-inducible expression system for Bacillus subtilis. *Journal of Biotechnology*, 128(3):486–499.

[120] Pope, B. and Kent, H. M. (1996). High efficiency 5 min transformation of Escherichia coli. *Nucleic Acids Research*, 24(3):536–537.

[121] Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., and Cormier, M. J. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 111:229–33.

[122] Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R., and Tobes, R. (2005). The TetR family of transcriptional repressors. *Microbiology and Molecular Biology Reviews*, 69:326–56.

[123] Rao, C. V. and Koirala, S. (2014). Black and white with some shades of grey: the diverse responses of inducible metabolic pathways in Escherichia coli. *Molecular microbiology*, 93:1079–83.

[124] Ravikumar, S., Ganesh, I., keun Yoo, I., and Hong, S. H. (2012). Construction of a bacterial biosensor for zinc and copper and its application to the development of multifunctional heavy metal adsorption bacteria. *Process Biochemistry*, 47(5):758 – 765.

[125] Reining, A., Nozinovic, S., Schlepckow, K., Buhr, F., Fürtig, B., and Schwalbe, H. (2013). Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature*, 499(7458):355–359.

[126] Rensing, C. and Grass, G. (2003). Escherichia coli mechanisms of copper

homeostasis in a changing environment. *FEMS Microbiology Reviews*, 27:197–213.

[127] Rensing, C. and Maier, R. M. (2003). Issues underlying use of biosensors to measure metal bioavailability. *Ecotoxicology and environmental safety*, 56:140–7.

[128] Rensing, C., Mitra, B., and Rosen, B. P. (1997). The zntA gene of Escherichia coli encodes a Zn(II)-translocating P-type ATPase. *Proceedings of the National Academy of Sciences of the United States of America*, 94:14326–31.

[129] Renzi, F., Panetta, G., Vallone, B., Brunori, M., Arceci, M., Bozzoni, I., Laneve, P., and Caffarelli, E. (2006). Large-scale purification and crystallization of the endoribonuclease XendoU: troubleshooting with His-tagged proteins. *Acta crystallographica. Section F, Structural biology and Crystallization Communications*, 62(Pt 3):298–301.

[130] Reuter, J. S. and Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC bioinformatics*, 11:129.

[131] Ridley, H. and Lakey, J. H. (2015). Antibacterial toxin colicin N and phage protein G3p compete with TolB for a binding site on TolA. *Microbiology (Reading, England)*, 161:503–15.

[132] Rieder, R., Lang, K., Graber, D., and Micura, R. (2007). Ligand-Induced Folding of the Adenosine Deaminase A-Riboswitch and Implications on Riboswitch Translational Control. *ChemBioChem*, 8(8):896–902.

[133] Rinaldi, A. J., Lund, P. E., Blanco, M. R., and Walter, N. G. (2016). The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. *Nature Communications*, 7:8976.

[134] Rodrigo, G., Landrain, T. E., Majer, E., Dars, J.-A., and Jaramillo, A. (2013). Full Design Automation of Multi-State RNA Devices to Program Gene Expression Using Energy-Based Optimization. *PLoS Computational Biology*, 9(8):1–11.

[135] Rodriguez-Montelongo, L., de la Cruz-Rodriguez, L. C., Farias, R. N., and Massa, E. M. (1993). Membrane-associated redox cycling of copper mediates hydroperoxide toxicity in Escherichia coli. *Biochimica et biophysica acta*, 1144:77–84.

[136] Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology*, 323(5):785–793.

[137] Ross, W., Aiyar, S. E., Salomon, J., and Gourse, R. L. (1998). Escherichia coli promoters with UP elements of different strengths: modular structure of bacterial promoters. *Journal of Bacteriology*, 180:5375–83.

[138] Russell, J. P., Chang, D. W., Tretiakova, A., and Padidam, M. (2006). Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *BioTechniques*, 40:460, 462, 464.

[139] Salis, H. M., Mirsky, E. A., and Voigt, C. a. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nucleic Acids Reseach*, 27(10):946–50.

[140] Savinov, A., Pan, J., Ghosh, P., and Hatfull, G. F. (2012). The Bxb1 gp47 recombination directionality factor is required not only for prophage excision, but also for phage DNA replication. *Gene*, 495(1):42–48.

[141] Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS microbiology reviews*, 34:779–96.

[142] Serganov, A., Huang, L., and Patel, D. J. (2008). Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature*, 455(7217):1263–1267.

[143] Serganov, A. and Patel, D. J. (2007). Ribozymes, riboswitches and be-

yond: regulation of gene expression without proteins. *Nature Reviews Genetics*, 8(10):776–790.

[144] Serganov, A., Yuan, Y.-R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R., and Patel, D. J. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry & biology*, 11:1729–41.

[145] Shcherbo, D., Merzlyak, E. M., Chepurnykh, T. V., Fradkov, A. F., Ermakova, G. V., Solovieva, E. A., Lukyanov, K. A., Bogdanova, E. A., Zaraisky, A. G., Lukyanov, S., and Chudakov, D. M. (2007). Bright far-red fluorescent protein for whole-body imaging. *Nature methods*, 4:741–6.

[146] Shcherbo, D., Murphy, C. S., Ermakova, G. V., Solovieva, E. A., Chepurnykh, T. V., Shcheglov, A. S., Verkhusha, V. V., Pletnev, V. Z., Hazelwood, K. L., Roche, P. M., Lukyanov, S., Zaraisky, A. G., Davidson, M. W., and Chudakov, D. M. (2009). Far-red fluorescent tags for protein imaging in living tissues. *The Biochemical Journal*, 418:567–74.

[147] Sheppard, D. E. and Englesberg, E. (1967). Further evidence for positive control of the L-arabinose system by gene araC. *Journal of Molecular Biology*, 25:443–54.

[148] Shi, W., Wu, J., and Rosen, B. P. (1994). Identification of a putative metal binding site in a new family of metalloregulatory proteins. *The Journal of biological chemistry*, 269:19826–9.

[149] Shimoga, V., White, J. T., Li, Y., Sontag, E., and Bleris, L. (2013). Synthetic mammalian transgene negative autoregulation. *Molecular systems biology*, 9:670.

[150] Shimomura, O., Johnson, F. H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hy-

dromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*, 59:223–39.

[151] Shine, J. and Dalgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. *Nature*, 254:34–8.

[152] Siuti, P., Yazbek, J., and Lu, T. K. (2013). Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology*, 31(5):448–452.

[153] Smith, M. C. M. and Thorpe, H. M. (2002). Diversity in the serine recombinases. *Molecular Microbiology*, 44(2):299–307.

[154] Sternberg, N. and Hamilton, D. (1981). Bacteriophage P1 site-specific recombination. *Journal of Molecular Biology*, 150(4):467–486.

[155] Szybalski, W. and Skalka, A. (1978). Nobel prizes and restriction enzymes. In *Gene*, volume 4, pages 181–2, Netherlands.

[156] Tchounwou, P. B., Yedjou, C. G., Patlolla, A. K., and Sutton, D. J. (2012). Heavy Metals Toxicity and the Environment. *Molecular, Clinical and Environmental Toxicology*, 101:133–164.

[157] Thattai, M. and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98:8614–9.

[158] Thevenot, D. R., Toth, K., Durst, R. A., and Wilson, G. S. (2001). Electrochemical biosensors: recommended definitions and classification. Technical report, England.

[159] Tobin, J. F. and Schleif, R. F. (1987). Positive regulation of the Escherichia coli L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. *Journal of Molecular Biology*, 196:789–99.

[160] Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer

of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America*, 76:4350–4.

[161] Tsien, R. Y. (1998). The green fluorescent protein. *Annual review of biochemistry*, 67:509–44.

[162] Viswanathan, S., Rani, C., Ribeiro, S., and Delerue-Matos, C. (2012). Molecular imprinted nanoelectrodes for ultra sensitive detection of ovarian cancer marker. *Biosensors & Bioelectronics*, 33:179–83.

[163] Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P. F., and Mörl, M. (2013). De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research*, 41(4):2541–51.

[164] Wacker, A., Buck, J., Mathieu, D., Richter, C., Wohnert, J., and Schwalbe, H. (2011). Structure and dynamics of the deoxyguanosine-sensing riboswitch studied by NMR-spectroscopy. *Nucleic Acids Research*, 39(15):6802–6812.

[165] Wang, Y., Yau, Y.-Y., Perkins-Balding, D., and Thomson, J. G. (2011). Recombinase technology: applications and possibilities. *Plant cell reports*, 30:267–85.

[166] Wen, G., Wen, X., Shuang, S., and Choi, M. M. (2014). Whole-cell biosensor for determination of methanol. *Sensors and Actuators B: Chemical*, 201:586 – 591.

[167] Wieland, M. and Hartig, J. S. (2008). Artificial Riboswitches: Synthetic mRNA-Based Regulators of Gene Expression. *ChemBioChem*, 9(12):1873–1878.

[168] Winkler, W., Nahvi, A., and Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, 419(6910):952–6.

[169] Wittmann, A. and Suess, B. (2011). Selection of tetracycline inducible self-

cleaving ribozymes as synthetic devices for gene regulation in yeast. *Molecular BioSystems*, 7(8):2419.

[170] Woestenenk, E. A., Hammarström, M., van den Berg, S., Härd, T., and Berglund, H. (2004). His tag effect on solubility of human proteins produced in Escherichia coli: a comparison between four expression vectors. *Journal of Structural and Functional Genomics*, 5(3):217–29.

[171] Wolf, D. and Mascher, T. (2016). The applied side of antimicrobial peptide-inducible promoters from Firmicutes bacteria: expression systems and whole-cell biosensors. *Applied Microbiology and Biotechnology*, 100:4817–29.

[172] Wu, J. and Filutowicz, M. (1999). Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. *Acta Biochimica Polonica*, 46(3):591–9.

[173] Xu, Z. and Brown, W. R. A. (2016). Comparison and optimization of ten phage encoded serine integrases for genome engineering in Saccharomyces cerevisiae. *BMC biotechnology*, 16:13.

[174] Yamamoto, H., Wittek, D., Gupta, R., Qin, B., Ueda, T., Krause, R., Yamamoto, K., Albrecht, R., Pech, M., and Nierhaus, K. H. (2016). 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 113:E1180–9.

[175] Yamamoto, K. and Ishihama, A. (2005). Transcriptional response of Escherichia coli to external zinc. *Journal of Bacteriology*, 187:6333–40.

[176] Yamamoto, K. and Ishihama, A. (2006). Characterization of copper-inducible promoters regulated by CpxA/CpxR in Escherichia coli. *Bioscience, biotechnology, and biochemistry*, 70:1688–95.

[177] Yang, L., Nielsen, A. A. K., Fernandez-Rodriguez, J., McClune, C. J., Laub,

M. T., Lu, T. K., and Voigt, C. A. (2014). Permanent genetic memory with greater1-byte capacity. *Nature Methods*, 11(12):1261–1266.

[178] Yang, Y., Guo, L.-H., Qu, N., Wei, M.-Y., Zhao, L.-X., and Wan, B. (2011). Label-free electrochemical measurement of protein tyrosine kinase activity and inhibition based on electro-catalyzed tyrosine signaling. *Biosensors & bioelectronics*, 28:284–90.

[179] Yeo, J., Park, J.-Y., Bae, W. J., Lee, Y. S., Kim, B. H., Cho, Y., and Park, S.-M. (2009). Label-free electrochemical detection of the p53 core domain protein on its antibody immobilized electrode. *Analytical Chemistry*, 81:4770–7.

[180] Zhang, F., Carothers, J. M., and Keasling, J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature biotechnology*, 30:354–9.

[181] Zhang, M., Mullens, C., and Gorski, W. (2005). Insulin oxidation and determination at carbon electrodes. *Analytical chemistry*, 77:6396–401.

[182] Zheng, G., Patolsky, F., Cui, Y., Wang, W. U., and Lieber, C. M. (2005). Multiplexed electrical detection of cancer markers with nanowire sensor arrays. *Nature Biotechnology*, 23:1294–301.

[183] Zhu, X., Wang, X., Zhang, C., Wang, X., and Gu, Q. (2015). A riboswitch sensor to determine vitamin B12 in fermented foods. *Food Chemistry*, 175:523–8.